Aus der
Klinik für Anästhesiologie mit Schwerpunkt operative Intensivmedizin (CCM/CVK),
Charité - Universitätsmedizin Berlin

DISSERTATION

Effects of Telemedical Support on Quality of Emergency Information Retrieval Considering Offshore Wind Power Infrastructure

Effekte von telemedizinischer Unterstützung auf die Qualität von Notfallinformationsgewinnung unter Berücksichtigung von Offshore Windkraftinfrastruktur

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité - Universitätsmedizin Berlin

von

Philipp Landgraf, M.Sc.
aus Berlin.

Datum der Promotion: 21. Juni 2020

Für meine Eltern und meinen Bruder.

Table of Contents

Abstrakt

Einleitung

In Deutschland wird eine Verbesserung der medizinischen Versorgungsqualität mit einer Verbesserung von Gesundheit und Langlebigkeit assoziiert. Sie wird allerdings durch ökonomische und demographische Entwicklungen gefährdet.

In Offshore-Windparks sind diese Gefahren besonders greifbar, da bereits eine groSSe Lücke in der medizinischen Notfallversorgung besteht.

Telemedizin könnte einen Lösungsansatz darstellen.

In Landgraf et al. [1] haben wir über die Auswirkung telemedizinischer Unterstützung auf den ersten Teilprozess der Erstreaktion im Notfall berichtet.

Methoden

In simulierten, identischen Windkraftanlagen wurden Teams von zwei medizinischen Laien (Offshore-Wartungsingenieure) und Teams von zwei medizinischen Fachleuten (Rettungsassistenten) in einem simulierten Offshore-Notfallszenario mit einem simulierten, polytraumatisierten Patienten konfrontiert. Die Teams wurden nach dem Zufallsprinzip in unabhängige Gruppen eingeteilt (Intervention versus keine Intervention). Die Teams in den Interventionsgruppen wurden telemedizinisch durch Übertragungstechnik mit Verbindung zu einem entfernten Arzt unterstützt.

Alle Teams sollten eine Primärbefragung des Patienten durchführen. Mithilfe von Videoaufzeichnungen bewerteten Rezensenten ihre in Einheiten aufgegliederte Leistung sowie benötigte Zeit. Diese Leistungswerte und Zeiten wurden auf Auswirkungen von Fachwissen und telemedizinischer Unterstützung untersucht. Kern der Untersuchung war eine Nichtunterlegenheitsprüfung von unterstützten medizinischen Laien gegenüber nicht unterstützten medizinischen Fachkräften.

Ergebnisse

36 Offshore-Ingenieure in Zweierteams besetzten 18, 34 Rettungsassistenten in Zweierteams besetzten 17 simulierte, identische Windkraftanlagen. Nach Randomisierung wurden neun Ingenieurteams sowie neun Rettungsassistententeams telemedizinisch unterstützt, sodass neun Ingenieurteams sowie acht Rettungsassistententeams nicht unterstützt wurden. Die Leistung der unterstützten Ingenieure war besser als die der nicht unterstützten Ingenieure ($p < 0.01$) und derjenigen von nicht unterstützten Rettungsassistenten (bei einer Unterlegenheitsschwelle von einem Leistungspunkt) nicht unterlegen ($p = 0.03$). Der Unterschied zwischen unterstützten und nicht unterstützten Rettungsassistenten war nicht signifikant ($p = 0.11$). Ohne Unterstützung haben Rettungsassistenten Ingenieure übertroffen ($p < 0.01$). Unterstützte Gruppen waren langsamer als nicht unterstützte Gruppen ($p < 0.01$) [1].

Fazit

Die erste Reaktion auf medizinische Notfälle in Offshore-Windparks mit erheblich verzögerter professioneller Hilfe kann durch telemedizinische Unterstützung verbessert werden. Zukünftige Arbeiten sollten unser Ergebnis in zusätzlichen Szenarien testen und interdisziplinäre sowie systemische Aspekte untersuchen [1].

Abstract

Introduction

In Germany, improvement of quality of care is associated with improvement of health and longevity but threatened in several dimensions by economic and demographic trends. In offshore wind parks, these threats are particularly tangible with an already large gap of emergency medical care.

Telemedicine may be an approach to a solution.

In Landgraf et al. [1] we reported on the effect of telemedical support on the first subprocess of emergency first response.

Methods

In simulated, identical wind power plants, teams of two medical non-professionals (offshore maintenance engineers) and teams of two medical professionals (paramedics) faced a simulated polytraumatized patient in a simulated offshore emergency scenario. The teams were randomized into independent groups (intervention versus no intervention). Teams in the intervention groups were telemedically supported by transmission technology with connection to a remote physician.

