



Universitetet
i Stavanger

Measuring non- technical skills in a Norwegian Air Ambulance medical scenario using the customized rating instrument AeroNOTS.

**Institutt for Helsefag
Master in Health Science**

**Student: Solveig Gjertsen
Supervisor: Marianne Storm**

Date: June 29th, 2017

Number of words: 10.937 + 5.077 = 16.014

UNIVERSITETET I STAVANGER

**MASTERSTUDIUM I HELSEVITENSKAP
MASTERSTUDIE**

SEMESTER: Våren 2017

FORFATTER/MASTERKANDIDAT: Solveig Gjertsen

VEILEDER: Marianne Storm

TITTEL PÅ MASTEROPPGAVE:

Måling av ikke-tekniske ferdigheter hos anestesileger i Norsk Luftambulans ved hjelp av det tilpassede måleverktøyet AeroNOTS.

Measuring non-technical skills in a Norwegian Air Ambulance medical scenario using the customized rating instrument AeroNOTS.

EMNEORD/STIKKORD:

Non-technical skills, Air ambulance, patient transport, clinical training, simulation, AeroNOTS.

ANTALL SIDER: 88

STAVANGER 29. juni, 2017.



” Team-based delivery addresses the concept of how we live and operate in this high-stakes environment- create a realistic environment; initiate a team-based application of psychomotor, cognitive, and affective skill sets presented in the context of real life events.”

David Matic, 2012

Content

PART 1

Content.....	4
Acknowledgement.....	7
Abstract.....	7
1.0 Introduction.....	8
1.1 Aim of the study.....	10
1.2 Key concepts.....	11
1.2.1 Medical simulation.....	11
1.2.2 Defining non-technical skills (NTS).....	12
1.3 Previous research on medical simulation and assessment of NTS.....	13
2.0 Theoretical framework:.....	15
2.1 Crisis resource management (CRM).....	15
2.1.1 Know your environment.....	17
2.1.2 Anticipate, share and review the plan.....	17
2.1.3 Ensure leadership and role-clarity.....	17
2.1.4 Communicate effectively.....	18
2.1.5 Call for help early.....	18
2.1.6 Allocate attention wisely- avoid fixation.....	18
2.1.7 Distribute the work-load.....	18
3.0 Simulation.....	20
4.0 Method.....	21
4.1 Design.....	21
4.2 Participants and field of research.....	21
4.3 Data collection and ethical considerations.....	22
4.4 Simulation scenario.....	23
4.5 AeroNOTS and translation.....	24
4.6 Observation.....	25

4.7	The observers role.....	26
4.8	Validation.....	27
4.8.1	Validation concerns.....	27
4.8.2	Statistical analysis and validity.....	28
4.8.3	Validation and behavioral markers.....	30
4.9	Data analysis.....	31
5.0	Results.....	32
6.0	Discussion.....	33
6.1	CRM and NTS.....	33
6.2	Validation and rating of NTS.....	37
6.3	Content, construct, external and face-validity of AeroNOTS.....	38
6.4	NTS rating instruments in HEMS.....	39
7.0	Conclusion.....	40

PART 2: PAPER

Abstract.....	43
Background.....	43
Method.....	45
Study design and participants.....	45
Validation.....	45
Observation tool and translation.....	47
Data collection.....	47
Simulation scenario.....	48
Statistical methods.....	49
Results.....	50
Ethical considerations.....	51
Discussion.....	52
Limitations.....	55

Conclusion.....	56
References.....	58

APPENDIX

List of figures:

Figure 1:	Bland Altman plot and Limits of agreement (LoA).....	69
Figure 2:	Cohens Kappa - agreement between the two assessors.....	70
Figure 3:	Boxplot of years of experience vs 15 elements in AeroNOTS...	71

List of tables

Table 1:	Demographics of participants.....	72
Table 2:	Original AeroNOTS observation form (2015).....	73
Table 3:	Translated AeroNOTS (Norwegian) observation form (2016)..	74
Table 4:	Translated table of scoring system, AeroNOTS (Norwegian)....	76
Table 5:	AeroNOTS observasjonsskjema – Norwegian short version.....	77

List of attachments

Attachment 1: Forespørsel om deltakelse i forskningsprosjekt.....	78
Attachment 2: Tilbakemelding på melding om behandling av personopplysninger (NSD).....	80
Attachment 3: Guidelines of Scandinavian Journal of Trauma and Rectitation.....	82

Acknowledgement

This paper has been realized in cooperation with the Foundation of the Norwegian Air Ambulance (SNLA). They have contributed to support the work by initiating the theme for the study and by being a gate opener to the field of research. They have also given financially support to make data collection possible. One employee in the SNLA, Knut Styrksson, has been crucial to collect the data. In addition to this, SNLA have given a great contribution to the project by offering personnel resources with statistical, research- and medical competence supporting this study. Thanks to Elisabeth Jeppesen, Jo Røislien and Håkon Abrahamsen for their efforts.

My supervisor, Marianne Storm, has throughout this work been of great importance guiding and supporting me to complete this project. Thank you for patience and support!

To refine the language, my co-student, Sigrun Anna Qvindesland, has made a great contribution. My husband, Håvard Vetrhus, has also contributed by correcting language and technical support. Thank you!

In order to translate the assessment tool AeroNOTS, Hans Erik Bugge, has done the re-translation from Norwegian to English. He is working as a 1st amanuensis at University of Stavanger (Faculty of humanism) and has excellent English proficiency.

All of these contributors have made this project an interesting and educational journey.

Abstract

The work presented in this paper was conducted in cooperation with the Foundation of the Norwegian Air Ambulance to explore to which extent it is possible to measure skills connected to teamwork and cooperation. Because real life observations in this environment was not possible of ethical and safety reasons, simulation was as close as we could get to assess the helicopter emergency medical services (HEMS) and assess their team work. The first part of the work should be an overview over previous research in this field concerning simulation based training and non-technical skills (NTS). The aim of this study was presented as were the theoretical framework, crisis resource management (CRM). The methodical aspects were presented and conducted information

about observation, participants, measurement instrument (= assessment tool) and the simulated scenario conducted outdoor by the waterside in the woods of Holmenkollen, outside Oslo.

The second part was a paper presenting background for the study, previous research on the theme, methods and field of research. In addition the result were presented and discussed.

As attachment 3 shows, the style of references in the journal were the paper should seek submittence (Scandinavian Journal of Trauma & Rhesusitation), is different from APA 6th style which have been used throughout all this work. This will be changed before submittence in accordance to my supervisor.

1.0 Introduction

In 2012, World Health Organization (2012) made a report on team work in health care called "To err is human". This report focus on the variety and complexness of teamwork and point out the nature of the teamwork based on implementation of crisis recourse management (CRM) principles which are defined by the aviation industry as:" using all available sources - information, equipment, and people - to achieve safe and efficient flight operations" (WHO, 2012, doc 1.5, p.3)

These principles have been adopted and used in health care to improve team work and initiate safe processes. Among the skills implemented in CRM-principles are NTS as they focus on cooperation and team communication. To enhance such skills, simulation has been used as a method. During the last two decades there has been an increasing interest in simulation as a method of learning, especially in the field of critical care treatment of patients (Bredemose, 2010; Cooper, 2011; Ballangrud, 2014; Garden, 2012). These studies show that simulation is an efficient method of learning Technical and Non-Technical Skills. (Abrahamsen, Sollid, Ølund, Røislien & Bondevik.,2015; Flin, 2000).

Anesthetics and surgeons work in a high-stake environment where work performance depends heavily on good team performance. The benefits of medical simulation to improve technical and interdisciplinary skills and team performance are highly recognized (Yule, Parker, Wilkinson, McKinley, McDonald, Neill & McAdam, 2015;

Ahmed, Miscovic, Darzi & Athanasiou, 2011). In some universities, students also have mandatory simulation training to enhance students' teamwork and clinical skills. (Husebø, 2012; Qvindesland, Bjørshol, Aase, Rossavik & Kluge, 2015). According to these authors, this training can prepare and enhance good team work among health care providers.

In the HEMS, simulation are used as part of annual training programme. The Norwegian government has guidelines (Nasjonal standard for luftambulanselager, 25.aug 2011. www.regjeringen.no/st.m.no.43/ URL) imposing pilots and HEMS paramedics to have yearly mandatory simulation training. The aviation industry has during the last few decades implemented simulation training to maintain and strengthen technical and NTS professionally. This means that this kind of training is included in their working schedule, and that courses have to be passed to continue their work duties. Also offshore workers in health care (nurses and medics) have such claims in Norway. This is a stronger claim than physicians have according to Abrahamsen et.al (2015). They do not have similar training requirement as they are not employed as flight personnel, but employed at their regional hospitals. HEMS physicians therefor have less simulation hours than pilots and HEMS paramedics. (Abrahamsen et.al., 2015). This study argues that there is a need for more training hours for the HEMS physicians including valid systems for evaluating simulation training.

During the last decade, there has been developed some rating instruments being able to measure the level of NTS (Jepsen, Ødegaard & Diekman, 2015; Rehim, DeMoor, Olmsted, Dent & Parker Raley, 2017). These skills refer to task management, team work, situation awareness, leadership and communication skills. The first assessment tool included behavioral markers for anesthetists in the operating theatre and intensive care units. Foster (1995) define behavioral markers broadly as "motor behavior, cognitive and affective events, and psychological responses".

This assessment tool was called anesthetists non- technical skills. From this instrument there were customized other assessment tools for other fields in health care, e.g. by surgeons, scrub nurses, nurse anesthetists and emergency teams (Cooper, Cant, Conell, Sims, Porter, Symmons, Nestel & Liaw, 2016; Fletcher, Flin, McGeorge, Maran & Patey, 2003; Shazrinizam & Neary, 2014). Julia Myers and colleagues (2016) developed and customized an instrument for HEMS: AeroNOTS. This assessment tool has much in

common with former frameworks for assessing behavioral markers. Myers assessment tool, AeroNOTS, have adapted most of the main categories of anesthetist non-technical skills form like task management, team-work, situation awareness and decision making. Still it differs in one important area: communication. AreoNOTS does not have a single element named communication, but include this in other elements of the assessment tool.

HEMS is an isolated and challenging work place in health care, and claims of safety and high performance are crucial (www.regjeringen.st.m 43, 2011). Valid and applicable assessment tools for this area of health care may contribute to safe and professional treatment and care.

1.1 Aim of the study.

This study will apply Myers framework to a HEMS simulated scenario. This study will use Myers assessment tool in assessing the participants non-technical skills performed in a simulated scenario. The purpose is to examine the tools applicability and validity. Validity will be investigated through statistical tests, and applicability assessed through the two observers experiences by using this specific tool. This will be done through our translation of the assessment tool. This process will demonstrate how well the content has been adapted.

Our scientific questions are:

Is AeroNOTS a valid assessment tool for NTS in a Norwegian Air Ambulance setting?

Are years of experience as a prehospital anesthetist a significant factor for high performance in NTS?

This study focuses on the air medical physicians and their performance on the NTS. Myers (2016) suggests that this rating instrument may have the potential being used for other participants of the HEMS crew, such as pilots and HEMS paramedics.

Hopefully this study can contribute to add an objective and valid rating instrument for the valuable simulation based training already performed in HEMS. This may also be a tool for the management to evaluate their employees NTS and follow up if scores are low.

This study will explore if Myers assessment tool can be a useful contribution to the existing assessment tools for NTS.

1.2 Key concepts.

1.2.1 Medical simulation.

Medical simulation is defined by David Matics as:

”Team-based delivery addresses the concept of how we live and operate in this high-stakes environment- create a realistic environment; initiate a team-based application of psychomotor, cognitive, and affective skill sets presented in the context of real life events; identify and measure outcomes based on objectives for the event; and provide the actual learning stage through a debriefing in which students can actively reflect in open dialogue” (Matic, 2015, s.261).

As medical simulation as a learning methods has been in use since the 1980`s, there is some evidence for its relevance. Qualitative and quantitative research claims this method can be useful for students and more experienced healthcare providers (Ballangrud, Hall-Lord, Persenius & Hedelin, 2015; Brindley, Suen & Drummond, 2007; Clapper, 2015). Simulations preference compared to classroom teaching is that the participant has an active role throughout the scenario. In the debriefing part the participants can be able to explore gaps in performance which should be locked or repeated (Garden, LeFevre, Wadington & Wellers, 2015).

The simulation setting gives the participants the ability to work in a safe environment which is as close to real life as possible. The situation (scenario) can be a daily action or a very rare one, depending on the purpose with the simulation training (Gaba, 2001). The facilitators (instructors) role is to plan the scenario to be as realistic as possible and to put the participants at scene. The facilitator task is also to conduct debriefing after the scenario. The purpose with this is to bring the participants experiences in the scenario to a conscious level trough reflection and discussion, and being able to learn from oneself or the other participants. Lack of knowledge or skills can be discovered, and gaps may be bridged before it causes any harm to a patient. Good or excellent performance should also

be recognized and hopefully repeated (Gorden, Darbyshare & Baker, 2012; Garden et.al., 2015).

1.2.2 Defining non-technical skills (NTS)

NTS can be defined as “the cognitive, social resource/skills that complement technical skills and contribute to safe and efficient task performance” (Flin, O`Connor & Crichton, 2008, p.84).

One example of a NTS may be planning and preparing for the situation to come. This includes picking up necessary equipment you may need, preparing medication for this particular case, sharing information about the patient you expect with the team and calling for extra personnel if required. If these tasks are not taken care of, it will affect the patient negatively, and the resources are not being used effectively.

Adoption of standardized tools has been an effective strategy in enhancing teamwork and reducing risk. Additionally, it also contribute to a culture change were the team`s work have a higher priority than the individual expertise (Leonard, 2004). The awareness of how important communication and teamwork is in addition to safe care and treatment, is a change of paradigm in health care (Leonard, 2004; Gaba, 2001).

During the last twenty years the development of assessment tools for NTS has raised. It started with anesthetist non-technical skills (ANTS) in 1997 to measure NTS in the operating theatre and at the recovery unit. From then several forms have been customized for different fields inside and outside hospitals. 35 of these tools have been evaluated in the systematic review studies of Cooper et.al. (2014) and Jepsen et.al. (2012). Validity seems to be difficult to prove, due to limited study samples and different methods.

Myers (2016) operationalizes NTS in four main categories in AeroNOTS. Each category connects three to five elements. The main categories are described as: Task management, teamwork, situation awareness and decision making. The assessment tool will be presented later. Table 1 gives an overview.

HEMS in Norway have been using simulation to strengthen the teamwork for several years. To find a valid and useful instrument for this part of health care could clearly enhance safety and work confidence among HEMS physicians.

1.3 Previous research on medical simulation and assessment of NTS

Literature search in following databases was performed:

CINAHL, Medline and Cochrane Library using keyword like simulation, non-technical skills, physicians, air- ambulance, AeroNOTS and crew resource management. Several articles were found, and the following articles were considered relevant for this study.

Garden et.al. (2015) review of 27 articles describes the benefits of simulation as the opportunity to provide efficient and timely acquisition of technical and NTS with the provision of feedback. This systematic review, including eight papers, also describes the benefits of debriefing especially for learning NTS. There is evidence for the improved skills by simulation training according to Gardens review. The conclusion is that even if scoring systems exist, they are rarely used in debriefing and simulation. Until these systems are being used there will be "*a lack of empirical data regarding to non-technical skill performance*" (Garden, 2015, s.306). This may be an argument for exploring these instruments and validate them. Then they can be a contributor in debriefing.

Even if medical simulation started among anesthetists in the late 80`s, this method of learning have spread to other areas of health care like elderly people, surgeons, nurses at wards, leaders and nursing and physicians students (Ross, Anderson, Kodate, Thomas, Thompson, Thomas, Key, Jensen, Schiff &Jaye,2012; Shazrinizam & Neary, 2014; Cooper, Porter & Peach,2014; Robertson & Bandali, 2008; Flynn/Sandaker, 2009).

Gordens systematic review (2012) includes 20 studies of NTS training. The included studies course concepts varied in time and content, but all included educational interventions to improve NTS of staff working in an acute health care environment. All course had elements of simulation based training. The method of the review was to put each study into a manuscript screening tool developed by the authors. This tool was characterized by author, year, location, study type, participants, intervention, outcome measures, results, conclusions, level of outcomes and strength of conclusion. The study concludes that the studies included had reasonable methodical quality and that there were a significant uniformity of the content of interventions which referred to five key areas: error, communication, teamwork, leadership and situational awareness.

According to the survey of 207 participants, Abrahamsen et.al. (2012) found significant differences between the amounts of simulation based training of physicians in the Norwegian Air Ambulance compared to pilots and rescue paramedics in HEMS. Their study underpins the value of simulation as a tool for increased NTS and indicates that all of the crew members lack sufficient training in these skills despite governmental claims (www.regjeringen.no st.m.43).