All teams were to conduct a primary survey of the patient. Using videorecordings, reviewers scored their itemized performance and required time. These scores and times were explored for effects of expertise and telemedical support, culminating in a non-inferiority trial of supported medical non-professionals against unsupported medical professionals.

Results

36 offshore engineers in teams of two staffed 18, 34 paramedics in teams of two staffed 17 simulated, identical wind power plants. After randomization, nine teams of two engineers as well as nine teams of two paramedics were telemedically supported, leaving nine teams of two engineers as well as eight teams of two paramedics unsupported. Supported engineers' performance was better than that by unsupported engineers ($p < .01$) and non-inferior (at one item margin) to that by unsupported paramedics ($p = .03$). The difference between supported and unsupported paramedics was not significant ($p = .11$). "Without support, paramedics outperformed engineers ($p < .01$). Supported groups were slower than unsupported groups ($p < .01$)" [1].

Conclusion

"First response to medical emergencies in offshore wind farms with substantially delayed professional care may be improved by telemedical support. Future work should test our result during additional scenarios and explore interdisciplinary" [1] as well as systemic aspects.

1 Introduction

We currently sum up a high degree in a multitude of dimensions as high quality of care. The dimensions vary as expressed by different frameworks for measurement and improvement [2]. In Germany, as an example, improvement of quality of care is associated with improvement of health and longevity but threatened in several dimensions by economic and demographic trends [3].

Of these threatened dimensions, clinical quality and therein the concept of clinical effectiveness links the process of care with quality of care [2]. The very first sub process of emergency first response is information retrieval.

Of these economic trends, an increase of costs [4] and an optimization of organizational processes associated with a limitation of caregiver freedom, resources [5] and risk are relevant to the topic.

Of these demographic trends, an increasing demand for caregivers and a declining number of physicians [6] are but the tip of an iceberg.

One particularly tangible process illustrates this context: The first response to medical emergencies on offshore wind power plants.

As described in the introduction to Landgraf et al. [1] and more thoroughly in Stuhr et al. [7] and Durstewitz and Lange [8]: The offshore wind industry is growing, along with a demand for personnel working off shore who need medical support in emergencies. Traditionally, a telephone call with emergency services may trigger a rescue with delay depending on distance, weather and structural inaccessibility. Tantalizingly, high speed network connections are routinely established for monitoring and control of wind power plants.

Following the economic trend, decreasing unnecessary costs of treatment or rescue, risk of rescue and optimization of organizational processes are necessary. The costs may increase with deterioration of a patient due to delays and false decisions. The economically optimal organizational processes continue a first response that followed medical guidelines and supplied reliable information relevant to the clinical path planning, selection and preparation.

Following the demographic trend, decreasing demand for physicians is necessary. Utilizing the time required for rescue, delegating workload altogether and enabling additional remote support minimize the demand for physicians.

Therefore, we seek a network-based solution to avoid patient deterioration by reducing delays and false decisions, to enable a first response that follows medical guidelines, to gather reliable information for the ensuing processes and to spare (physician) resources.

The project Sea and Offshore Safety (SOS project, projekt-sos.charite.de, funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), funding code BMU-41V6169) developed and tested in a simulator study a concept for telemedical support in medical emergencies on offshore wind parks. The concept was aimed at improvement of quality of medical decisions pertaining to rescue urgency, initial treatment and preparation for further treatment utilizing real time transmission of audio, video and vital-sign data to a realistically feasible extent.

1.1 Objectives

In Landgraf et al. [1] we investigated whether “real-time teleconsultations for [medical] non-professional first responders in an emergency scenario improves guideline adherence and time to treatment surrogative of quality of care. Under the assumption of this scenario being commonplace for [medical] professionals, we used this group [of paramedics]’ performance as a benchmark.” Looking only at the very first sub process, that of information retrieval, the primary metric was a scoring of correctly attained information by teams of participants according to current emergency first response guidelines.

At the center of this question, the one-sided null hypothesis of no equivalence $H_0 : \Delta \leq \delta$, where Δ is the difference of the means of performance by supported medical non-professionals versus unsupported professionals with $\delta = -1$ as the chosen non-inferiority margin, is tested against the hypothesis of one-sided equivalence $H_1 : \Delta > \delta$; The difference in performance being greater than -1 item or in other words supported non-professionals being not inferior to unsupported professionals.

The secondary metric was time required for this retrieval. Moreover, inter-rater reliability in rating the correctness of information retrieval per team, the correctness per information item and required time were reported.

In this synopsis (German “Manteltext”[1]), current perspectives on telemedicine and information retrieval, additional details of the employed methods, findings, their applications and further research questions are reported[2]. The order and structure of this report follows closely that of Landgraf et al. [1] while avoiding repetition of published material except where necessary for coherence.