Cooper et.al. (2011) studied 24 teams incorporating nurses, physicians and other health related workers. They focused on some NTS like leadership and team-working skills. Their study indicated increased skills at the end of the course according NTS. The self-reporting system the participants made, were described as satisfying by the participants. To explore the effect and give simulation training an objective assessment, rating instruments have been developed during the last two decades. Kirkpatric did an innovational work on evaluation programmes in the 60`s - this has been refined and used in the field of simulation until today (Kirkpatric, 1996). His program focuses on simple and practical, let participants describe feelings and reactions of what happened in the scenario. The participants evaluate their learning effect of the scenario, to which extent they change their behavior at work, and how they could improve their skills to conduct a better job (Kirkpatric, 1996). This evaluation was a self-rating system, and till today this has been the most common way of rating the participants in simulation training.

Even if self-rating systems give some directions of the efficiency of simulation, there has been proved that this method has weaknesses. They have a high degree of being subjective, and are not necessary connected to the learning goals. Kirkpatric`s (1996) advice is to develop more specific criteria for evaluation and use control groups, if possible, to make sure results are valid and reliable.

Jepsen, Østergaard and Dieckman (2015) identified assessment tools for NTS in different areas of health care, but mainly for handling emergency situations, in their critical review. Most of the instruments consist of almost the same categories of NTS and many use behavioral markers. The study contains an overview of 23 rating instruments characterized by name, purpose/users, main sources of data and scoring system. Validity procedures and reliability are also being assessed. The authors conclude that there is a need to focus on validity of assessment tools and training of raters in using the tools

(Jepsen et.al., 2015). However, none of these assessment tools referred by Jepsen et.al are customized for air ambulance personnel.

Julia Myers et.al. (2016) developed a HEMS assessment tool for NTS called AeroNOTS. Customizing the instrument from anesthetist non-technical skills is the first phase of the study. Myers framework is the only one so far that is developed and adapted for HEMS. Second phase of this study is testing the framework at 16 physicians in an air medical critical care environment using videos from scenarios conducted at a hospital in New Zealand. The results show that the framework was useful and may examine differences in the performance of NTS, especially to identify when specific non- technical factors are likely to break down.

Myers` observational study is using two observers who had calibrated the framework. Data indicates that experienced physicians had higher scores than less experienced physicians, and that there was a correlation between general performance and observation forms. Myers also did self-rating and proved differences between the experienced and less experienced physicians. Results indicate that the inexperienced doctors assessed their performance to appear better than the experienced compared to the observed score.

There were also seen correlations between the two observers indicating that they agreed on the ratings. Myers` framework can be supplementary to other rating systems for NTS and it may be part of the safety culture in the air ambulance environment.

2.0 Theoretical framework.

2.1 Crisis resource management (CRM)

Gaba states that the CRM can be summarized as: “*the articulation of principles of individual and crew behavior in ordinary and crisis situations that focuses on skills of dynamic decision making, interpersonal behavior, and team management*” (Gaba, 2010, p.3).

CRM has been highlighted as one of the basic theoretical elements of simulation based training (Hughes,Benenson,Krichten, Clancy,Ryan & Hammond, 2014; Carne,

Kennedy & Gray, 2011) and has background from the aviation industry after row of serious accidents in the 70's. Through investigations there were demonstrated a need of tools and frameworks that could enhance safety (Hughes et.al., 2014). CRM principles and rating instruments developed for use in the aviation industry have been adapted and customized for healthcare environments. Prehospital emergency medicine and anesthetists working in the operating theatre were among the first to implement these principles in their work. Their need for being optimally prepared for emergency situations can explain why they have embraced these theories. (Gaba, 2010) CRM and NTS are heavily dependent on each other to underpin the strategy of the team management. To focus on these strategies and techniques, simulation has been used to make health care workers realize how they use their knowledge and skills. The way they lead a group, or organize the treatment or support their colleagues can make a great difference in the outcome for the patient. Still we don't know how much training is needed and how frequent, but we certainly know that "one shot" is not enough (Gaba, 2010).

CRM highlights the necessity of efficient communication and how to perform good teamwork to avoid adverse situations for the patients. CRM focus on the individual perspective of cooperating. Team resource management (TRM) describes how to strengthen the team through attitudes, organization and communication skills.

CRM, TRM and NTS are strongly connected and dependent on each other. CRM are mainly based on principles from Gaba's and Sala's work (Carne et.al., 2011). These principles have been modified several times, but the key principles are:

- Know your environment
- Anticipate, share and review the plan
- Ensure leadership and role clarity
- Communicate effectively
- Call for help early
- Allocate attention wisely, avoid fixation
- Distribute the work load, monitor and support team members

(Carne et.al., 2011, p.8).

The connection between CRM and NTS is seen when quality healthcare is provided. NTS is a set of behavioral markers that characterize a person, e.g. "using assertiveness" or "gathering information" or "assessing capabilities". These may vary according to each person and situation. The CRM principles may be understood as a "gold standard" for treatment and care provided. (Carne et. al., 2011). High levels on NTS may deliver efficient and safe health care according to the CRM principles. Description of main CRM-principles as described by Carne et.al., 2011:

2.1.1 Know your environment

This principle refers to get to know the physical workplace, where things are being stored, locations you may need to know and time penalty of finding equipment.

It also refers to what human resources you have access to, what competence your colleagues have, leader-style of the leaders, and if there is a collaborative approach to dealing with conflicts.

2.1.2 Anticipate, share and review the plan

Emergency medicine, rely on planning and preparation when dealing with intra or inter-hospital transport of unwell patients. It means to anticipate delays, inform patient and plan for what to do in the meantime. Personal factors like being hungry, angry, late, tired and stressed, lack of knowledge, illness, inexperience and environmental issues like interruptions, handovers production pressure and equipment failure can be avoided by planning well and share it with the other involved. Working together to achieve the common goal, is critical to effective teamwork. Reviewing the plan and maintaining dynamic skepticism towards previous diagnosis are important to detect errors and adapt planned actions.

2.1.3 Ensure leadership and role clarity.

The leader's role is critical and should be performed in a participative manner and with least conferential approach. The leader should also ensure that roles are defined for each team member. Leaders who allocates the team members, maintain the team's shared cognition and monitor the internal and external environments will probably ensure that the team "keep on track". The team members' task is to support the leader with relevant information, responsible task work, monitor and support other team members and corrective action if needed.

2.1.4 Communicate effectively

Communication failures are the leading cause of patient harm in health care. In the team setting, effective communication distributes needed information to other team members and facilitates continuous updating of the shared mental model. Cultural factors and task load in a stressful situation could be reasons why messages are missed or interpreted incorrectly. Communication tool like closed loop communication and SBAR can be an effective way of combat information exchange difficulties. Closed loop information involves that sender and receiver of a message state the message loud and clear to avoid misunderstandings. It is also called “confirmed communication”.

SBAR is a recipe when transferring information to colleagues or other health providers.

S = Present yourself and give a brief summary of the **situation**.

B = Do the person know the patient previously? If not, give some **background** information.

A = what is the **actual** reason to call for help / need more help?

R = what kind of **recommendation** do you need?

In order to avoid communication failures, it is important to acknowledge that even if we treat the same patient, we may have different perspectives. These tools may enhance patient safety because communication can be more standardized.

2.1.5 Call for help early.

The timely involvement of appropriate expertise can impact upon patient outcomes in cases of serious illness. Emergency departments and wards have “alert buttons” to call for help of colleagues and in-hospital “emergency alarm” when patients have a cardiac arrest. Unexperienced physicians normally have the possibility to call a senior for advice when unsecure. Cultural attitudes can prevent action to be taken, and this could, in worst case, effect the patient condition.

2.1.6 Allocate attention wisely - avoid fixation.

There is recognized tendency when stressed to focus on one particular issue that may lie in control of the stressed individual. For anesthetists an example could be to focus on intubation in a cardiac arrest situation, and not to get oxygen to the patient. The

compression may be delayed and the patient life at risk. This is called “fixation error” and reduces situation awareness.

The role of the leader may also remain empty because the leader himself is busy with providing CPR. A wise allocation would be that the leader delegates tasks to others to reduce the work load and be able to see the “big picture” and make his decisions.

2.1.7 Distribute the workload - monitor and support team members.

The leader should maintain the awareness of the big picture and contribute to overall patient management by ensuring that his/her staff is supported to safely see patients in appropriate time frames rather than taking large workloads themselves.

In a trauma team in emergency departments there is often formalized distribution of tasks, and this may contribute to an integrated and comprehensive approach to critically ill patients.

The job of emergency physicians is a complex one, where application of the key components of CRM can result in better performance and outcomes in the management of patients with acute illness and injury.

The development of CRM from the beginning of 1980 have focused on NTS as a factor to improve health care and patient security (Reeves et.al.,2013) Health care services, particularly in emergency medicine, have found it beneficial to look upon the behavioral markers that can be crucial when treating a serious ill patient in a team. These markers are individual characteristics and skills that are involved in the group process and affect the quality and outcome for each patient. The individual characteristic includes knowledge, attitudes, motivation and personality. It also includes the ability to do certain tasks, teamwork, decision making, and situation awareness and stress management. All together these skills will affect the outcome of team performance, individual performance and individual job satisfaction. (Reeves et.al., 2013)

Gaba claims that these principles should be applied in health care for all personnel groups throughout their careers using simulation based training repetitively to increase technical and non- technical skills, attitudes and ultimately for outcomes of patient care processes (Gaba, 2010).

Adoption of standardized tools has been an effective strategy in enhancing teamwork and reducing risk. Additionally, it also contribute to a culture change were the team`s work have a higher priority than the individual expertise (Leonard, 2004). The awareness of how important communication and teamwork is in addition to safe care and treatment is a change of paradigm in health care (Leonard, 2004; Gaba, 2001).

Anesthetists have developed these principles into a course concept called Anesthetists crew resource management (ACRM) starting up in 1990 (Gaba, 2010). Other parts of health care have also adapted these concepts and customized it for their domain. Especially skills in communication and leadership have been addressed and found useful in these curriculums (Gaba,2010). An important question is: Do these techniques of teamwork improve patient care and outcome? Yet we don`t know the answer to this question, because to measure reliability and validity in these domains has been challenging. This is also the conclusion of Jepsen`s et.al. review of 23 different assessment tools for NTS (2013) - these studies had mostly low quality and validity and reliability was hard to confess.

3.0 Simulation

Simulation is based on the theoretical frameworks of learning theory and adult learning as described by Kolb and Bloom among others in the mid-fifties (Kolb, 1984). Their ideas have inspired to a different way of learning where the teacher is a facilitator who should let the participant skills and knowledge come to the surface (Kaufman, 2004). To achieve this, all the participants need to feel safe and comfortable. The facilitator needs skills and knowledge about CRM to establish this. According to Kolbs theories (Kolb, 1984) the participants need support and challenges to expose skills and knowledge they have, and learn from one and others. If failures occur during a scenario, the important part to learn is *why* it happened. If the participant understands this, he is able to change his behavior in real life. This means that instead of criticizing the participant, the facilitator asks questions to explore and understand the background for action. Positive and negative gaps in performance can be explored and recognized (Rudolph et.al., 2006).

The development of assessment tools for NTS has risen from 1997(anesthetist non-technical skills) . From then several forms have been customized for different fields

inside and outside hospitals. 35 of these tools have been evaluated in the systematic review studies of Cooper et.al. (2014) and Jepsen et.al. (2012). Validity seems to be difficult to prove, due to limited study samples and different methods.

4.0 Methods

4.1 Design

The design for this study is univariate and descriptive including a structured observation with two observers using protocols with a rating system. As the study takes place outdoor in a simulation based scenario, it is very close to real life and the participants normal work situation. Ringsdal (2001) claims the aim of the study should guide the design and mention mainly three categories of design: Explorative, descriptive or explanatory. Our design is quantitative and descriptive, meaning that we describe the statistical data of our rating instrument.

”The purpose of descriptive studies is to observe, describe, and document aspects of a situation as it naturally occurs and sometime to serve as starting point for hypothesis generation or theory development.” (Polit & Beck, 2012, p. 226).

4.2 Participants and field of research

The participants of the study are physicians working in the Norwegian Air Ambulance and signed up for this training course which is mandatory for all physicians in this field. The course lasted for four days in total and one of these days included medical simulation. The course is run 6 weeks in a row every year. All the physicians were anesthetists working at the regional hospitals in addition to their duty at the air ambulance transfer. All of the physicians who could be included volunteered. Six physicians who had volunteered had to be excluded as they attended three crews in the scenario. This meant less workload on each participant and may be a possible bias due to better scores in our data collection. Both observers shared the opinion on this exclusion and found it important that all participants worked under the same conditions. In this study 24 physicians were included. They were from all districts of Norway, all together nine different aircraft bases out of 12. Some had little experience and other had been working

in the field for decades. Differences in age and how much training they had access to, were also seen.

4.3 Data collection and ethical considerations

The observation of the participants took place every Wednesday between August 24th and September 28th 2016. Two AeroNOTS observers performed the observation. The same scenario was performed every week. The facilitators were different every week, but they presented and evaluated the case mostly in the same manner.

The participants were invited to the study 1-2 weeks ahead of the course, and information about the aim and the method were given in an information letter (attachment 1). Arriving the course, brief information was given, and they were able to sign up by putting their signature at a consent statement where they also put notes about age, gender, working place and experience.

The permission was distributed after a short meeting the morning before the observations. The evening before observation we were able to socialize informally with the participants to get an idea of the culture in the research field. This might have made it easier for the physicians to volunteer to the study. All of the invited physicians accepted participation.

The observation started when the participants arrived at the simulation site by the lake, and stopped at the end of the scenario when they had planned their departure. The ratings were done during and as soon as possible after the scenario had ended.

In cases where HEMS crew split up and treated the patient (mannequin) in two different places, the assessors followed one crew each. This was a challenge we had not discussed in front of the observation. The fact that both of the assessors could not observe the second crew working simultaneously meant that the scores for these crews were unsatisfactory. This was the case with three of the observed crews. We tried to make notes on how unsatisfactory (by using percentage score) these scores were, but this was difficult to transfer to the database. One could discuss whether these observations (six physicians) should be excluded, but we decided to include them because of lack of observations and that we had some relevant data that could be useful.

4.4 Simulation scenario

The data was collected during one medical scenario located outdoor by the waterside of a small lake in a hillside near Oslo (Holmenkollen). The scenario was made as an extraordinary stressful situation for the participants and they needed two crews to handle the situation. The learning goals for the participants were to perform the new AHLR logarithm for children and present good teamwork using all resources.

This is the scenario that took place:

Three children are missing by the waterfront. The mother for one of them is present, shouting and crying for help. A neighbor of her is also present and has called for help. The mother and her neighbor is played by volunteers. Sink- mannequins were used as the missing children. They look like children at the age of 8-10 years old. They have been put into the water before the scenario starts with no visible signs.

As the first crew enters the waterside, they meet the mother and the other woman (neighbor). The crew arrive the scene after a 3 minutes' drive in a vehicle. This is the time they have for preparing and planning their actions. Observation starts when they arrives the waterside. Depending on their planning, their actions are different, but all of the observed crews used their HEMS paramedics in the water wearing diving suits. This implicates that the physicians and the pilots have to cooperate adequately to succeed. They will have to scramble for more help (another crew, rescue team, ambulances) and they have to report to the coordinating instance (AMK) on their progress in work, discussing which hospital to be chosen to which child. At the same time they have to maintain standards to treat the children correctly and organize the two crews internally. They have a shortage of people and have to use all resources adequate to show excellent performance. Even if each of the three occupational groups has their own domains, they have to be creative and communicate clearly, to get all tasks done.

The scenario last for 30-40 minutes and ended up as both crews had planned and prepared for their department. The way they solved the scenario varied significantly. This had implications to their NTS scores which are described under the chapter results.

The two observers are hidden behind a car when the first team arrives to not disturb the participants from their tasks. When they start to work, we are standing as close to the teams as possible without interrupting, usually 1-2 meters from their treatment area. We can hear the dialogues and see their actions. In three of the scenarios they were treating

the patients at two different places at the same time. Then the two observers had to split up and followed one crew each in periods of the scenario. This meant that the opposite crew got a less valid score, but still we could keep track of the situation as a hole.

After the scenario was ended, all the participants had a debriefing session with trained facilitators. This lasted for about 45 minutes and all the participants had to reflect on their actions to increase their learning. The facilitators` role was to point out gaps in performance and make each participant realize what had been good, or what could have been improved, but this debriefing had no impact on the AeroNOTS score given by the assessors. The facilitators did not score the participants, but made notes for their debriefing.