1.2 Emergency Telemedicine

According to Perednia and Allen [9], “TELEMEDICINE can be broadly defined as the use of telecommunications technologies to provide medical information and services. [...] this definition includes medical uses of the telephone, facsimile, and distance education”. The efforts reported by Hudson and Parker [10] in 1973, therefore, are some of the earliest efforts to telemedically support medical professionals by more expert medical professionals.

---

[1] This is the translation according to https://promotion.charite.de/en/procedure/regulations_2017/dissertation/, accessed 02.08.19 11:02 GMT+2.

[2] In accordance with https://promotion.charite.de/en/procedure/regulations_2017/dissertation/, accessed 02.08.19 11:02 GMT+2.

As Zachrison et al. [11] report, "In 2016, telemedicine was used in most US emergency departments (58 %)". The reasons for this trend are summarized by Culmer et al. [12] in a review of 1564 abstracts concerning prehospital telemedicine: "[C]ost was comparable or less than controls. Care quality was also found to be in line with or slightly preferable to face-to-face care with some advantages in response time and quality. Patients and providers were satisfied with the systems." This strengthened prehospital link of the chain of emergency survival enables the entire chain to achieve best results [13]. Some of this benefit may be due to "pervasiveness of informational and communicational information, [and][...] universal access to updated information"[14] as well as the reduced delays [12].

To maximise this reduction of delays, pervasiveness and universality of access to information along the chain of survival, the very first link - the patient and bystander - would also need to be telemedically supported. For most medical scenarios, this empowerment works well [15], often reducing false decisions and unnecessarily triggered chains [16].

Between actually low urgency scenarios of a patient themselves still conversing with a remote physician and probably fatal scenarios of a patient being unresponsive which have been investigated as use cases of telemedicine [12, 15], there are medical scenarios in which delay and reliable information could be decisive.

The polytraumatized patient considered by Landgraf et al. [1] is one of these scenarios. Distance, lack of physician resources but high speed network connection corroborate its potential usefulness for telemedicine research.

1.3 Emergency Information Retrieval

The initial step of first response should be to gather information about the patient [17] following the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) scheme and the Symptoms, Allergies, Medications, Past medical history, Last oral intake, Events (SAMPLE) patient history survey.

An incompleteness of this information was found by Stiell et al. [18] in 32.2 % of 1002 emergency department visits. Of these 32.2 %, 34.1 % had been brought in by an ambulance. More recently, Mashoufi et al. [19] found in their review on medical data quality reports between 2000 and 2015 that completeness, accuracy, consistency and accessibility were rated as low as 30 % in the completeness dimension.

In consequence, patients about whom the information was incomplete stayed an average of 1.2 hours longer in the emergency department [18]. On the other hand, "For each 1-hour reduction in [information] access time, visit length was 52.9 minutes shorter, the likelihood of imaging was lower [...], the likelihood of admission was 2.4 percentage points lower, and average charges were $ 1187 lower ($P \leq .001$ for all)"[20]. Furthermore, Janakiraman et al. [21] report on an inverse correlation of information quality with length of stay, 30-day readmission rate and likelihood of an additional physician being sought.

In summary, high quality - particularly complete - information may be relevant for decreasing time and physician requirements with a plausible increase in patient safety [22]. Therefore, these are metrics chosen by Landgraf et al. [1].

1.4 Trauma Response Empowerment

The scope of empowerment considered in this work includes only giving confidence and enabling control of a situation. It does not consider giving legal power or permission as neither is a limiting factor during simulation based research.

Bakke et al. [23] report on a majority of medical non-professional bystanders performing first aid in a majority of trauma cases observed in Norway. They did not report on quality of information retrieval or timeliness but did find a higher quality of first aid given by trained bystanders compared to untrained ones.

Training, therefore, empowers bystanders to perform aspects of first aid.

Latifi et al. [24] report on telepresence support by trauma surgeons improving quality of care by rural hospital staff for trauma patients.

From a consistency of improvement across levels of training, it follows that telemedicine may additionally empower bystanders even with training.

The offshore engineers considered here receive regular first aid training. In Landgraf et al. [1] we assume an improvement in information retrieval to signify an empowerment to retrieve information due to telemedical support.

1.5 Simulation

Within this work, simulation refers to the artificial representation of a simplified model of reality and its evolution. As an example, we may model an emergency scenario as a set of starting and boundary conditions with a limited number of possible trajectories and therefore a very finite realism - we refer to the model's extent of reproduction of reality / fit for purpose as fidelity [25]. Some of the functions of the human patient in this example are emulated, such as reacting to a stimulus, exhibiting vital signs according to a patient state model or anchoring participants in the emergency mindset.