4.5 AeroNOTS and Translation

Myers (2016) operationalizes NTS in four main categories in AeroNOTS. Each category connects three to five elements. These are:

1. **Task management** refers to planning and preparing equipment and tasks, prioritizing and re-evaluates the situation, maintaining standards/ procedures and identifying and using resources.
2. **Team work** refers to cooperation of team activities, exchanging information, using authority and assertiveness, assessing capabilities and supporting others.
3. **Situation awareness** means to gather information, recognizing and understanding and anticipating
4. **Decision making** refers to identifying options, balance risks and selecting options and re-evaluating.(Table 1)

To each of the 15 elements there is made a description to explain the content of this exact behavioral marker. This is to help the assessor/observer detect this particular behavior easier. An example of description connected to “exchanging information” could be:

“Giving and receiving the knowledge and data necessary for team coordination and task completion”.

Another one connected to “Supporting others” is: “Providing physical, cognitive or emotional help to others members of the team”. These descriptions are seen in other rating instruments as well (Jepsen et.al.,2015; Patey, Flin, Fletcher, Maran & Glavin,2017).

Refinements must be done through evaluation and use to make categories, elements and descriptions as specific as possible.

The translation from English to Norwegian was primarily done by one of the assessors. The form contains of 4 main areas and 15 elements with an attached description. The other assessor also contributed with his comments after the first version were made. We had no controversies in our understanding of the translation. Then the translated version and the English version were handed to a person without medical background but with good English proficiency. He agreed in the translation of the two assessors. Finally the Norwegian translation was handed to a professional translator. He commented on some grammar diversities in the extended version of the form related to the use of present and past. This was corrected due to the translator's recommendations. No major diversities were found in the main areas and the elements in the translated form (Table 3).

There was also made a translation of the assessment scale for clinical performance (Table 4). The translation was made as described above attached to the main form.

For practical reasons we also made a short version of the form (Table 5). This contains of the main categories and the 15 elements but without descriptions.

4.6 Observation

Observation was chosen on the background of Myers study (2016) that was an important inspiration for this study. Additionally, judgement of the participants, time, geographic and proximity to the research field were done, as described by Ringdal (2001). We were able to focus on a small geographic area (a small lake), a reasonable amount of participants (25-30) and the training camp were limited in time (6 weeks) All these factors made observation possible. Compared to Myers study, two observers used videos of the participants in a simulated scenario in a hospital setting, and made their scores on the basis of these.

Foster & Cone (1996) argues that to establish accuracy in research of behavioral markers, real-time observations of ongoing performance of the behavior in the natural environment is preferred. Accuracy means "the extent to which scores on a measure reflect the "true" properties of the behavior" according to Foster & Cone (1996, p.254) Our observation is as close to real-time observation you can get without observing the physicians at work.

The scenarios were carried out outdoor identic to scenarios the physicians experience at work. A lot of efforts and planning were carried out to make each scenario as realistic as possible. Evidence demonstrate that simulation with a high degree of realism is efficient to establish a good learning environment for technical and non- technical skills (Ross et.al.,2012; Shazrinizam & Neary,2014; Cooper et.al.,2011; Robertson & Bandali,2008; Flynn & Sandaker, 2009).

Sorensen & Stanton (2015) argues that when focusing on situation awareness in real world tasks, inter-rater reliability has most value. Situation awareness is one of the main categories in our rating instrument. On this background we established to observers to be able to measure inter-rater reliability. This was a parallel to Myers study (2016).

4.7 The observers role

Observation could include different paradigms, such as quantitative designs, depending on the observers role (Glesne, 2006; Ringdal, 2001). In this study the observer has a role “without participation” and is as little involved in the scenario as possible (Thagaard, 2013). Both observers did their scores as close to the participants as possible to see and to hear them act, but we were aware that we did not disturb them when in action. The observers never spoke to the participants during the scenario. The aim was to score their NTS according to the form, without any intervention in the scenario. Thagaard (2013) points out that if the participants are occupied and concentrated on their tasks, they might be less disturbed by the scientist presence. The scenario was extremely complex and the participants had to concentrate on a high level for their task management. This could mean that the disturbance of the scientist was minor.

Thagaard also mention the scientist role in the field. The importance of being accepted in the field you study is essential (Hammersley &Atkinson, 2004). Both observers were “outsiders”, which means that the observers were not part of the participant’s daily work. We decided to spend time with the participants socially the day before the observations were carried out. In this way we established a relation to them, and we also used our prior work experience to connect to their professional stand. Thagaard (2013) claims this can be important to establish comfort, safety and confidence.

The two observers uses Myers observation form with four main categories divided into three- five elements each (Table 3, 4 and 5). The observers are both nurses, one a nurse

anesthetist with no experience in prehospital work, but 20 years of in hospital experience in the operating theatre. The second, a paramedic nurse who have most of his experience in prehospital service, during the last 30 years as an ambulance paramedic.

This was an open observation, which means that the participants knew that they were being observed. According to Thagaard (2013) it is preferred that the scientist announces his precens. On the other hand, there might be a chance for the participants to act differently because they know they are being observed, and this could be a bias.

Even if one of the observers had his work in the Foundation of The Norwegian Air Ambulance, none of us were colleagues of the participants. There has been some discussion on how the observers position as an “insider” or an “outsider” influences on the participants. We were both “outsiders” with a basic understanding of the participants work. We had both long experience on working in teams, and since the purpose of the study was in this field, it`s reasonable to claim that we had “positioned insight” as Thagaard (2013) mention as important to be able to interpret what is seen.

4.8 Validation

This part will describe characteristics of validity, and these will be discussed further in the discussion compared to the results of this study.

Shadish, Cook & Campbell (2002) define validity in the context of research design as “the approximate truth of an inference” (p. 34). They also claims that validity always is a matter of degree, not an absolute. Chronbach and Meel`s have a broader view according to Shadish et.al.(2002), that validity describes the meaning of scores produced by a measurement instrument or procedure (Foster & Cone,1995) As validity is the property of an inference, the design elements only effects the inference that can be made. This means that validity is not a research design, but is strongly connected to the choice of design.

4.8.1 Validation concerns

Five areas of validity will be briefly described: Statistical conclusion validity, construct validity external validity, content validity and face validity as describes by Polit & Beck (2012), chapter 10 and 14.

1. Statistical conclusion validity means that the researcher concludes that there truly is an empirical relationship between the presumed cause and the effect. The researcher must provide evidence as strong as possible to prove this.
2. Construct validity involves to which degree an measurement scale is a good representation of the underlying construct that was theorized as having the potential to cause beneficial outcomes, and if the independent variable are good operationalized of the construct for which they are intended.
3. External validity concerns about replication and whether an observed relationship will hold over variations in persons, across settings, time or measure of outcome. External validity is about the generalizability of causal inferences.
4. Content validity may be defined as “the extent to which an instrument`s content adequately capture the construct- that is, whether an instrument has an appropriate sample of items for the construct being measured” (Polit & Beck, 2012, p.310). An expert group or guidelines may be used to enhance this part, and statistical test measure their agreement.
5. Face validity refers to whether an instrument looks like it is measuring the target construct according to Polit & Beck(2012, p.310). This is not a strong evidence of validity, but can be useful if the participants have resistance to being measured because they do not consider it relevant for themselves. All these parts of validity will be discussed according to output data. As our method contains two observers, we are especially interested in the inter-rater reliability.

4.8.2 Statistical analysis and validity.

This study tends to validate an assessment tool for behavioral markers as defined by Foster & Cone (1995) and Myers et.al. (2016).

Foster & Cone also claims that if measure is sound, its` scores provide generalizable information about occurrence, duration, latency, or magnitude of the performance being assessed (Foster & Cone, 1995). To demonstrate this we used three different statistical tests in our assessment tool, AeroNOTS.

Bland Altmans analyze

In his paper Bland & Altman (1986) argue for using this test instead of correlation tests for certain issues. They claim that correlation test is misleading because they measure the degree of differences of equal methods - this does not necessary mean that they agree.

They state that measurement method for comparison data, correlation coefficient or regression analysis is not appropriate.

Bland Altman's analyze agreement between two tests performed in the same manner. As there were two assessors measuring the same situation, this method fits the purpose. The test demonstrates the total scores for two assessors in a boxplot with a central 45-degree line. If the two assessors confirm, the dots will be presented along the 45-degree line mostly. If they do not confirm, the dots will be spread. The longer distance from the line, the less confirm the assessors. This is the first phase of Bland Altman's analyze-where the two assessors are presented separately (figure 1A).

Next step in the Bland Altman's analyze is to plot the mean of the differences between the two assessors to assess skewness- this is seen as a dotted line in the boxplot. If normally distributed Limits of Agreement (LoA) of these mean values are $\pm 1,96 \cdot SD$ (figure 1B). This means that 95% of the differences in scores for the assessors is plotted in this area.

Cohens Kappa (*k*)

To assess agreement between the two assessors, Cohens Kappa is used. This test is commonly used in the medical literature to assess inter-rater agreement or consistency of two different diagnostic tests. Cohens Kappa is an estimate of the proportion of agreement between two raters (or instruments) that takes into account the amount of agreement that could have occurred by chance (Pallant, 2012). The reference values for *k* variates between 0-1, where 0.5 represent moderate agreement, 0.7 good agreement and 0.9 very good agreement.

Cohen's kappa is used to calculate the sensitivity and specificity of a measure. In our study, sensitivity refers to the portion of cases that are correctly observed as behavioral described in our assessment tool, AeroNOTS.

Cohen's Kappa is designed for measuring ratings made by two raters and was considered a preferred method for our study. Sorensen & Stanton (2015) claims that inter-rater agreement may be useful to measure situation awareness which is an aspect of NTS. Situation awareness is one of our four main categories to be assessed.

Cohens Kappa can be compared with p-value, but have different numbers on the scale. A p-value of 0.005(significant) can be compared with a *k*-value of 0,5 which means moderate agreement.

(Pallant, 2012)

Kruskal Wallis test

The most common use of the Kruskal-Wallis test is when you have one nominal variable and one measure variable, and does not presume that data are normally distributed (McDonald, 2014). The test is sometimes called non-parametric one-way ANOVA. This test was used to explore the connection between the physicians experience and the total scores onNTS reported by the two assessors. The p-value is corresponding to the chi-square test, mostly used for parametric data. The reason for using this test is primarily because it is non-parametric and do not presume normally distribution, as was the case for our data.

4.8.3 Validity and behavioral markers

Foster & Cone (1995) mention that contextual variables may affect the result of measurement, meaning that relationship between measures can vary, depending on the context in which they are examined. In our observation setting the physicians experience in prehospital medicine could be challenged because of other persons in the scenario failed or misunderstood what was going on. They also address two important issues according to validity with measures of behavior: content validity and accuracy (Foster & Cone, 1995)

In the AeroNOTS assessment tool the content validity refers to in which extent the behavior is described as a specific trait that can be scored in its own right, not as a sign of some abstraction. AeroNOTS tool has four main categories, divided into three to five elements connected to each category. This is the way the instrument try to enhance content validity. In addition, there is a description for each of the 15 elements of the assessment tool to describe how the behavior be observed.

An example of this is the element “decision making”:

Decision making	Identifying options	Generating alternative possibilities or courses of actions to be considered in making a decision or solving a problem
	Balancing risks & selecting options	Assessing hazards to weigh up the threats or benefits of a situation, considering the advantages and disadvantages of different courses of actions based on these processes
	Re-evaluating	Continually reviewing the suitability of the options identified, assessed and selected; and reprocessing the situation following the implementation of a given action

This description may be helpful for the observers to identify and exclude certain behavior of the participants, and make sure that scores are fulfilling. In order to assess content validity, this may be evaluated as positive about the instrument.

Foster & Cane (1996) also mention accuracy in connection to behavioral assessment. Three ways of establishing accuracy are described:

- a) Physical evidence of behavior (e.g. sweat, reading written journals)
- b) Real-time observations in natural environment, or
- c) Controlled stimuli that can be created that depict variations in the behavior to which the measure should be sensitive.

They also describe direct observation as the hallmark of objective assessment.

To enhance accuracy, a well-planned and organized simulated scenario, including two observers to detect the behavior described in our assessment tool simultaneously, can be particularly beneficial.

4.9 Data analysis

All the data were organized in the SPSS as a dataset. The information about each participant on gender, age, experience each in one row. Then the observation form were divided into each element and registered with the scores. There were four main categories of the observation form. These were: task management, teamwork, situation awareness and decision making. The 3-5 elements in each category were summed and divided on

elements in each group to a mean value. There were also made a total score for the two assessors by summarizing all values from each participant.

To analyze the data there were used frequencies test, Kruskal Wallis test and Cohens Kappa from SPSS, and one test from R: Bland- Altman analysis. These were preferred due to evidence on non-parametric and continuous data (Altman & Royston, 2006).

Our focus is to assess similarities between the two assessors. For this purpose a Bland Altman analysis was preferred. This test gives more reliable outcome when looking for concurrence

(Altman & Royston, 2006). We made a stand that the clinical relevance of the individual rating should be above three, which mean good performance and above. This was done due to face validity and because rating variance above three should have no clinical effect for the patient outcome. Clinical relevance is important to evaluate the observation form not only from a statistically view, but also how the form may be used later on. From our perspective, a score under three, most of the elements may lead to reports and follow-ups.

5.0 Results

The results of this study reflects other studies during the last ten to fifteen years: To measure NTS using rating instruments may be difficult, and that validity may be hard to prove (Jepsen et.al.,2013 Gorden et.al.,2012).

Demographic data demonstrates that the 24 participants (two women and 22 men) had average age of 46 years old and came from 9 out of 14 helicopter bases from all over the country. They had an average in experience as a anesthetist for 13.57 years, and the mean time spend for training non-technical skills were 3.25 hours per month.

The Bland Altman analysis had an overall low agreement measuring the total scores of the two assessors. The agreement interval ranges from – 21 to 20 of a total score up to 75 (Figure 1). This means that the assessors score varies and without a particular pattern.

Cohens Kappa (k) tests each element (15) of the rating instrument and compares the two assessor results on each element/item of the assessment tool. k is the value measuring agreement among these two, ranging from 0-1 in 0.2 points steps. 0-0.2 means “slight agreement” and 0.8-1 mean ”almost perfect agreement”.

In this study five elements had fair agreement ($k = 0.2-0.4$). These were: “Coordinating activities”, “Exchanging information”, “Using authority and assertiveness”, ”gathering information” and “balancing risks and selecting options”. Even if agreement is seen as fair, the elements that confirms are sensitive for communication skills, which is an interesting finding. This will be further discussed.

Most of the other elements were measured as “no agreement” (0), ”Slight agreement”(0-0,2).

The Kruskal Wallis test was used to look for connections between the physicians experience and their NTS with a significant level at $< 0,005$. Only one element was measured as significant: ”planning and preparing”. Statistically this finding has low strength, but there might be a slight indication of that long experience may alter good NTS as most of the participants had more than ten years of experience and they all had scores mainly above 3, which means acceptable skills.

6.0 Discussion

This chapter will discuss the result on the background of theoretical framework and validity concerns.

6.1 CRM and NTS

CRM can be seen as the basis of the instruments that has been developed for NTS. This has happened as a parallel to the extended use of simulation as a method to increase technical- and NTS. Bosseau, Murray & Foster (2000) points out certain principles that are at stake when a team is going to cooperate, such as:

- ✓ Leadership /function of a leader
- ✓ Communication
- ✓ Continuous reassessment
- ✓ Use all available resources
- ✓ Avoidance of fixation
- ✓ Consideration of personal traits

These principles are mainly the same as those described by Carne et.al. (2012). In order to demonstrate that CRM is not specifically defined but is more a set of principles, these have been picked and found appropriate for our setting.

All of these characteristics or behavior markers can be recognized in AeroNOTS assessment tool.

Leadership represent “ Prioritizing”, “ identifying and utilizing resources”, “ coordinating activities in the team”, “using authority and assertiveness” and “assessing capabilities” and are key leadership characteristics in HEMS.

Study result shows that the physicians most often were scored above 3 (table 2), which is acceptable (Figure 3). They have long experience, and the HEMS physicians have the main responsibility for the medical decisions. This means they have to adapt skills to get track of the situation (situation awareness) and lead the other team members. This means that most of the physicians make good prioritizing, identify and use resources, coordinate others, are assertive and authoritarian and assess capabilities throughout this scenario.

Communication among HEMS team members are captured in the elements “Exchanging information”, “supporting others”, “gathering information” and “re-evaluation” in AeroNOTS. Communication is a challenge for HEMS personnel. The helicopter itself is a noisy environment, which is an obstacle for communication between the team members. Even if they do not work in the helicopter, there may be factors in the environment that can challenge exchange of information, e.g. wind and waterfalls, traffic noise or noise from the rotor. In the simulated scenario, the assessors observed traffic noise from the road along the waterfront. Distance between the team members, which was particular challenging when the two crews split up and worked with their patient in two different stations, was a hinder for communication.

Cohens Kappa demonstrated, the communication skills were rated by both assessors to have “fair agreement” ($k = 0.2-0.4$). This means that this rating instrument fairly good enhance the communication skills of these physicians. This is one of the positive findings in our study, as the area of communication is where AeroNOTS differs from most other rating instruments (Myers, 2016, Jepsen et.al.,2013, Gorden et.al.,2012).