This technique can be employed in education to great success [26]. Building on this experience, immersive simulations can enable rapid reproduction of rare and risky situations and process intervals which in turn enables prospective investigation - assuming the simulation is valid. One example highlighting this assumption is reported by Slakey et al. [27]; Models of real adverse outcomes were simulated to reveal more system errors and how complex decisions were made than conventional root cause analysis. The authors insist on the validity of the model (construct validity), that findings were due to the intervention (internal validity) and adequate to inform improvement of the real system (external validity) as has been shown for the transferability of skills. This validity remains a weakness of the technique and sustains it as a topic of research [28, 29].

The reproduction of rare and risky situations and process intervals, however, may justify this limitation where patient safety is concerned [25].

2 Additional Details of the Methods and Materials

As compiled by Lamé and Dixon-Woods [25], an observed situation could be modeled to evaluate candidate interventions. Here, the situation is the offshore emergency, the candidate intervention is telemedical support and the evaluation is a prospective, controlled simulation study. In the simulation study, qualitative findings supported by quantitative data "collected either in real time or through use of video recordings"[25] would be generalized and validated in a multicenter study.

The topic of this work is only the first part of the simulation study with data collected using video recordings, quantitative and qualitative analyses and generalization perspectives.

2.1 Fidelity

To immerse participants into a reproduction of a rare and risky situation, the real situation was modeled in three dimensions of simulation fidelity [30] as was practical within the scope of the SOS project: The patient, the clinical setting and the facilities.

2.1.1 Patient Dimension

To emulate a conscious, polytraumatized patient convincingly enough, the following simplifications were chosen[3]:

Externally, the shape, size and weight approximated that of a healthy male, while the surface material, consistency and mobility did not. Affixed to one arm and one leg were textured rubber phantoms of an open fracture and a severe burn, respectively. See figure 2.1 for an overview.

Internally, the palpability of radial and carotid pulses, interfaces with an arbitrary pulse oximetry sensor, for five-lead electrocardiography, defibrillation and intermittend non-invasive blood pressure monitoring were reproduced

Functionally, basic eye lid activity, pupillary light reflex, chest excursion, carbon dioxide exhaustion, heart and lung sounds but neither temperature, nor skin tone variation, nor facial expression, nor bleeding were emulated. Moaning, interactive speech and cognition were enacted and projected by audio transmission into the mannequin's head.

2.1.2 Clinical Setting Dimension

In any emergency, following the primary assessment in the ABCDE order [17], a secondary, history assessment covering the SAMPLE questions should be conducted alongside a survey of vital signs and where applicable the Onset, Provocation, Quality, Radiation, Severity and Time (OPQRST) pain assessment.

---

[3] Features not currently technically feasible were not considered and therefore left out of this list.