Continuous re-assessment represents the element in AeroNOTS called “Re-evaluating”. Re-evaluating means that the physicians have to use “time-outs” in the

situation, plan next step and sharing relevant information with the rest of the team. In the simulation scenario, there was limited time to stop working - everyone had lifesaving tasks to perform. It was extremely important though to keep track of the situation for everyone and to distribute the resources where they were needed. In particular there were three phases in the scenario where re-evaluating was important; when the first child were rescued, when the second child were found, and before the helicopter take-off. The Cohens Kappa test demonstrated slight agreement ($k = 0,055$) between the two assessors, on this element and the scores were most often above 4. This result could indicate that the three phases described above were captured by both assessors in a similar manner. It could indicate that the simulation participants had an overall good score and were able to re-assess the situation at these three points.

Use all available resources has an equivalence in AeroNOTS called “identifying and utilizing resources”, but also “assessing capabilities” and “coordinating activities in the team” can be categorized here. This means that everyone in the team should have tasks to do, and that team members can contribute with information and tasks that is relevant. The importance of speaking up as a team member must not be underestimated according to Carne et.al.(2012). The reason for this is that information may be lost, tasks not taken care of and patient`s safety challenged. This means that the resource refers to mental support and task management. If participants in the team for different reason do not come forward with their thoughts and ideas, this can affect the patient`s condition. It could also mean that if a team member is afraid of authorities or think that their information or question is not of any importance, they certainly will not speak up. An open and collaborative climate may be established to enhance patient safety.

In our study these elements had slight agreement according to Cohens Kappa. Still the scores for the elements were mostly around four, which is good. These skills are complicated to achieve and claims experience and training in real life settings. The participants had more than 10 year of experience and in addition they have monthly training hours which may be the reason for their good skills.

Avoidance of fixation may be recognized in elements in AeroNOTS as ”prioritizing”, “recognizing and understanding”, and “identifying options”. Avoidance of fixation means that the participant get busy with details and loses the sight of the “big picture”. This is

an important issue that can be the reason of defocusing and putting the patient's life at risk (Carne et.al.,2012).

In our study these elements demonstrates slight agreement and fair agreement with scores 3-4 according to Cohens Kappa which can be considered as acceptable. It is reasonable to believe that in a less stressful scenario, these score will improve on these participants. Stress is able to defocus attention and cause fixation error (Carne et.al.,2012). This scenario was very stressful, but compared to the scores at Cohens Kappa it seems like these physicians were focused on their tasks and had a good overview of the situation.

Consideration of personal traits can be recognized in AeroNOTS, but not as a specific element. This behavioral marker refers to using people in the team wisely, as some are better skilled in some areas. For the safety and efficiency of the patient, everyone should perform as good as possible.

In HEMS, the number of persons is limited and roles predefined as a three-piece crew is represented by a physician, pilot and HEMS paramedic. The roles are set in advance in most situations. Even though, support and allocations of tasks may occur between the paramedic and the physician. Also the pilot may be relocated.

But the HEMS paramedic also has a lot of tasks that the physician do not have regarding rescue procedures.

In our scenario there were two crews simultaneously in action. The allocation of people was done mostly in order to get more resources, e.g. when the physicians allocate from search and rescue procedures performing CPR when a child is found.

The HEMS physicians are extremely dependent on their teams' skills to provide safe and efficient treatment. The unique working conditions of HEMS personnel, the environmental challenges and limited recourses make their situation vulnerable. This is why it is important to ensure that training is standardized and equal for all crew members. Abrahamsen et.al. (2016) underline the importance of simulation-based training for the physicians in HEMS. This should be at the same levels as pilots and HEMS paramedics, which it is not today.

It is clear, though, that AeroNOTS enhance the CRM principles and are recognized in all the 15 elements of the rating instrument. NTS as described in AeroNOTS seems to be relevant and recognizable in our group of participants.

6.2 Validation and rating of NTS

As referred validation of existing assessment tools and rating instruments for NTS seems hard to prove (Jepsen et.al.,2013, Gorden, Darbyshire & Baker,2012), our study had challenges connected to validity. Some will be discussed here.

Statistical conclusion validity was seen to some extent in the three tests that were performed. The Bland Altman plot demonstrates how the two assessors total scores variates separately and as a mean. The LoA has a wide range between - 20 to 19 (table 1) which means the assessors scores is varying to a certain extent. Before conducting this analysis we set a lower value of score (3) for clinical relevance as presented by Bland & Altman (1986).This was done to separate the acceptable performance from the less good. This meant that all scores above 3 (totalscoe 45) could be considered as acceptable. Most of the participants had scores above 3 in all 15 elements (Figure 2).This may be evidence for using AeroNOTS in addition to debriefing to identify if some of the team members have gaps in performance that need to be followed up. The reason for the variation of the assessor's scores may have other reasons than the rating instrument itself. The fact that we choose not to calibrate how to use it and our background as "outsiders" with positioned insight may have biased our result in this test. AeroNOTS is simple to use in the field and may be a neutral and objective assessment tool for HEMS personnel.

The Cohens Kappa had the strongest proof as five of 15 elements had fair agreement between the two assessors (Figure 2). All of the five elements that had fair agreement were connected to communication. Even if the agreement were not too strong, the finding indicates that this rating instrument is sensitive to measure communication skills, which can be considered positive. The greatest difference among AeroNOTS and other rating instruments for NTS, is the lack of a specific element named "communication". As referred HEMS personnel works in a noisy and stressed environment where communication can be challenged. The HEMS personnel's ability and skills to communicate effectively is vital (Myers, 2016). The AeroNOTS assessment tool seems to be an instrument able to assess vital elements of these communication strains.

The Kruskal Wallis test had significant result in one of 15 elements at a p-level < 0,005.

This cannot be considered as a strong statistical proof of the connection between experience and a high level at NTS. This test demonstrated some areas with scores of 4-5, which is very good. There were clearly demonstrated that they were high-level

performers in vital areas (Figure 3) such as “ maintaining standards”, “exchanging information”, “ using authority and assertiveness”, “assessing capabilities”, “supporting others” and “gathering information”. Also in this test the items concerning communication stands out, which is positive. One reason for this finding may be that most of the physicians had more than ten years of experience and could be considered experienced. The biases mentioned due to Bland Altmans analyze may also interfere the result at this analysis, as there were some variations in the scores by the two assessors.

6.3 Content, construct external and face validity of AeroNOTS.

Assessment of content validity

AeroNOTS has been developed from a validated instrument: anesthetists’ non-technical skills and customized for HEMS personnel. In addition to content validity, this is important, because most of the elements in the rating instrument has been used and tested for more than twenty years.

Content validity refers to in which extent the elements can be interpreted and measured as they are seen (Polit & Beck, 2012). AeroNOTS` implementation of already validated categories and elements may serve as a validation of content.

The explanations to each element should also enhance content validity. These may contribute to a broader understanding of each element to understand the behavioral markers, and make it easier for the observer to understand the certain behavior which is scored. We did not use an expert panel to enhance content validity, neither statistical method. This was already conducted by Myers in the development process of this instrument (Myers, 2016). Through the translation we tried to enhance the content as far as possible.

Assessment of construct validity

According to construct validity there was used a Likert scale to evaluate each item. This was easy to interpret, and is a common and validated scale for measurement in quantitative designs. Our scale was a 1-5 point scale ranging from poor performance – marginal -acceptable - good- excellent performance. These were found suitable for these behavioral markers to describe different levels of performance.

Assessment of external validity

External validity is concerned about the representativeness of the study. A crucial aspect of external validity is replication, - the generalizability of the results can be attained if findings are replicated in several sites (Polit & Beck, 2012).

In our study we partly replicated a study from New Zealand in the same environment and participants.

Even if the participants were mainly the same (physicians working in HEMS) and the assessment tools were the same (AeroNOTS) some differences in our methodical choice were made: The two observers were outsiders and did not calibrate the rating instrument in advance, different choices of statistical methods, the observation was conducted outdoor (Myers studied videos from a hospital setting) and translation may have affected the results.

The fact that we studied Norwegian physicians towards physicians from New Zealand may be another disturbing factor related to external validity.

Even if our results do not seem to be as positive as Myers`s study, this study may contribute to explore the rating instrument AeroNOTS to enhance external validity by trying out different ways of using it. Still we need more research on this to enhance replicability.

6.4 NTS Rating Instruments in HEMS

Is this assessment tool relevant for HEMS physicians?

The reason for developing rating instrument for NTS is described and evaluated in several studies (Gorden et.al., 2012; Garden et.al., 2013; Jepsenet.al.,2012; Rehim et.al.,2016). The conclusions of these studies are that they may be an important contribution to assess health providers in different areas to enhance quality and patient safety.

HEMS are an isolated and high-stake environment to provide health care, compared to other healthcare institutions (hospitals, nursery homes, GP offices). This means that the standards for medical treatment and care must be high. In Norway there are claims of education and experience for the HEMS physicians (www. stortingsmelding 43, 2012),

but there are no claims of how to develop necessary traits to lead a group of high-performing experts in rural areas to save lives.

The Foundation of Norwegian Air Ambulance has, in cooperation with Norwegian Air Ambulance, offered their personnel annual courses to enhance this important part of their work.

The training contains technical procedures as well as NTS training to strengthen the team. Their “crew-concept” is a part of this, and contains vital elements of CRM. These principles are implemented in this annual training course which last for one week.

In order to document and evaluate these and other courses, the rating instrument AeroNOTS should be considered, even if there still seems to be more work to be done on how to use it.

7.0 Conclusion

Our study confirms other studies about validation on rating instruments for NTS: It is difficult to establish statistical proofs on validity (Jepsen et.al, 2013; Rehim et.al,2016). However, this study do have some interesting findings. AeroNOTS may be able to distinguish acceptable performance from less good performance as demonstrated in the Bland Altman and Cohens Kappa analyses. Most participants had scores above 3 which is acceptable (Figure 1 + 2). This instrument may be used to discover major gaps in performance that should be followed up by the management. AeroNOTS could be useful for the management in HEMS to explore certain areas for internal training. This training may be more adapted to personal needs and challenges. There will be variations in experiences and training among the employees in HEMS. AeroNOTS may contribute to already existing training at the aircraft bases. AeroNOTS may be useful for the management and for the employee as they can work toward a common goal.

This rating instrument may be able to enhance communication skills to a certain point according to Cohens Kappa (Figure 2). This is probably the most valuable finding in this study because communication is a challenge and also extremely important in HEMS. The fact that the rating instrument enhances this part is crucial.

We will highlight the use of the instrument to be taken into consideration before use. As other studies demonstrate, validity is hard to prove (Jepsen et.al, 2015, Rehim et.al, 2016).

These studies also find that the assessor's role and competence is vital for valid results. More efforts should be put into this area. It is vital that the assessors know the rating instrument and how it should be used. According to our study, the rating instrument itself is not self-declaring, and the rater's background should also be considered. Retrospective it is clear that these factors may have biased our study.

We hope that use of AeroNOTS further on may contribute to a closer follow-up of the physicians in HEMS to make their treatment and collaboration even more safe and efficient for the patient.

In order to assess how experience may affect the physicians NTS, this study cannot conclude. Most of our participants must be considered experienced, and most of them had scores above 3, which is acceptable. This may be proof for that experience as a HEMS physician is connected with improved NTS.

Due to small sample size and possible biases, more evaluation has to be done, but in our experience this rating instrument may be a good supplement to existing training.

As Myers suggests, it would also be interesting to apply AeroNOTS for other parts of the HEMS crew, for example HEMS paramedics and pilots in addition to the physicians. Teamwork is complex and challenging, and dependent on the rest of the participants in the team. This may be a reason for measuring all the crew members to straighten their NTS in order to give safe treatment and care.

PART 2

PAPER

*Measuring non-technical skills in a Norwegian Air Ambulance
medical scenario setting using the customized rating instrument
AeroNOTS*

Abstract

The provision of life aid support to critically ill and injured patients in helicopter emergency medical services (HEMS) is a complex process strongly dependent on the team's non-technical skills to provide sufficient health care. These skills include task management, situation awareness, decisionmaking and teamwork. However there has been a lack of frameworks for measuring such non-technical skills for air ambulance personnel. In 2015 Julia Myers performed a customisation of the former validated observation tool Anaesthetists Non Technical Skills (ANTS) for use in HEMS called AeroNOTS (aeromedical non- technical skills). In the present study we apply this tool in a simulated clinical setting in Norway, to explore whether the tool is applicable in our environment, assess the tool's validity, and measure whether NTS improve according to physicians' experience.

Background

Non- technical skills can be defined as “the cognitive, social and personal resource skills that complement technical skills and contribute to safe and efficient task performance” (Flin, 2008). Safe and effective delivery of care relies heavily on teamwork (Westli et.al.,2010, Mills et.al.,2008, Yule et.al., 2015). Teamwork can be defined as a “set of interrelated behaviors, actions, cognitions and attitudes that facilitate the required task work that must be completed” (Flin R, et.al., 2008). Since the 1970`s there has been an increased focus on NTS in teams in order to avoid harm to patients and optimize patient safety (Flin & Maran 2004, Gordon 2012). There has been use of medical simulation to enhance learning of technical and NTS for clinicians working in a high-stake environment to optimize patient safety. In particular, poor interdisciplinary communication seems to be a significant factor in adverse events in air ambulance services (Geogiou,A.& Lockett DJ., 2010).

Simulation as a method to improve healthcare professional's technical and NTS is common in high-stake environments such as pre-hospital care (Flynn & Sandaker 2014, Abrahamsen et.al. 2014, Myers et.al., 2014). Simulation training can be conducted as an efficient way to learn and develop such skills according to Fletcher et.al.(2003). The study analyzed videos and questionnaires on anesthetists using the behavioral marker system

anesthesiologists non-technical skills (ANTS) They described that this assessment tool can be rated with acceptable levels of agreement and internal consistency.

O’Leary’s study (2015) points out that the assessment tools like Objective Structured Clinical Observation (OSCE) can be useful, but reliability depends on to which extent the results will be reproduced by different raters (interrater reliability) Jepsen, Østergaard and Dieckman (2015) identified 23 different assessment tools for NTS in different areas of health care. In their review they assess the tools purpose, development of instrument, scoring scale and validation procedure. Finally they assess reliability and validity. They also discuss the aim of using assessment tools: that is to provide a feedback to the participant and increase clinical standards over time. This means that the participants need to repeat the training. It also means that different assessors have to be involved to rule out biases and analyzes the normal variations among those measured and measuring. The tool also needs to be implemented in a context, and the facilitator has to be considerably skilled to provide constructive feedback. The tools themselves cannot be put into any context, but have to be customized for their use. Jepsen et.al (2015) addresses the lack of a gold standard for these rating instruments.

However, none of these assessment tools referred by Jepsen et.al are customized for air ambulance personnel. The authors conclude that there is a need to focus on validity of assessment tools and training of raters in using the tools in order to increase the awareness of the importance of human factors and improve patient safety.

Some of the assessment tools investigated by Jepsen et.al., are created to aid facilitators in their assessment of simulation participants (Jepsen et.al., 2015, Fletcher et.al., 2003). The assessment tools anesthetists non-technical skills (ANTS) and nurse-anesthetists non-technical skills (NANTS) include several categories of NTS like situation awareness, decision making, teamwork and leadership including a scoring system (Fletcher et.al., 2003, Flin and Maran, 2004, Flynn and Sandaker, 2014, Ballangrud, 2014).

Myers et al (2015) have adopted the main categories from ANTS and customized the NTS framework for air ambulance personnel with a particular emphasize on assessing communication in the noisy air ambulance environment and tested it with 16 physicians in an observational study in New Zealand.

“Well-designed training for air- ambulance clinicians should aim to prepare them for the recognized risks to patient care during all phases of transfer, and an assessment

framework based on non-technical skills would clearly have a high degree of relevance for this purpose.” (Myers et.al., 2016).

The primary aim of this study is to apply Myers assessment tool AeroNOTS in a Norwegian Air Ambulance simulation based training to explore its validity as an assessment tool for NTS. Further, we also want to explore whether AeroNOTS is an applicable instrument for the observer, and whether years of experience as prehospital anesthetist is a significant factor for high performance in NTS.

Method.

Study design and participants.

This study has a non-experimental research design and may construct a picture of the participants NTS at a certain point in time (LoBiondo-Wood & Harper,2002). It is an observational study of 24 anesthetists during a complex outdoor simulated medical scenario. The participants were crew members (anesthetic physicians) from nine out of twelve helicopter bases in Norway. The bases are located in rural and city areas of Norway and they have different work tasks and responsibilities depending on their location and population. One helicopter base is situated in a populated area in the eastern part of the country. Four (17%) of the participants had their daily work at an Emergency Ambulance in Oslo and were not employed in the Air Ambulance Services. They are normally invited to the camp for training in advanced prehospital medicine, and were included in this study.