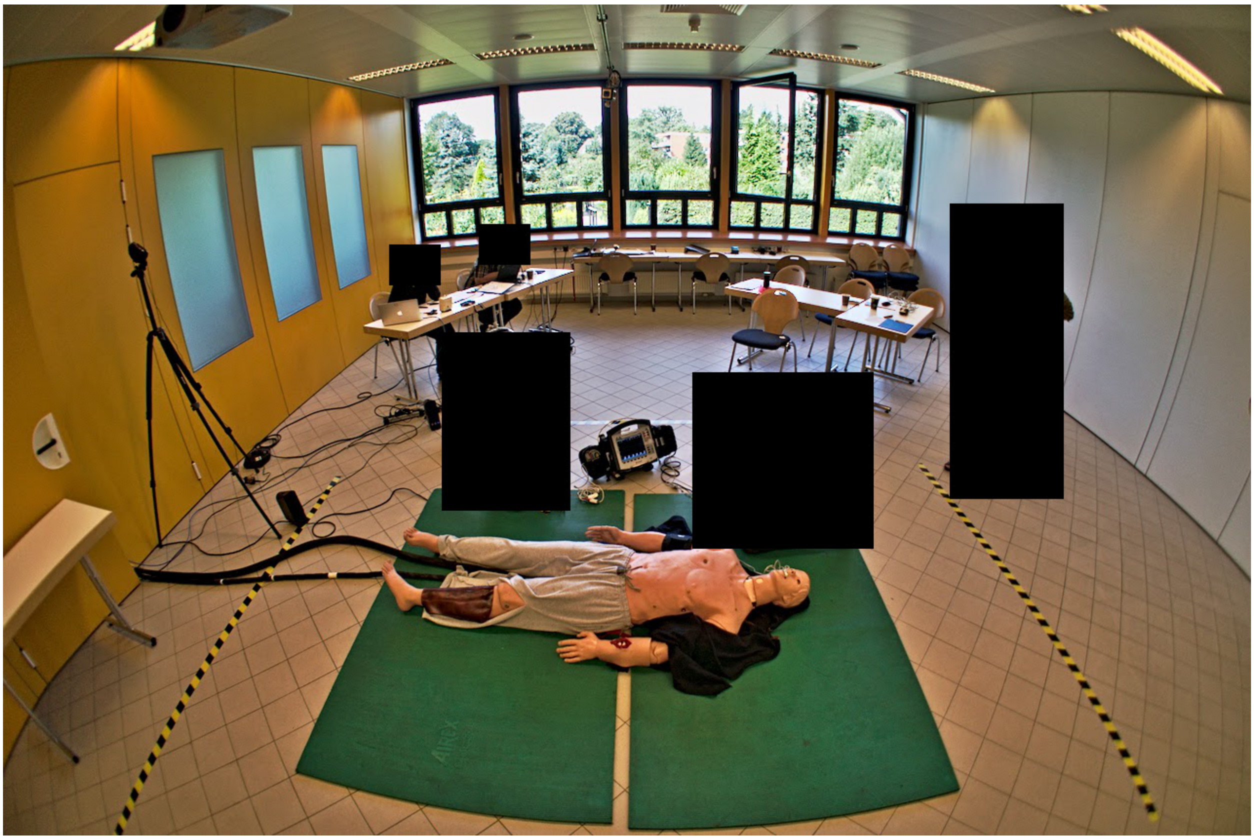

Fig. 2.1: The mannequin within the simulation area marked by tape models the space available in an offshore wind power plant. Black rectangles are used to protect identities of involved individuals. The picture was taken by one of the authors of Landgraf et al. [1], permission to use was given.

In a polytraumatized patient, primary and secondary assessment would lead into a rapid trauma assessment to identify life threats (bodycheck) and then into prioritized diagnostic or therapeutic actions. By repeating the same assessments, insight into status trends would be gained.

Taking advantage of the reproducibility of boundary conditions and to improve variable isolation, the process the project members would follow was cut into five sequential subprocesses. The first subprocess included the ABCDE scheme, the SAMPLE history and vital sign survey. The following subprocesses covered the bodycheck, polytrauma treatment, intraosseous drug administration and finally recording and transmitting a 12-lead electrocardiogram.

The scope of this work is only the first subprocess.

2.1.3 Facilities Dimension

The facilities in an offshore wind power plant are marked by confined space, often grated floors, fall hazards, electrocution hazards, moving parts and fire hazards. Taking full advantage of simulation to avoid any of these hazards, while adding comfort and enabling voluntary and safe participation, only the space was confined as shown in figure 2.1.

The advantage of an offshore wind power plant being connected to a high speed network for monitoring, control and maintenance was reproduced to enable usage of telemedical technology. A vital sign monitor able to transmit vital signs (monitor; *corpuls-3* with *corpuls.web*, GS Elektromedizinische Geräte G. Stemple GmbH, Kaufering, Germany) and a system of audio headset, head-mounted camera and hand held camera (commlink; *Frontline Communicator*, MDAI mobile solutions GmbH, Wolfratshausen, Germany) enabled support by a remote physician. The used technology's feasibility was an additional endpoint of the SOS project and had been deemed plausible among project members. Its application was intended to increase fidelity in comparison to mock-ups. Telemedical technology was only made available to the intervention group.

The availability of a standard emergency case was also reproduced after plausibility had been confirmed among project members.

Typical offshore protective gear was not modeled to avoid distraction and frustration of participants, examination gloves are part of a standard emergency case.

2.2 Selected Cohorts

Among the SOS project stakeholders and their networks, two cohorts stood out as relevant to the selected use case:

Offshore Maintenance Engineers are typically the patients and bystanders when offshore emergencies do occur. They share exceptional health as this is an eligibility criterion for their work. They share a language and large fields of experience because they have worked as colleagues at one of the project partners. Much of this experience is in efficiently, effectively and cautiously acting and following manual instructions and structuredly overcoming challenges. They routinely participate in life support training but are generally medical non-professionals.

Paramedics are typically among the first responders to medical emergencies, routinely conducting primary and secondary assessments, selecting and being an essential link of the chain of survival. Like the engineers, they too worked as colleagues at one employer and efficiently, effectively and cautiously care for patients until a physician arrives or assist them thereafter, often overcoming acute challenges. They have completed and regularly refresh extensive emergency medicine training and are state-certified emergency medicine professionals.

From these, volunteers were paired up into teams of two and randomized by lot into a total of four independent groups of teams: Medical non-professionals with or without and medical professionals with or without support[4]. Figure 2.2 illustrates this process. Neither may be representative of larger cohorts.