According to informal conversations with the participants prior to the observations, it was clear that most of them had been working in the HEMS, in addition to being on call at in-hospital intensive care units and operating theatres throughout their career. This experience could mean the potential to develop their NTS in relevant clinical settings.

Validation.

As the aim of the study is to explore validation of the assessment tool AeroNOTS, some important elements on this issue should be mentioned. Polit and Beck (2017) describes four main elements of validity:

1. Statistical conclusion validity means that the researcher must provide evidence as strong as possible.
2. Construct validity involves to which degree the intervention is a good representation in the underlying construct.
3. External validity is concerned about the generalizability of causal inference. Will the observed relationship hold over variations in persons, settings and time?
4. Content validity may be defined as “the extent to which an instrument`s content adequately capture the construct- that is, whether an instrument has an appropriate sample of items for the construct being measured” (Polit & Beck, 2012, p.310).
5. Face validity refers to whether an instrument looks like it is measuring the target construct according to Polit & Beck(2012, p.310)

Polit and Beck (2017, chapter 10) also describes threats to validity. How the participants are included in the study is a matter of choosing a suitable design, and ensure that the researcher get **the requested data collected**.

Threats to statistical conclusion validity may be errors in interpreting the data correctly according to the statistical tests that have been used. Type 1 or type 2 errors are the most common.

Threats to construct validity includes reactivity to study situation, known as the Hawthorne effect, novelty effects because of the researchers enthusiasm or skeptical attitude, compensatory effect where the participants compensate for not receiving beneficial treatment. Treatment diffusion or contamination is when the participants drop out and put themselves in the control group, or categorization of groups is not appropriate.

Threats to external validity may occur because of interaction between relationships and people for example that different gender or ethnical background may give different results. Interactions between causal effects and treatment variations are also a threat if the treatment is paired with other effects witch influences the outcome.

According to **content validity and face validity** there is possible errors according to how a instrument is being used and of whom, and the way the items are interpreted. Face validity is the weakest proof of validity and mostly used to cope with participants resistence.

Some of these issues concerning validity will be discussed on the background of this study's results.

Observation tool and translation.

The AeroNOTS assessment tool was translated into Norwegian. The translation from English to Norwegian was done by one of the assessors. This translation was verified and corrected by the other assessor and one other person with good English proficiency but no work experience in health care. After the first translation and corresponding correction rounds no further changes to the translation were made. The translated form was handed to a translator who retranslated from Norwegian back to English. The translator made comments on some grammatical errors regarding the use of verb tense in the explanation of different elements in the extended version of the tool. Main concepts and content were not changed. A short version of the Norwegian translation for simplicity in the field was made. It included an outline of the main areas and elements in the assessment tool (Table 5). The AeroNOTS tool consists of 15 items, all given a value on a five point Likert scaler from 1 to 5 resulting in a total score between 15 and 75 (Table 2).

Data collection.

The observation of the participants by two AeroNOTS assessors was performed every Wednesday between August 24th and September 28th 2016. The same scenario was performed every week but with different crew members. 24 observations were measured twice. In total 48 observations. Six physicians were excluded as they attended in three crews at two scenarios – this meant less workload on each participant and probably a better score. These six physicians were additional to the 24 as referred. A written permission was given from those who accepted to participate. All of the invited physicians accepted participation. The assessors' observation started when the participants arrived at the simulation site, and stopped at the end of the scenario before take-off. The rating was done during and as soon as possible after the scenario ended. In cases where HEMS- crew split up and did their treatment at different places, the two assessors split up and each followed a subgroup separately.

Simulation scenario

The simulation took place outdoor in a hillside by a small lake outside Oslo (Holmenkollen). This is the annual campsite for all employees in the Norwegian Air Ambulance (NLA) to facilitate their training in medical, rescue techniques and CRM competence, named Camp Torpomoen. The place for the camp and focus areas varies from year to year. The main learning goals for the participants in this year scenario were to perform adequate communication and follow the new guidelines for advanced heart and lung resuscitation (CPR) for children.

The crew members received their information on their base about three minute driving distance from the lake, and their time for planning of organization and actions. This was a search- and- rescue-assignment, and the information given at radio communication contains information about three kids aged 10-12 years old missing, probably drowned in the lake with no visible signs. The crew had to gather all possible information themselves about the position of the children. The children were hidden under the surface of the water (sink mannequins). One of the children's mother and a neighbor were there to support them with information. If they called for firemen support with divers, the AMK (113) central answered they were busy with another assignment and couldn't come until later. Police were not at the scene, and ambulance had a delay and could come in about 30-40 minutes, if requested.

The medical scenario was set up with two active, three piece crews on scene (pilot, HEMS paramedic and physician). The crew arrived at the scene by car, not by helicopter for safety and economic reasons. The HEMS paramedic had to go into the water in diving suits to search for the children, while the pilot and the physicians organized themselves in a way they determined most suitable.

The crew had to organize their resources and call for help while performing ongoing lifesaving treatment of the children. They also had to consider using the people on site in the most efficient way to succeed, effectively testing their skills within all domains of NTS.

Each scenario last from 30-40 minutes and there were three scenarios each day. The participant participated in the described scenario only once, but they had two other scenarios the same day. In total the assessors observed this particular scenario 12 times.

Each crew member was observed and independently scored by the two assessors; one nurse anesthetists with 20 years of in-hospital experience, and one para-medic nurse with 30 years of experience in prehospital care. The two did not calibrate or discuss their assessment scores of the content of each element before rating.

Statistical methods.

Continuous data are summarized as mean (SD) for symmetric data, and median (quartiles) for skewed data. Categorical data are summarized as numbers (%).

The total score for the AeroNOTS tool can be considered as a continuous variable, and agreement between the total score from the two independent assessments on each participant was calculated by a Bland Altman (BA) analysis, introduced by a seminal paper by Bland and Altman (Bland & Altman 1986, Altman 2006, Giavarina 2015). The method highlights that agreement is a stronger claim than merely high correlation, and an essential aspect of the method is the comparison between the *observed* variation in the data and the *clinically acceptable* variation. The latter is a clinical judgment that should be defined a priori based on clinical necessity (Giavarina, 2015). We defined this as a score above or under 3 which is the cut off for acceptable clinical performance.

The agreement between the two different measurements on the same subject, here the total score as judged by the two independent assessors, are visualized in a BA plot, where the difference between the two measurements is plotted against their mean. The corresponding limits of agreement (LoA) are the limits between which 95% of the observed differences lie, representing the actual variation in the data (Giavariana,2015). The LoA enable a comparison between the actual variations in the collected data with the clinical acceptable variation.

Using Bland- Altmans plot is preferred when searching for similarities, while correlation tests are preferred when looking for differences in data (Giavarian, 2015, Bland & Altman, 1983).

Inter-assessor agreement for each of the 15 individual items was calculated using Cohens Kappa (Pallant, 2013). Guidelines on interpreting the values of Cohens Kappa characterize values 0-0.20 as “slight agreement”, 0.21-0.40 as “fair agreement”. 0.41-0.60 as “moderate agreement”, 0.61 - 0.80 as “substantial agreement” and 0.81 - 1 as ”almost

perfect agreement” (Pallant,2012). Limitations with using k , can be that k tends to underestimate on the rare category, and for this reason it is considered as conservative measure of agreement (Pallant, 2013).

To explore whether experience was associated with improved NTS, we plotted the bivariate association using boxplots, and applied a Kruskal Wallis test. This test is preferable when data have one nominal variable (experience) and one ranked variable (Likerts scoring scale 1-5) to show the variance in the data (Mc Donald, 2009).

The null-hypothesis of the Kruskal Wallis test is that the mean rank of the groups are the same assuming that the shape of the distribution in each group is the same. It does not assume that data are normally distributed, which is an advantage in our study. As this test explains any differences among the groups that would make the mean ranks different, the Kruskal Wallis test could be preferred (Mc Donald, 2009).

The statistical analyses were performed using SPSS (IBM; Armonite NY version 21) for PC and R 3.11 (R development Core Team, 2007). P-values < 0.005 were considered statistically significant.

Results.

Demographic data demonstrates that there were 22 men and two women included in the study (Tab 1). The participants age varied from 36- 66 years old (mean = 46.16). Their experience as an anesthetist varied from 6 to more than 30 years including time of educating as anesthesiologist (mean= 13.57). Their training hours in NTS varied from 0-10 hours per month (mean = 3.25). These results indicate that most of the participants had long prehospital medical experience, and that they also had some training in NTS in addition to their calls.

The bivariate scatterplot of the total score from the two assessors seems to indicate quite large variation (Figure 1). Corresponding Bland Altman (BA) plot with LoA superimposed (Figure 1) confirms this. The LoA are from - 20.x to 19.y, which is considerably larger than the a priori defined acceptable clinical difference of ± 3 .

The BA analyze shows the variance between the two raters measured by their total scores (15-70) compared to a mean value. The variance was between 40 – 70 (Fig 1A).

Corresponding Cohens Kappa values confirms this, with agreement ranging from slight to fair as k varied from 0.022- 0.314 (fig 2). This result is not significant for all of the 15 elements with a significant k -level > 0.5 . Five elements demonstrates fair agreement: “collecting information”, “exchange information”, “using authority and assertiveness”, ”coordinating the team” and ”risk management”.

Plotting years of clinical experience versus the score for individual items in the assessment tool (Fig 3) demonstrates a tendency towards higher scores for longer experience. Most of scores with a high rating of 4 or more stem from individuals with more than 10 years of experience. The Kruskal Wallis test could however not confirm this tendency as statistically significant as only one element (“planning and preparing”) had a significant level demonstrating $p < 0.005$.

The Kruskal Wallis test indicates how experience influenced their NTS according to our observation form. There were mainly scores above 3, which means “acceptable performance” and above. Especially the elements “maintaining standards”, ”exchange information”, ”assessing capabilities” and “supporting others” had a high frequency and the highest score (4-5). ”Maintaining standards” includes their technical skills and procedures as well as the way these are performed. This group of physicians had long professional experience and even hours of training every month as they are expected to be high-level performers. Our findings support this.

Ethical considerations.

All the participants were informed about the study 7-10 days prior to their arrival at the camp by an information letter (Table 6). The day the observation took place; there was a 10-minutes information meeting about the aim of the study. Those who volunteered had to sign a formal consent statement, and provide demographic and professional information (ex. age, experience, amount of training hour per month). Names were separated from other information and kept on a database at the university’s computer.

3 months before starting up the study we applied for permission at the Norwegian Statistic Data Services (NSD) as required. Notification for the study was given 12th of August 2016, and the notification number is 49139.(attachment 2)

Discussion.

The aim of our study is to apply Myers assessment tool AeroNOTS to explore its validity for HEMS physicians in Norway and if physicians prehospital experience affects their NTS. The assessment tools applicability will also be considered.

Most of the study participants had more than 10 years of experience including specializing as an anesthesiologist, and 75% had an age of 49 years old.. There seems to be a connection between the high scores on the AeroNOTS element and the physicians age and number of training hours. Most of the physicians were rated by both assessors as high level performers.

Bjørndal & Hofoss (2015) suggest that it is reasonable to think that clinical experience of 10 - 15 years affects the clinician NTS

To explore limits of agreement for AeroNOTS, a Bland-Altman plot to compare the two assessors scoring of the simulation participants were conducted. The Bland-Altman plot demonstrated low agreement between the two assessors. In Myers study (2015) there were performed a Spearmans rank correlation test witch is most suitable looking for differences. In Myers study most of the statistical tests demonstrated significant levels due to distinguish higher and lower levels of non-technical performance. Different choice of statistical method will clearly affect the results of the two studies.

We found that the total scores varied from 40 - 70 points (15-70). This means that the scores are spread and that statistically is this tool (AeroNOTS) not very precise and accurate in measuring NTS of the HEMS physicians the way it was used in this study. It also demonstrates that the two assessor's judgements of the scores were different. The reason for this may be connected to the observers background (not HEMS personnel), not calibrating the instrument before use and that some observations were conducted inappropriate according to observing only one of the crews each .This may have affected our scores. All these factors may have been biased validity.

If these differences have clinical relevance, is not obvious. Giavarina (2015) claims that limits of agreement for clinical relevance have to be set a priory the study. Our limit was set at 3, which means that all scores between 3 and 5 are acceptable performance without any clinical disadvantages. Referred to the Bland Altman analyses there is statistically low agreement among the two assessors, but clinically

moderate agreement which is acceptable according to NTS and clinical performance. According to Jepsen et.al. (2015), validity varied in the 23 studies on assessment tools for emergency medicine. This means that the criteria for internal, external and content validity are challenged in most of the studies included. Their conclusion is that the raters should be better skilled and trained, and that assessment tools should be retested and refined to be valid. Our study confirm these challenges and point out some of them as described.

The comparison of the two assessors by Cohens Kappa indicates low agreement on 10 of the 15 elements in the observation form. There was fair agreement for the five elements: "coordinating activities in the team", "exchange information", "using authority and assertiveness", "gathering information" and "balancing risk and selecting options". Both assessors had the same perception and assessed these five elements equally ($k = 0.226 - 0.314$). All of these five elements is strongly connected to communication. This may indicate that these elements are sensitive for communication skills even if they are not labelled in the form as such. This is an interesting finding because the differences between this assessment tool and others used for NTS are connected to communication.

In the category "re-evaluation" the ratings between the assessors conformed. An explanation for this could be that this element exists throughout the scenario at several points: in the beginning of the scenario, after finding the first missing child, after finding the second child and before take-off. This means that this particular element was assessed several times and may have higher scores. The re-evaluating was mainly a task for the physicians as they were the leader of the team. There was an overall high score on this element and due to the Cohens Kappa test this element had "slight agreement" ($k = 0.055$).

The physicians' ability to organize and keep track of the tasks in the complex and challenging simulation scenario were overall good. As the results demonstrate, the participants were most often scored above 3, which is acceptable. Studies underpin that simulation based training on NTS tends to improve these skills, and especially team performance (Siu et.al., 2016; Jepsen et.al., 2016; Eppich et.al.9 2015). The participants demonstrated these skills through their communication and task performance during the simulation.

These tests indicate that the observation tool is statistically of limited value for this group, or that the observers had different ideas on how to use the tool. However, it could be relevant to ask how much the result differs, and in which areas they agree. The elements that are highlighted with a high score (mostly above 4) are:

"Maintaining standards", "exchanging information", "assessing capabilities" and "supporting others". These areas reflect requirement for complex and coordinated actions from the simulation participants, necessary to perform at the highest level. It also demonstrates good team performance, which is crucial in such a complex case.

Most of the participants had scores above 3, could be interpreted as an overall acceptable performance of NTS, which is underpinned by their long experience (mean: 13.75 years), age (mean: 49) and efforts to maintain their high standards through the training program (mean: 3.25 h/month). Evidence support how training in NTS improves technical and NTS skills in the air ambulance services (Abrahamsen et. al., 2014, Lamb, 2007).

Observation and self-reporting assessment after simulation may cause different results. Inexperienced physicians tend to evaluate themselves with higher scores than more experienced physicians (Myers et al., 2015). A review on assessment tools demonstrates that self-assessment is the most common way to measure NTS (Cooper et.al., 2014; Jepsen et.al.,2016) and Myers study question if this is the most reliable way of assessment. This study did not use self-reporting assessment. We wanted to investigate how the observation tool worked out without coordinating the use of the tool before the study started. Even if both assessors found the form easy to interpret and understand, the results of our statistical investigation demonstrated that we used the form slightly different as Cohens Kappa (Fig.2) and Bland Altman analyzes (Fig 1) demonstrates. However, we found it useful in order to divide acceptable performance from marginal and poor performance, likewise Myers study (2016).

Our study had only physicians participating. The assessment tool should, according to Myers (2016), also be applicable for other clinical groups involved in critical care transfer. In the first phase we considered including HEMS paramedics into our study, but then we would have needed more observers. Some of the pilots also showed their interest for the assessment tool. Our reason for excluding them was primarily due to

lack observers. We recommend more studies to investigate AeroNOTS for pilots and HEMS personnel to measure NTS for these groups. In order to investigate the assessment tool ability to measure how experience influenced on NTS, we found that there is a slight indication that long experience as a HEMS physician, may increase NTS. The assessor's background and their lack of pretesting / configuration of the tool could clearly affect the result. This was an expected finding which is underpinned by evidence from other studies.

Assessing AeroNOTS according to validity, we found some confounding factors as mentioned (background, no calibration and observational laps). These will clearly challenge validity. Content validity seems to be undertaken through our translation to Norwegian and all items seem relevant. Face validity was used to assess clinical acceptable values conducting the Bland Altman analysis.

Limitations.