---

[4] Please find additional information in [1], [31] and [32].

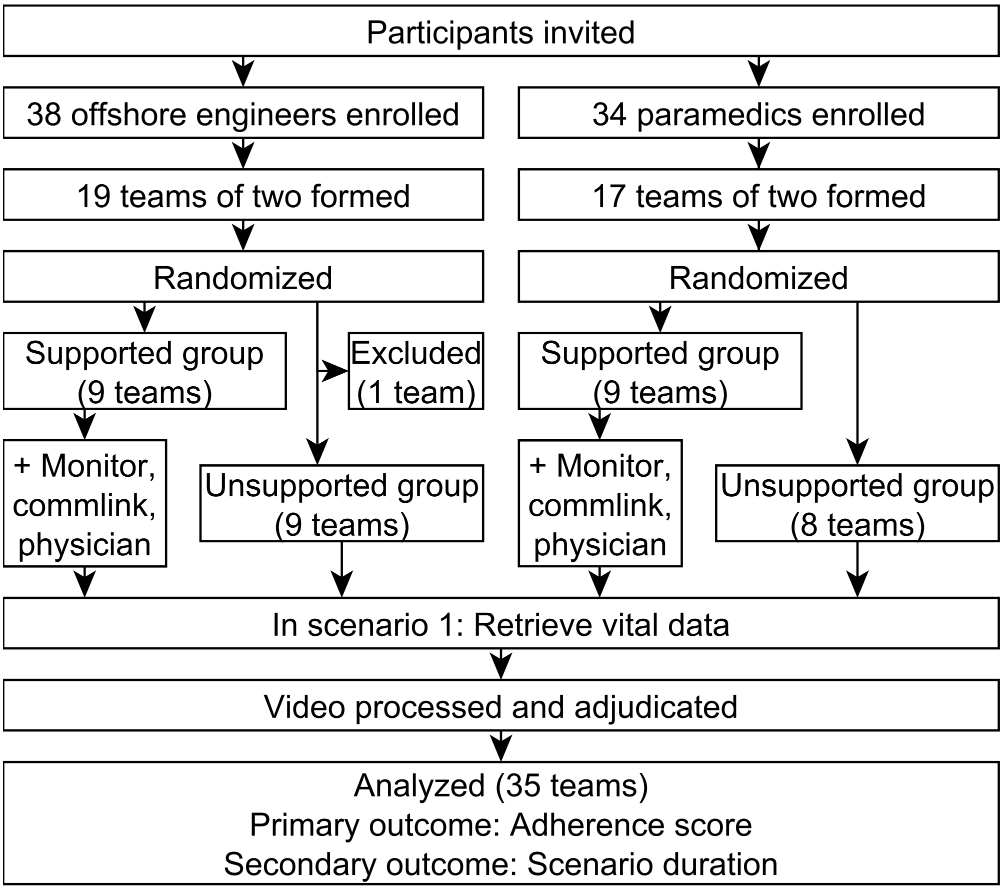

Fig. 2.2: "Flow diagram of the simulator study. Randomization into groups, half of them supported by a remote physician via a telemedically enabled monitoring unit and bidirectional commlink. One team dropped out due to illness." Figure (modified to show team size) and caption from Landgraf et al. [1].

2.3 Data Processing

Prior to the validation of metrics, measurement and the validation and analysis of data[4], the recordings from two perspectives on the simulation space and the telemedical support workstation were combined into a single, synchronous video perspective with the addition of a stop watch to improve consistency, precision and accuracy of time ratings as shown in figure 2.3.

These videos, approximately six hours in length, were redundantly and securely stored and reviewers individually introduced to quantifying performance and duration by the author[4].

The checklist in table 2.1 had been compiled by co-authors within the SOS project prior to the simulator study[4]. "In operationalizing the ERC guidelines, all applicable Good ReseArch for Comparative Effectiveness (GRACE) principles (*graceprinciples.org*) were considered. The checklist reliability and validity were probed in two test simulations. No efforts were undertaken to validate the duration metric." [1] This checklist was used by both reviewers. Their filled-out checklists were programmatically aggregated into a master sheet and organized into structures compatible with SPSS 24 (IBM, Armonk, USA), EquivTest (StatCon GmbH, Witzenhausen,


[redacted] [redacted] [redacted] [redacted]

Fig. 2.3: Two perspectives on the simulation space (top), one perspective on the telemedical support workstation (bottom) and a stop watch (bottom left). Black rectangles are used to protect identities of involved individuals. The picture was taken by one of the authors of Landgraf et al. [1], permission to use was given.