In order to explore how experience influenced NTS, more information on this should have been provided - we asked for age of year, and years of experience. Compared to Myers study (2016) she used numbers of assignments per year with a cut-off 45, and dichotomized "experienced" and "inexperienced" at this point. On the other hand we had more variation when analyzing our data, and according to evidence as referred (R Development Core Team, 2007) the Kruskal Wallis test is preferred in our data. This made comparison with Myers study difficult though. Still we could have needed more precise information on kind of experience; we were primarily interested in prehospital experience. Collecting data on experience, we were inaccurate in our questionnaire about where they had been working. We tried to compensate for this through specifying that experience included work as an anesthetist (including specializing). We recommend explicit information about the physicians former work experience, as we found it complicated not knowing how long they had been working in HEMS because we have not been asking for it. Informal discussions before the observation with the participants were useful to interpret the information we got.

A significant challenge we met, was that some of the scenarios turned out to be located at two different places at the waterside, because the child was brought there when rescued from the water. This meant that the two assessors had split up to follow one crew each

because they were not able to observe both crews simultaneously, so our assessment of the opposite crew members was inadequate. This was a challenge we met on site, and we had to make a choice at the time how to solve this. Even if we marked each observation form how fulfilling the observation were, we were not able to highlight this in our dataset. This is a weakness in our study.

The two assessors did not calibrate the view of how to use the assessment tool because we wanted to see if we used it differently or not. This may be a vital limitation in our study, and demonstrates great variations in the way we scored the participants (Figure1) This choice was taken consciously as we considered it to serve as an internal validation criteria in addition to other.

Other factors that could explain the differences in using this assessment tool, are the two assessor`s background. Both were nurses, one was an experienced pre-hospital nurse; the other a nurse anesthetist with mostly in-hospital experience. The way we both looked upon the different elements and the participants work, would be seen through our previous experiences and interpreted by these (Fangen, 2011). Most likely the observer with great prehospital experience would recognize and assess the participant`s actions more accurate. The assessors prepared themselves by reading and talking socially to the participant the day before observation, which also was a gate opener to the field (Fangen, 2011). This may have an impact on our scores and the final result.

Because our study contains a small sample size conclusions cannot be drawn (Bjørndal & Hofoss, 2015, Ringdal, 2001). Further examinations have to be done.

Conclusion.

In order to use and validate Myers assessment tool forNTS, AeroNOTS, we found it useful to identify major gaps in performance, as Myers also points out. AeroNOTS can be able to select poor from acceptable performance in a simulated outdoor clinical setting similar to work for Air Ambulance physicians. Statistically we did not find evidence for validity. There is indications of that this could be a useful tool, but further evaluation on this rating system is merited.

Even if there was low inter-rater agreement according to Cohens Kappa and the Bland Altman analysis among the two assessors, this also could be explained by other reasons

than the assessment tool itself. The results could also indicate that NTS may improve according to experience and some added training .Our sample is too small to conclude on this.

We recommend further evaluation of this framework, and would consider it interesting to include other relevant clinical groups like HEMS paramedics and pilots in the data.

This study finds AeroNOTS assessment tool to be useful to define good from poor performance in NTS. However, the statistical test cannot confirm the tools validity according to Bland Altmans analysis and Cohens Kappa. The clinically impact of the results is positive, as they are measured as "fair agreement". We recommend the assessors to coordinate their use of the AeroNOTS form before using it, and that the assessors may be HEMS physicians themselves to make sure their understanding of the observed situation are at the same level. This may have an influence on the ratings.

Keywords: Non-technical skills, Air ambulance, patient transport, clinical training, simulation, AeroNOTS.

REFERENCES

Abrahamsen, H., Sollid, S., Ølund, L., Røislien J., Bondevik Tchudi, G. (2014). Simulation-based training and assessment of non-technical skills in the Norwegian helicopter Emergency Medical services: a cross-sectional survey.

Emergency Medical Journal 0:1-7. – doi: 10.1136

Ahmed, K., Miscovic. D., Darzi, A., Athanasiou, T., Hanna, G. (2011). Observational tool for assessment of procedural skills: A systematic review.

The American Journal of Surgery, vol. 202, issue 4, p 469-480.

Altman, D. G. (2006). The cost of dichotomizing continuous variables. *British Medical Journal*, 332(7549),1060.

Ballangrud R., Hall-Lord, M.L., Persenius, M., Hedelin, B. (2014). Intensive care nurses' perceptions of simulation-based team training for building patient safety in intensive care: A descriptive qualitative study.

Intensive and Critical Care Nursing Volume 30, pages 179–187.

Observational tool for assessment of procedural skills: A systematic review.

The American Journal of Surgery, vol. 202, issue 4, Oct 2011, p. 469-480

Bevan A, Joy R, Keeley S, Brown P. (2015). Learning to nurse: combining simulation with key theory.

British Journal of Nursing, Vol. 24, No. 15

Bland, M. & Altman, D.G. (1986): Statistic methods for assessing agreement between two methods of clinical measurement.

Lancet, i: p. 307-310

Bloom, B.S. (1984) *Taxonomy of Educational Objectives*.

Allyn and Bacon, Boston, (Copyright 1984 by Pearson Education.)

Bredmose, P. Habiq K., Davies, G. & Lockey, D. (2010). Scenario based outdoor simulation in prehospital trauma care using simple mannequin model.

Scandinavian Trauma and Resuscitation Emergency, 18:13

Brindley, P.G., Suen G.I., Drummond, J. (2007) Medical Simulation: How to build a successful and long-lasting program.

Canadian Journal of Respiratory Therapy, winter.

Bjørndal, A. & Hofoss, D. (2015) *Statistikk for helse- og sosialfagene*.

Oslo: Gyldendal Norsk Forlag.

. . (1986) p.

Bosseau Murray, W. & Foster, P. (2000). Crisis resource management among strangers: Principles of organizing a multidisciplinary group for crisis resource management.

Journal of clinical anesthesia, 12, p. 633 - 838

Carne, B., Kennedy, M., & Gray, T. (2012). Review article: Crisis resource management in emergency medicine.

Emergency medicine Australasia, 24, p. 7-13

Cooper, S., Porter, J. & Peach, L. (2014). Measuring situation awareness in emergency settings: a systematic review of tools and outcomes.

Open access Emergency medicine, 6, p.1-7

Cooper, S., Cant, R., Conell, C., Sims, L., Porter, J., Symmons, M., Nestel, D. & Liaw, S.Y. (2016). Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams.

Resuscitation, 101, 2016, p. 97-101.

Eppic, W., Nannicelli, A., Seivert, N., Sohn, M., Rozenfeldt, R., Woods, D. & Holl, J.(2015). *Journal of continuing education in the health professions, 35(2), p. 83-90*

Derbyshire J., Machin A., Crozier, S. (2015). Facilitating classroom based inter professional learning: A grounded theory study of university educators perceptions of their role adequacy as facilitators.

Nurse Education today, 35, p. 50- 56

Fangen, K. & Sællerberg, A.M. (2011) *Mange ulike metoder.*

Oslo: Gyldendal Akademisk.

Fletcher, G., Flin, R., McGeorge, P., Glavin, R. Maran, N. & Patey R(2003): Anesthetists Non-technical-skills (ANTS) evaluation of a behavioral marker system.

British Journal of Anesthetists, 5, p. 580-588

Lamb, D. (2007): Could simulated emergency procedures practiced in a static environment improve the clinical performance of a Critical Care Air Support team (CCAST): A literature review.

Intensive and Critical care Nursing, Vol 23, Issue 1, page 33-42

Flin, R., O'Connor, P., Mearns, K. (2000) Crew resource management: Improving team working high reliability industries.

Team Perform management, 8, p. 68-78

Flin, R., Patey, R. (2011). Non-technical skills for anesthetists: Developing and applying ANTS.

Practicion & Resuscitation Clinical Anesthesiology, 25, p. 215-27

Flin, R.H., O'Connor, P., Crichton, M. (2008). *Safety at the sharp end: A guide to Non-technical skills.*

Farnham, UK: Ashgate

Flin, R.H. & Maran, H. (2004). Identifying and training non-technical skills for teams in acute medicine.

Quality & Safety of Health Care, 13. Supplement 1.

Flin R, O'Connor & P, Mearns K, (2008). Crew resource management: Improving team working high reliability industries. *Team Performance management.* 12; p.68-78

Flynn, F. og Sandaker, K. (2009) Å sikte mot eksellense - en simuleringsbasert studie om utvikling av ikke-tekniske ferdigheter hos anestesisykepleierstudenter.

Masteroppgave, Høgskolen i Gjøvik, Gjøvik.

....(1995)*Psychological p.*

D. . s.K. .&. (2001). , p.

Garden A.L., LeFevre D.M., Waddington H.L. & Wellers J.. (2015) Debriefing after simulation based non-technical skills training in health care: a systematic review of effective practice.

Anesthesia Intensive care, 43:3.

Geogiou, A. & Lockey, D.J. (2010). The performance and assessment of hospital trauma teams.

Scandinavian Journal of Trauma and Resuscitation and Emergency Medicine, 18, 66

Giavarina, D. (2015). Understanding Bland Altman analysis.*Biochemical Medicine*, 25(2), p. 141- 151

Glesne, C. (2006) *Becoming Qualitative Researchers. An introduction*.

Boston: Pearson Allyn and Bacon.

Gorden,M., Darbyshire, D. & B. Non-technical skills training to enhance patient safety: a systematic review.

Medical education, 46, p. 1042-1054

Hammersley, M. & Atkinson, P. (2004). *Feltmetodikk. Grunnlaget for feltarbeid og feltforskning*.

Oslo: Gyldendal.

Helmreich, P.L. (1997): Managing human error in aviation. *Scientific America*, 276(5), p. 62-70.

Helmreich, P.L. & Merrit, A. C. (1998): *Culture in work in aviation and medicine*.
Aldershot: Ashgate.

Husebø, S.(2013). *Training for team-based resuscitation in nursing education*.
PhD Thesis, Universitetet i Stavanger, Stavanger.

Jepsen,R., Ødegaard,D., Diekmann,P. (2013). Development of instruments for
assessment of individual`and teams` non-technical skills in health care: a critical review.
Cognitive and technical work. 17 p.63-77. DOI 10/1007/s10111-014-0306y

Johannesen, A. (2009). *Introduksjon til SPSS*.
Oslo, Abstrakt forlag, 4.utgave

King, A. E., Conrad, M, & Ahmed, R. (2013). Improving collaboration among medical,
nursing and respiratory therapy students through inter-professional simulation.
Journal of Inter-professional Care, 27, p. 269-271

Kolb, D.A. (1984). *Experiential Learning: Experience as the Source of Learning and
Development*.
Englewood Cliffs, NJ: Prentice-Hall.

Lamb, D. (2007). Could simulated emergency procedures practiced in a static
environment improve the clinical performance of a Critical Care Air Support team
(CCAST): A literature review.
Intensive and Critical care Nursing, Vol 23, p. 33-42

Løkken, G. (2012): *Levd observasjon. En vitenskapsteoretisk forklaring til observasjon som forskningsmetode.*

Oslo: Cappelen Damm

Martinsen J. (2015). Observation and rating HEMS Crew in Non-technical-Skills: CRM Medical Simulation in Norwegian Air Ambulance.

Scandinavian Journal of Trauma Resuscitation and Emergency medicine, 23, Supplement 2

Matics, D. (2015). Implementing Simulation in Air Medical Training: Integration of Adult Learning Theory.

Air Medical Journal, Volum 34, p. 261- 262

McDonald, J. (2009) *Handbook of biological statistics.*

Baltimore, Maryland, USA: Sparky House Publishing, 2.edition

Myers, J., Powell, D.M.C., Psirides, A., Hathaway, P., Aldington, S., & Haney, M.(2016).

Non-technical skills evaluation in the critical care air ambulance environment: Introduction of an adapted rating instrument- an observational study.

Scandinavian Journal of trauma, Resuscitation and Emergency Medicine 24:24.

Pallant, J. (2013). *SPSS, 5th edition- Survival manual.*

Berkshire, England: Open University Press, McGraw-Hill Education

. L. (2011),,

Polit, D. & Beck, C. T. (2017). *Nursing research- generating and assessing evidence for nursing practice.*

Wolters Kluwer, 10th edition.

Qvindesland, S. A., Bjørshol, C., Aase, I., Rossavik, B & Kluge, M.(2015). Å skape fremtidens tverrprofesjonelle helseteam. www.tidsskriftet.no/ URL. November.

R Development Core Team (2007). R: A language and environment for statistical computing.

Vienna, Austria; R Foundation for Statistical Computing, <http://www.R-project.org/> URL.

Rehim, S., DeMoor, S., Olmsted, R., Dent, D., & Parker- Raley, J. (2016). Tools for assessment of communication skills of hospital action teams: A systematic review.

Ringdal, K. (2001). *Enhet og mangfold.*

Bergen: Fagbokforlaget Vigmostad og Bjørke.

Robertson, J. & Bandali K. (2008). Bridging the gap: Enhancing inter-professional education using simulation.

Journal of inter-professional care, 22:5, p.499-508

Ross,A., Anderson, j., Kodate, N., Thomas, L., Thomsen, K., Thomas, B., Key, S., Jensen, H., Schiff,R. & Jaye, P.(2012)Simulation training for improving the quality of care for older people: an independent evaluation of an innovative programme for interprofessional education.

British Medical Journal Quality and Safety, 0:1-9.DOI:10.1136/bmjqs-2012-0000954/URL

Rudolph, J.,Simon, R.,Dufresne, R. & Raemer, R. (2006). There`s no such thing as “non-judgmental” debriefing: A theory and method for debriefing with good judgement. *Simulation in health care, Vol 1. No. 1*

Shadish. W.R., Cook, T.D., & Campbell, D.T (2002) *Experimental and Quasi-experimental designs for generalized causal inference.*

Boston: Hughton Mifflin.

Siu, J., Maran, N. & Paterson- Brown, S. (2016). Observation of behavioral markers of non- technical skills in operating room and their relationship to intra-operative incidents.

The Surgeon, volume 14, issue 3, p. 119-128

.. & . (2015). *1 p ,*

DOI:10. 1080/1463922X.2015.1106619

Thagaard, T. (2013). *Systematikk og innlevelse- en innføring i kvalitativ metode.*

Oslo: Fagbokforlaget, 4.utgave.

Tin, M.B. (2000). *Etterord (s.85-126).* Merelau- Ponty, M.: *Øyet og ånden*

Oslo: Pax.

Valentine M., Nembhard, I. & Edmondson, A. (2015). Measuring teamwork in healthcare settings. A Review of survey instruments.

Medical care, Vol. 53, no 4.

Wax, R. (2006). *The ABCs of Air Medical Resource Management. In: Fredriksen K editor. Principles and direction in Air medical Transport*

Salt Lake City: Air medical Physician Association.

orld Health Organization. (2012).Doc 1.5.

WHO. /URL

Woolfolk , A. (2004). *Pedagogisk psykologi.*

Trondheim: Tapir Akademisk Forlag.

Yule, S., Parker, S. H., Wilkinson, M.S., Mc Kinley, A., Mc Donald J., Neill, A.& McAdam T(2015). Coaching non-technical skills improves surgical rezidents performance in a simulated operating room.

Journal of Surgical education, vol. 72, issue 6.

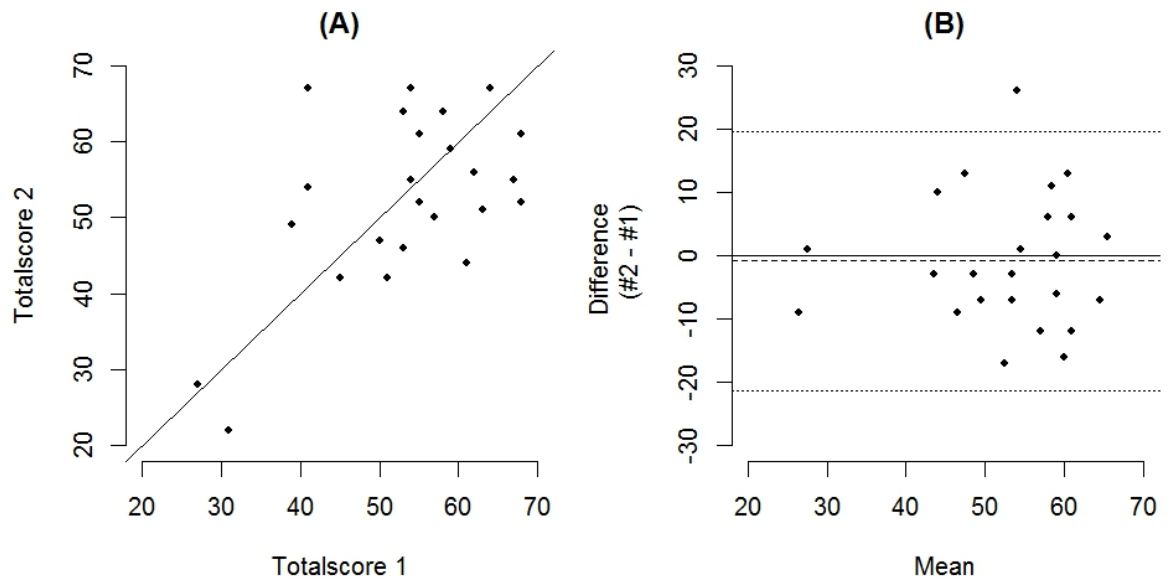
Helse- og omsorgsdepartementet (2000). Stortingsmelding no. 43.

<https://www.regjeringen.no/no/dokumenter/stmeld-nr-43-1999-2000-/id193493/sec2/>

URL.

APPENDIX

Figure 1: Bivariate scatterplot of the total score from the two assessors (A) and corresponding Bland Altman plot (B) and Limits of Agreement (dotted lines) super-positioned.



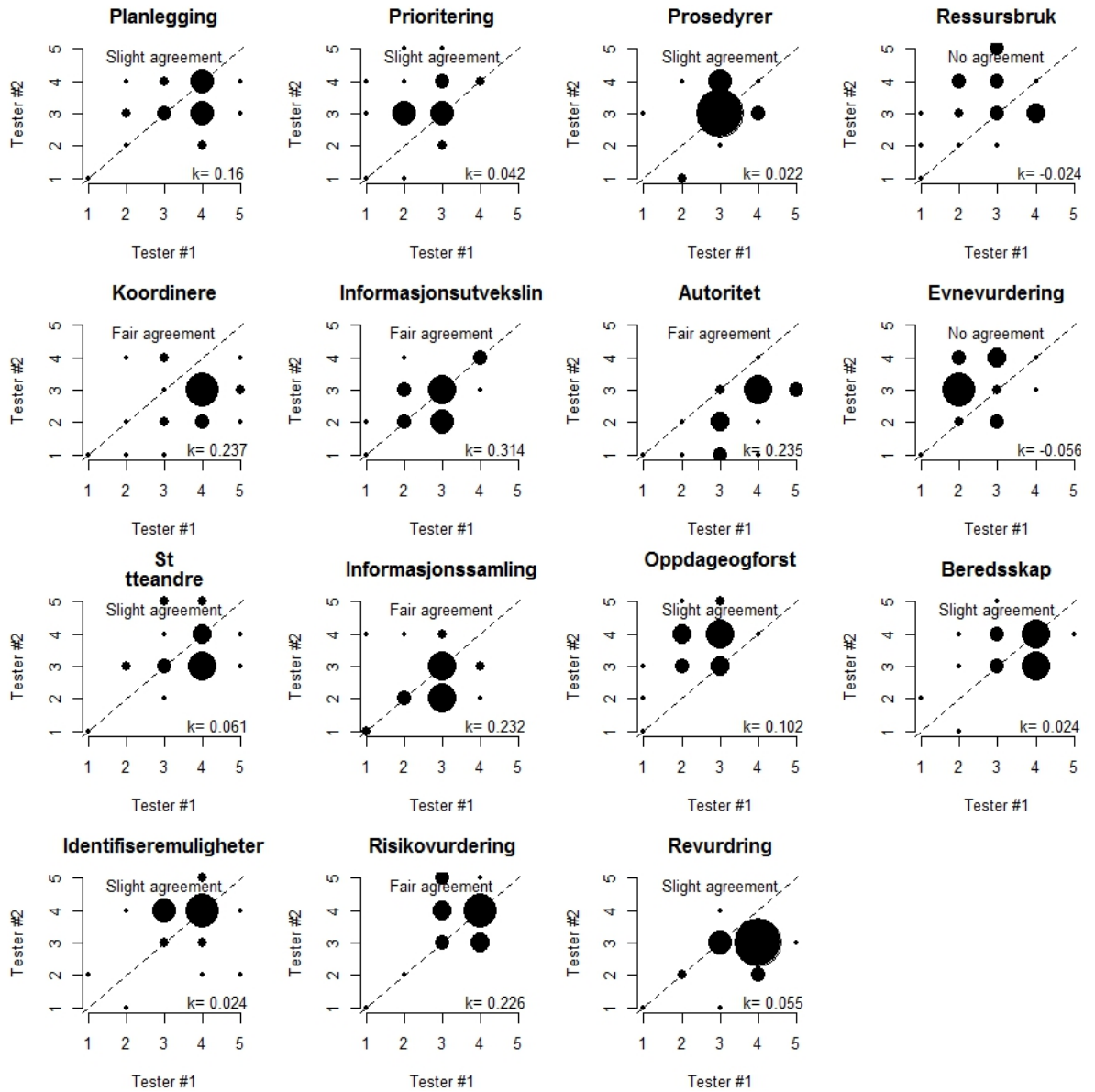


Figure 2: Agreement between the two assessors for the categories of each of the 15 sub-categories of Myers assessment tool. Size of circle represents number of times the two assessors scores conforms. For full agreement all circles should lie along the line of agreement (dashed line). Values and interpretation of Cohen's Kappa superimposed

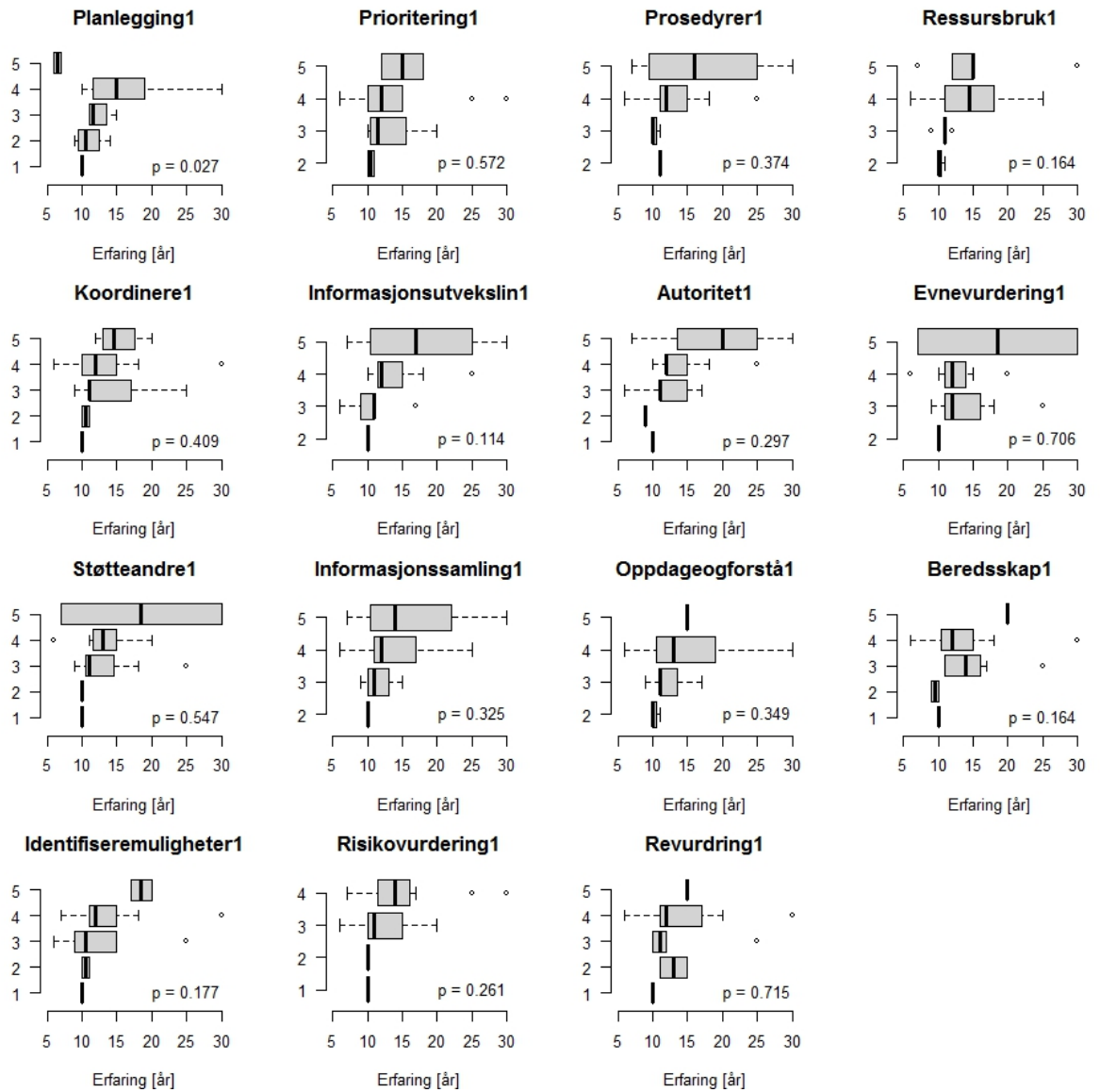


Figure 3: Boxplot of years of experience vs all 15 elements in Aeronots.

P-value from Kruskal Wallis test superimposed.

Table 1: Demographics of participants.

	N = 24	Missing	Mean	Percentiles - 75
Age (years)			46,16 (36-66)	48,75
Experience*		1	13,57 (6-30)	15
Training hours in NTS per month			3,25 (0-10)	5
Females	2			
Men	22			

*Mean years of work including time of specializing as an anesthesiologist.

Table 2: Original AeroNOTS observation form as presented in Myers et al (2015).

Task management The skills of management of resources and organization of tasks to achieve goals, be they individual case plans or longer term scheduling issues	Planning and preparing	Developing in advance primary and contingency strategies for managing tasks, reviewing these and updating them if required to ensure goals will be met, making necessary arrangements to ensure plans can be achieved
	Prioritizing	Scheduling tasks, activities, issues, information channels, etc. – according to importance (eg. due to time, seriousness, plans); being able to identify key issues and allocate attention to them accordingly, and avoiding being distracted by less important or irrelevant matters
	Maintaining standards	Supporting safety and quality by adhering to accepted principles of patient transport, following where possible, codes of good practice, treatment protocols and guidelines, and mental checklists
	Identifying and utilizing resources	Establishing the necessary, and available, requirements for tasks completion (eg. People, expertise, equipment, time) and using them to accomplish goals with minimum disruption, stress work overload or underload (mental and physical) on individuals and the whole team
Team working The skills of working with others in a team context, in any role, to ensure effective joint task completion and team satisfaction; the focus is particularly on the team rather than the task	Coordinating activities With the team	Working together with others to carry out tasks, for both physical and cognitive activities; understanding the roles and responsibilities of different team members; and ensuring that a collaborative approach is employed
	Exchanging information	Giving and receiving the knowledge and data necessary for team coordination and task completion
	Using authority & Assertiveness	Leading the team and/or the task (as required), accepting a non-leading role when appropriate; adopting a suitable forceful manner to make a point, and adopting for the team and/or situation
	Assessing capabilities	Judging different team members skills, and their ability to deal with a situation; being alert to factors, that may limit these and their capacity to perform effectively (eg. Level of expertise, experience, stress, fatigue)
	Supporting others	Providing physical, cognitive or emotional help to other members of the team
Situation awareness The skills to develop and maintain an overall dynamic awareness of the situation based on perceiving the elements of both the clinical and the aviation environment: patient, team, time, displays, equipment, aircraft operation, understanding what they mean and thinking ahead possible implications	Gathering information	Actively and specifically collecting data about the situation by continuously observing the whole environment and monitoring all available data sources and cues and verify data to confirm their reliability
	Recognizing & understanding	Considers and interprets information in light of the environment, identifies the match or mismatch between the situation and the expected state, updates one's current mental picture
	anticipating	Asking "what if" questions and thinking about potential outcomes and consequences of actions, intervention, non-intervention, etc.; running projection of current situation to predict what might happen in the near future
Decision making	Identifying options	Generating alternative possibilities or courses of actions to be considered in making a decision or solving a problem
	Balancing risks & Selecting options	Assessing hazards to weigh up the threats or benefits of a situation, considering the advantages and disadvantages of different courses of actions based on these processes

	Re-evaluating	Continually reviewing the suitability of the options identified, assessed and selected; and reprocessing the situation following the implementation of a given action
--	---------------	---

Table 3: Translated AeroNOTS (Norwegian) observation form (2016)

Oppgavehåndtering Evne til å organisere ressurser og oppgaver for å nå målene, både når det gjelder planer for enkelttilfeller eller planlegging av omfang over lengre tid.	Planlegging og forberedelse	Forhåndsplanlegge første skritt og videre strategier for oppgavehåndtering, vurdere og oppdatere planene om nødvendighet for å sikre måloppnåelse, iverksette nødvendige tiltak for å sikre at planen oppnås.
	Prioritering	Prioritere oppgaver, aktiviteter, områder, informasjon og kanaler osv i forhold til viktighet (for eksempel relatert til tid, alvorlighetsgrad og planer); være i stand til å identifisere de viktigste områdene og fokusere relevant på disse samt å unngå å bli distraheret av mindre viktige eller irrelevante gjøremål.
	Opprettholde standarder/prosedyrer	Styrke sikkerhet og kvalitet ved å holde seg til anerkjente prinsipper for pasienttransport, når det er mulig følge god praksis, behandlingsprosedyrer og retningslinjer samt mentale sjekklister.
	Identifisere og bruke ressurser	Etablere nødvendige og tilgjengelige krav for oppgaveutførelse (for eksempel personell, ekspertise, utstyr, tid) samt bruke disse til å fullføre målsettingen med et minimum av forstyrrelser, stress, over- eller underarbeid (mentalt og fysisk) på individ- eller gruppenivå.
Teamarbeid Evne til å arbeide sammen med andre i team, i alle slags roller, for å sikre en samlet, effektiv oppgavehåndtering og god lagånd; fokus er særlig på teamet heller enn på oppgaven.	Koordinere aktiviteter i teamet.	Arbeide sammen med andre for å utføre en oppgave, for både fysiske og kognitive aktiviteter; forstå rollene og ansvaret til hvert enkelt teammedlem samt forsikre seg om at felles forståelse i tilnærmingen etableres.
	Utveksle informasjon	Gi og motta kunnskap og data som er nødvendig for å koordinere teamet og oppgaveutførelsen.
	Bruke autoritet og myndighet.	Lede teamet og/eller oppgaven (som påkrevd) og akseptere en ikke-ledende rolle når det passer; skaffe seg relevant gjennomslagskraft tilpasset teamet og/eller situasjonen.
	Vurdere den enkeltes evner.	Vurdere ulike teammedlemmers ferdigheter og deres mulighet for å håndtere situasjonen; Å være var for faktorer som kan begrense disse samt deres evne og kapasitet til effektivt å utføre arbeidet (feks ekspertnivå, erfaring, stress, utmattelse)
	Støtte andre	Sørge for fysisk, kognitiv og emosjonell hjelp til andre medlemmer i teamet.
Situasjonsforståelse Evne til å utvikle og opprettholde en overordnet dynamisk forståelse av situasjonen på bakgrunn av oppfattet informasjon om kliniske og luftfartsmessige forhold: pasient, team, tid, instrumenter og utstyr, flytekniske forhold, forstå hva disse innebærer og forhåndsplanlegge for mulige komplikasjoner.	Samle informasjon	Aktivt og særskilt innhente data om situasjonen ved hele tiden å observere hele miljøet og overvåke alle tilgjengelige data- og informasjonskilder samt stadfeste data for å bekrefte at disse er pålitelige.
	Oppdage og forstå	Vurderer og tolker informasjonen i lys av miljøet, identifiserer samsvar eller manglende samsvar mellom situasjonen og forventet nivå, korjierer det mentale bildet.

	Beredskap	Stille "hva om"- spørsmål og vurdere potensielle utfall og konsekvenser av handlinger; inngrep ikke-inngrep osv, fortløpende bruk av her-og-nå-situasjonen til å forutsi hva som kan hende i nær fremtid.
Beslutningstaking	Identifisere muligheter	Utarbeide alternative muligheter eller mulige handlingsforløp i beslutningstaking eller problemløsning.
	Vurderer risiko opp mot valg av muligheter	Vurdere farer og veie disse opp mot trusselnivå eller fordeler i situasjonen, vurdere fordeler og ulemper ved ulike handlingsalternativ basert på disse prosessene.
	Revurdering	Sjekke kontinuerlig om identifiserte, vurderte og valgte muligheter er relevante; revurdere situasjonen basert på utførelsen av en valgt handling.

Table 4: Vurderingsskala for klinisk utførelse

Etter å ha sett dette simulerte luftambulanseoppdraget, hvordan vil du vurdere deltakernes kliniske utførelse. Deltakerne scores for hvert enkelt underelement i skjemaet over. Disse summeres og danner en totalscore.