Germany) and nQuery Version 8.4.0.0 (Stat. Solutions Ltd. & South Bank, Crosses Green, Cork, Ireland), respectively.

The comparisons of groups as shown in figure 2.4 and described below were reported on in Landgraf et al. [1] (enumeration changed to better distinguish the main objective from explorative tests).

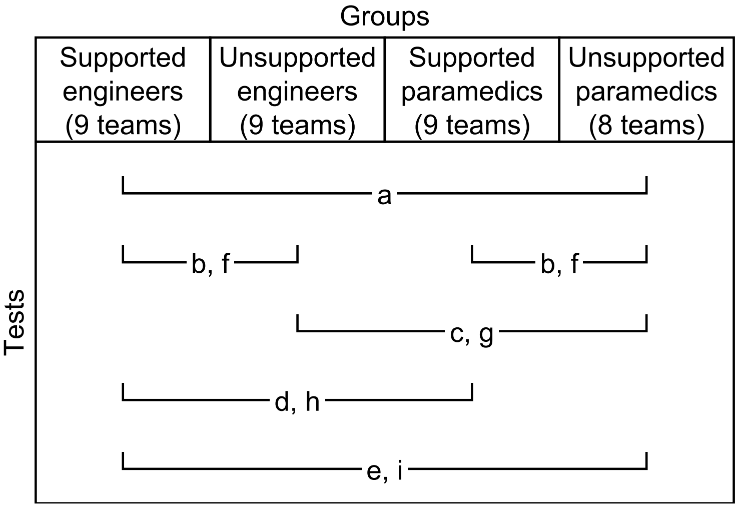

Fig. 2.4: Mapping of test indizes (vertical axis) to tested groups (horizontal axis) to indicate which groups were compared in which test. Where two tests compared the same groups, they compared different measures (performance or required time).

Index	Performance item	Description of correct result
1	Patient conscious	Yes
2	Patient orientated	At least one information retrieved: name, location, time
3	Glasgow Coma Scale	14-15
4	Does the patient breathe normally?	No
5	Respiratory rate	18-22 breaths per minute
6*	Respiratory rate correctly surveyed	Counted for 20 seconds or using monitoring device
7	Radial and if not carotid pulse	Both pulses perceptible
8*	Radial pulse correctly surveyed	By palpation
9*	Blood pressure correctly measured	Unsupported: Cuff placed correctly, stethoscope used in Riva-Rocci fashion Supported: Cuff placed correctly
10	Blood pressure	Unsupported: Korotkoff sound inaudible Supported: 110/60 mmHg using monitoring device
11	Blood pressure normal	Unsupported: Unclear evidence Supported: Yes
12	Symptoms	Patient was asked
13	Allergies	Patient was asked
14	Medication	Patient was asked
15	Past medical history	Patient was asked
16	Last meal	Patient was asked
17	Environment / Course of accident	Fall from 4 m height after burning lower leg

Tab. 2.1: “Checklist of itemized performance quality attributes as used by the remote physician and for evaluation per team and scenario. They are ordered according to the sequence the authors would perform them in.”
“* Item not asked at the end of the scenario but observed in the video material.”
Table and caption from [1].

The main objective of the study consisted in testing the one-sided hypothesis

a. Performance by supported engineers compared to unsupported paramedics is not inferior using the confidence interval inclusion method as well as Schuirmanns one-sided test at 5 % significance level with a 1 item ($\delta = -1$) non-inferiority margin.

Beside the (confirmative) test of the main hypothesis, the following tests in the area of explorative testing were performed:

“Average rater rating of number of items correctly performed per team (itemized performance) was analyzed with two-sided exact Mann-Whitney U tests for independent [groups][...] comparing

b. supported against unsupported engineers [as well as][...] supported against unsupported paramedics,

c. unsupported engineers against unsupported paramedics,

d. supported engineers against supported paramedics and

e. supported engineers against unsupported paramedics. [...]

Average rater measurement of scenario duration was evaluated using [two-sided exact Mann-Whitney U tests for independent groups][...] comparing

f. supported against unsupported engineers [as well as][...] supported against unsupported paramedics,

g. unsupported engineers against unsupported paramedics,

h. supported engineers against supported paramedics and

i. supported engineers against unsupported paramedics.

[Bonferroni-Holm-adjustment of the 5 % significance level was applied in each family of multiple comparisons[5] without additional tools.]

Finally, [...] [t]he number of teams per group who, by average rating, correctly performed each item was analyzed descriptively. [...]

Inter-rater reliability was estimated using average intraclass correlation coefficients in a two-way random-effects model for absolute agreement [...]. Quantile-quantile plots and minima in proportionate positive inter-rater agreement were explored." [1]

Output from all tools was added to the master sheet, as were data from interviews with project participants about their demographics, experience and evaluation of the applied technology. Only summative data were included in Landgraf et al. [1]. This master database was closed and submitted to clinical monitoring for archiving.