5 = Utmerket	Utførelse på høyeste nivå, alle oppgaver svært godt ivaretatt inkludert pasientsikkerhet.
4 = God	Kompetent utførelse, oppgaver utført adekvat og pasientsikkerhet opprettholdes.
3 = Akseptabel	Adekvat utførelse på et nogenlunde ferdighets- og sikkerhetsnivå som forventet.
2 = Marginal	Utførelse litt under forventet standard, noen feil som kunne hatt potensielle konsekvenser for pasientsikkerheten
1 = Ikke akseptabel	Utførelse godt under forventet standard, åpenbare hull i ferdigheter og pasientsikkerhet

Table 5 AERONOTS OBSERVASJONSSKJEMA (MYERS, 2016)
Norwegian, short version

Dato: Navn Crew:

Oppgavehåndtering	Planlegging og forberedelse	
	Prioritering	
	Opprettholde standarder/prosedyrer	
	Identifisere og bruke ressurser	
Teamarbeid	Koordinere aktiviteter i teamet.	
	Utveksle informasjon	
	Bruke autoritet og myndighet.	
	Vurdere den enkeltes evner.	
	Støtte andre	
Situasjonsforståelse	Samle informasjon	
	Oppdage og forstå	
	Beredskap	
Beslutningstaking	Identifisere muligheter	
	Vurderer risiko opp mot valg av muligheter	
	Revurdering	

Attachment 1:

Forespørsel om deltakelse i forskningsprosjektet

Vurdering av ikke-tekniske ferdigheter hos leger i Norsk Luftambulansse

Bruk av et vurderingsverktøy i simuleringsbasert trening.

Bakgrunn og formål

Mitt navn er Solveig Gjertsen og jeg er mastergradsstudent i helsevitenskap ved Institutt for helsefag på Universitetet i Stavanger. Jeg er anestesisykepleier med 20 års erfaring fra Stavanger Universitetssykehus. Min masteroppgave gjøres på Universitetet i Stavanger i nært samarbeid med Norsk Luftambulansse (NLA). Oppgaven er en del av et større forskningsprosjekt om simulering. Hensikten med studien er å se på om vurderingsverktøyet AeroNOTS kan være et redskap for å vurdere legenes ikke-tekniske ferdigheter i NLA. Studien vil være en replikasjonsstudie av Myers (2015) sin studie på New Zealand i 2015 hvor hensikten var å se om ikke-tekniske ferdigheter hos leger i luftambulansen varierte i forhold til alder, erfaring og utdanning.

Hva innebærer deltakelse i studien?

Det vil gjennomføres en observasjonstudie hvor observatørene vil benytte scoringsverktøyet AeroNOTS til å score anestesilegers ikke-tekniske ferdigheter (oppgavehåndtering, teamarbeid, situasjonsforståelse og beslutningstaking) i gjennomføring av simuleringstrening på camp Torpmoen.

To observatører (Knut Styrkson, NLA og undertegnede) vil være tilstede under et av de to medisinske scenarioene som gjøres under campen. Vi vil være plassert slik at vi er synlige og har mulighet til å observere scenarioet uten å hindre deltakerne. Deltakerne vil bruke radiokommunikasjon som de normalt bruker under slike oppdrag under scenarioet. Lydopptakene vil også være en del av datamaterialet.

Data som samles inn vil behandles i dataverktøyet SPSS og analyseres ved hjelp av ulike statistiske tester, før analyse og drøfting av resultatene. Inklusjonskriterier i studiet er alle anestesileger som deltar på Camp Torpmoen i år fra uke 34- 39 i 2016 Observasjonene finner sted hver onsdag i perioden av de to observatørene.

Hva skjer med informasjonen om deg?

Alle personopplysninger vil bli behandlet slik at de ikke kommer på avveie.

Samtykkeerklæringer vil bli nummerert, og det lages en liste over navn med kobling til nummer som oppbevares på min datamaskin. Observasjonsskjema og samtykkeskjema får samme nummerering og oppbevares adskilt fra hverandre under behandlingen av det statistiske datamaterialet.

Data vil være tilgjengelig for Norsk Luftambulansses forskningsmiljø og Universitetet i Stanger. Ingen deltakere vil kunne gjenkjennes ved eventuelle senere publikasjoner.

Frivillig deltakelse

Det er frivillig å delta i studien. Dersom du ikke ønsker å delta vil treningen avvikles som normalt. Du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Personopplysninger vil bli slettet ved innlevering av masteroppgaven i juni 2017.

Dersom du ønsker å delta eller har spørsmål til studien, ta kontakt med Solveig Gjertsen.

E-post: solveig.gjertsen@uis.no. Telefon: 992 93 295

Min veileder ved UiS er: Marianne Storm. E-post: marianne.storm@uis.no

Kontaktperson i NLA: Elisabeth Jeppesen E-post: elisabeth.jeppesen@norskluftambulanse.no

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Attachment 2:

Marianne Storm
Norsk hotellhøgskole (NHS) Institutt for økonomi og ledelsesfag Universitetet i
Stavanger
4036 STAVANGER
Vår dato: 12.08.2016 Vår ref: 49139 / 3 / AMS Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 01.07.2016.
Meldingen gjelder prosjektet:

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>.

Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet. Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://pvo.nsd.no/prosjekt>.

Personvernombudet vil ved prosjektets avslutning, 01.07.2017, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen
Kontaktperson: Anne-Mette Somby tlf: 55 58 24 10

*49139 Vurdering av ikke-tekniske ferdigheter hos leger i Norsk Luftambulansse.
Bruk av et vurderingsverktøy i simuleringsbasert trening.
Behandlingsansvarlig Universitetet i Stavanger, ved institusjonens øverste leder
Daglig ansvarlig Marianne Storm
Student Solveig Gjertsen*

Kjersti Haugstvedt
Anne-Mette Somby

Vedlegg: Prosjektvurdering

Kopi: Solveig Gjertsen solveig.gjertsen@uis.no

Personvernombudet for forskning

Prosjektvurdering - Kommentar

Prosjektnr: 49139

Utvalget informeres skriftlig om prosjektet og samtykker til deltakelse. Informasjonsskrivet er godt utformet. Personvernombudet legger til grunn at forsker/student etterfølger Universitetet i Stavanger sine interne rutiner for datasikkerhet. Forventet prosjektslutt er 01.07.2017. Ifølge prosjektmeldingen skal innsamlede opplysninger da anonymiseres. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan gjenkjennes. Det gjøres ved å:

- slette direkte personopplysninger (som navn/koblingsnøkkel)
- slette/omskrive indirekte personopplysninger (identifiserende sammenstilling av bakgrunnsopplysninger som f.eks. bosted/arbeidssted, alder og kjønn).

Attachment 3:

Criteria

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine strongly encourages that all datasets on which the conclusions of the paper rely should be available to readers. We encourage authors to ensure that their datasets are either deposited in publicly available repositories (where available and appropriate) or presented in the main manuscript or additional supporting files whenever possible. Please see Springer Nature's information on recommended repositories.

Preparing your manuscript

The information below details the section headings that you should include in your manuscript and what information should be within each section. Please note that your manuscript must include a 'Declarations' section including all of the subheadings (please see below for more information).

Title page

The title page should: present a title that includes, if appropriate, the study design e.g.: "A versus B in the treatment of C: a randomized controlled trial", "X is a risk factor for Y: a case control study", "What is the impact of factor X on subject Y: A systematic review" or for non-clinical or non-research studies a description of what the article reports list the full names, institutional addresses and email addresses for all authors if a collaboration group should be listed as an author, please list the Group name as an author. If you would like the names of the individual members of the Group to be searchable through their individual PubMed records, please include this information in the "Acknowledgements" section in accordance with the instructions below indicate the corresponding author.

Abstract

The Abstract should not exceed 350 words. Please minimize the use of abbreviations and do not cite references in the abstract. Reports of randomized controlled trials should follow the CONSORT extension for abstracts. The abstract must include the following separate sections:

Background:	The context and purpose of the study
Methods:	How the study was performed and statistical tests used
Results:	The main findings
Conclusions:	Brief summary and potential implications
Trial registration:	If your article reports the results of a health care intervention on human participants, it must be registered in an appropriate registry and the registration number and date of registration should be stated in this section. If it was not registered prospectively (before enrollment of the first participant), you should include the words 'retrospectively registered'. See our editorial policies for more information on trial registration.

Keywords: Three to ten keywords representing the main content of the article.

Background:

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine Original Research

<http://sjtrem.biomedcentral.com/submission-guidelines/preparing-your-manuscript/original-research-articles>[15.06.2017 12.55.06]

The Background section should explain the background to the study, its aims, a summary of the existing literature and why this study was necessary or its contribution to the field.

Methods

The methods section should include: the aim, design and setting of the study the characteristics of participants or description of materials a clear description of all processes, interventions and comparisons. Generic drug names should generally be used. When proprietary brands are used in research, include the brand names in parentheses the type of statistical analysis used, including a power calculation if appropriate.

Results

This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

Discussion

This section should discuss the implications of the findings in context of existing research and highlight limitations of the study.

Conclusions

This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study reported.

List of abbreviations

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations should be provided.

Declarations

All manuscripts must contain the following sections under the heading 'Declarations':

Ethics approval and consent to participate

Consent for publication

Availability of data and material

Competing interests

Funding

Authors' contributions

Acknowledgements

Authors' information (optional)

Please see below for details on the information to be included in these sections.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section. Scandinavian Journal of Trauma,

Resuscitation and Emergency Medicine - Original Research

<http://sjtrem.biomedcentral.com/submission-guidelines/preparing-your-manuscript/original-research-articles>[15.06.2017 12.55.06]

Ethics approval and consent to participate

Manuscripts reporting studies involving human participants, human data or human tissue must: include a statement on ethics approval and consent (even where the need for approval was waived) include the name of the ethics committee that approved the study and the committee's reference number if appropriate

Studies involving animals must include a statement on ethics approval. See our editorial policies for more information.

If your manuscript does not report on or involve the use of any animal or human data or tissue, please state "Not applicable" in this section.

Consent for publication

If your manuscript contains any individual person's data in any form (including individual details, images or videos), consent for publication must be obtained from that person, or in the case of children, their parent or legal guardian. All presentations of case reports must have consent for publication. You can use your institutional consent form or our consent form if you prefer. You should not send the form to us on submission, but we may request to see a copy at any stage (including after publication). See our editorial policies for more information on consent for publication. If your manuscript does not contain data from any individual person, please state "Not applicable" in this section.

Availability of data and materials

All manuscripts must include an 'Availability of data and materials' statement. Data availability statements should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the findings reported

in the article. We recognise it is not always possible to share research data publicly, for instance when individual privacy could be compromised, and in such instances data availability should still be stated in the manuscript along with any conditions for access. Data availability statements can take one of the following forms (or a combination of more than one if required for multiple datasets): The datasets generated and/or analysed

during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS]. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

All data generated or analysed during this study are included in this published article [and its supplementary information files]. The datasets generated and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

The data that support the findings of this study are available from [third party name] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine | Original Research <http://sjtrem.biomedcentral.com/submission-guidelines/preparing-your-manuscript/original-research-articles>[15.06.2017 12.55.06] permission of [third party name]. Not applicable. If your manuscript does not contain any data, please state 'Not applicable' in this section.

More examples of template data availability statements, which include examples of openly available and restricted access datasets, are available [here](#).

BioMed Central also requires that authors cite any publicly available data on which the conclusions of the paper rely in the manuscript. Data citations should include a persistent identifier (such as a DOI) and should ideally be included in the reference list. Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style. Dataset identifiers including DOIs should be expressed as full URLs. For example: Hao Z, AghaKouchak A, Nakhjiri N, Farahmand A. Global integrated drought monitoring and prediction system (GIDMaPS) data sets. figshare. 2014. <http://dx.doi.org/10.6084/m9.figshare.853801>

With the corresponding text in the Availability of data and materials statement:
The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].

Competing interests

All financial and non-financial competing interests must be declared in this section. See our [editorial policies](#) for a full explanation of competing interests. If you are unsure whether you or any of your co-authors have a competing interest please contact the editorial office. Please use the authors initials to refer to each author's competing interests in this section. If you do not have any competing interests, please state "The authors declare that they have no competing interests" in this section.

Funding

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared.

Authors' contributions

The individual contributions of authors to the manuscript should be specified in this section. Guidance and criteria for authorship can be found in our editorial policies. Please use initials to refer to each author's contribution in this section, for example: "FC analyzed and interpreted the patient data regarding the hematological disease and the transplant. RH performed the histological examination of the kidney, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript."

Acknowledgements

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.

[Reference number]

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine | Original Research

<http://sjtrem.biomedcentral.com/submission-guidelines/preparing-your-manuscript/original-research-articles>[15.06.2017 12.55.06]. See our editorial policies for a full explanation of acknowledgements and authorship criteria. If you do not have anyone to acknowledge, please write "Not applicable" in this section. Group authorship (for manuscripts involving a collaboration group): if you would like the names of the individual members of a collaboration Group to be searchable through their individual PubMed records, please ensure that the title of the collaboration Group is included on the title page and in the submission system and also include collaborating author names as the last paragraph of the "Acknowledgements" section. Please add authors in the format First Name, Middle initial(s) (optional), Last Name. You can add institution or country information for each author if you wish, but this should be consistent across all authors. Please note that individual names may not be present in the PubMed record at the time a published article is initially included in PubMed as it takes PubMed additional time to code this information.

Authors' information

This section is optional.

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background

information. Please refer to authors using their initials. Note this section should not be used to describe any competing interests.

Endnotes

Endnotes should be designated within the text using a superscript lowercase letter and all notes (along with their corresponding letter) should be included in the Endnotes section. Please format this section in a paragraph rather than a list.

References

All references, including URLs, must be numbered consecutively, in square brackets, in the order in which they are cited in the text, followed by any in tables or legends. The reference numbers must be finalized and the reference list fully formatted before submission.

Examples of the BioMed Central reference style are shown below. Please ensure that the reference style is followed precisely. See our editorial policies for author guidance on good citation practice. Web links and URLs: All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database. <http://tumor.informatics.jax.org/mtbwi/index.do>. Accessed 20 May 2013.

If an author or group of authors can clearly be associated with a web link (e.g. for blogs) they should be included in the reference. Example reference style:

Article within a journal

Smith JJ. The world of science. *Am J Sci*. 1999;36:234-5.

Article within a journal (no page numbers)

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine | Original Research

<http://sjtrem.biomedcentral.com/submission-guidelines/preparing-your-manuscript/original-research-articles>[15.06.2017 12.55.06]

Rohrmann S, Overvad K, Bueno-de-Mesquita HB, Jakobsen MU, Egeberg R, Tjønneland A, et al. Meat consumption and mortality - results from the European Prospective Investigation into Cancer and Nutrition. *BMC Med*. 2013;11:63.

Article within a journal by DOI

Slifka MK, Whitton JL. Clinical implications of dysregulated cytokine production. *Dig J Mol Med*. 2000; doi:10.1007/s801090000086.

Article within a journal supplement

Frumin AM, Nussbaum J, Esposito M. Functional asplenia: demonstration of splenic activity by bone marrow scan. *Blood* 1979;59 Suppl 1:26-32.

Book chapter, or an article within a book

Wyllie AH, Kerr JFR, Currie AR. Cell death: the significance of apoptosis. In: Bourne GH, Danielli JF, Jeon KW, editors. *International review of cytology*. London:

Academic; 1980. p. 251-306. OnlineFirst chapter in a series (without a volume designation but with a DOI)

Saito Y, Hyuga H. Rate equation approaches to amplification of enantiomeric excess and chiral symmetry breaking. *Top Curr Chem*. 2007. doi:10.1007/128_2006_108.

Complete book, authored

Blenkinsopp A, Paxton P. Symptoms in the pharmacy: a guide to the management of common illness. 3rd ed.

Oxford: Blackwell Science; 1998.

Online document

Doe J. Title of subordinate document. In: The dictionary of substances and their effects. Royal Society of Chemistry. 1999. [http://www.rsc.org/dose/title of subordinate document](http://www.rsc.org/dose/title%20of%20subordinate%20document). Accessed 15 Jan 1999.

Online database

Healthwise Knowledgebase. US Pharmacopeia, Rockville. 1998.

<http://www.healthwise.org>. Accessed 21 Sept 1998.

Supplementary material/private homepage

Doe J. Title of supplementary material. 2000. <http://www.privatehomepage.com>. Accessed 22 Feb 2000.

University site

Doe, J: Title of preprint. <http://www.uni-heidelberg.de/mydata.html> (1999). Accessed 25 Dec 1999.

FTP site

Doe, J: Trivial HTTP, RFC2169. <ftp://ftp.isi.edu/in-notes/rfc2169.txt> (1999). Accessed 12 Nov 1999.

Organization site

ISSN International Centre: The ISSN register. <http://www.issn.org> (2006). Accessed 20 Feb 2007.

Dataset with persistent identifier

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, et al. Genome data from sweet and grain sorghum (*Sorghum bicolor*). GigaScience Database. 2011. <http://dx.doi.org/10.5524/100012>.