2.4 Non-inferiority Trial

The non-inferiority test looks for a test sample (supported non-professionals) performing no worse than a reference sample (unsupported professionals) by a defined margin. This margin should be defined prior to data analysis and medically justified. Usually, it is referred to as a margin of clinical importance.

As shown in figure 2.5, a test sample performed no worse than a reference sample by an amount of clinical importance, if the defined non-inferiority (lower equivalence) margin δ is smaller than the lower confidence interval limit. If a defined upper equivalence margin were also considered (in a two-sided test of equivalence instead of a one-sided test of non-inferiority), equivalence would be claimed if the confidence interval were covered by the equivalence interval.

---

[5] Each family (tests of performance and tests of required time) contained six tests; The five reported here plus the comparison of unsupported engineers versus supported paramedics which was of no interest being a matter of course.

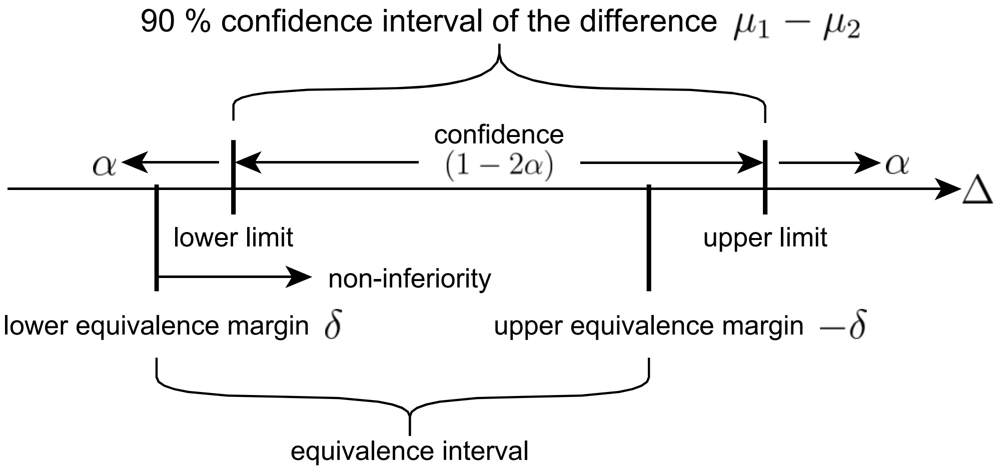

Fig. 2.5: Relationship of equivalence interval from lower equivalence margin $\delta < 0$ to upper equivalence margin $-\delta$ and confidence interval from lower limit to upper limit visualized along an axis for $\Delta = \mu_1 - \mu_2$, where μ_1 and μ_2 are means of independent samples. Testing inferiority $H_0 : \Delta \leq \delta$ against one-sided equivalence (non-inferiority) $H_1 : \Delta > \delta$ using Schuirmann's single one-sided test results in non-inferiority for lower confidence interval limits greater than the lower equivalence margin.

The probability of a type II error in testing for one-sided equivalence decreases (the power increases) with a diminution of the non-inferiority margin (lower equivalence bound) and/or a reduction of the common standard deviation (growth of effect size as well as diminishment of the confidence interval of the underlying difference of means).

Technically, the protocol followed in Landgraf et al. [1] in accordance with Chow and Liu [33] is an inferential, univariate, single one-sided t-test of equivalence for two independent samples of unequal size assuming equal variance (Schuirmann's "single one-sided test"). Furthermore, the confidence interval inclusion method for testing equivalence was used. We know from Liu and Weng [34] that the two-sided test, that is a one-sided test with a lower equivalence margin and another one-sided test with an upper equivalence margin, is robust against small deviations from normal distribution.

3 Additional Details of the Results

Except where declared otherwise, all citations in this section are from [1].

The sample sizes resulting from the recruitment and randomization into four groups of teams is visualized in figure 2.2.

"Medians (Md), 95 % confidence intervals of medians [as [lower limit, upper limit]][...], 1- / 2-sided exact significance (p) and effect size (d) are reported" where meaningful.

Significance after family-wise Bonferroni-Holm-adjustment of the 5 % significance level is denoted ($*$).

As visualized in figures 3 and 4 of Landgraf et al. [1]:

a. "Schuirmann's one-sided test rejects the null hypothesis of non-equivalence. Thus, non-inferiority of supported non-professionals' to unsupported professionals' performance within the specified equivalence bounds (difference of means less than or equal one item correctly performed) can be claimed."

supported non-professionals ($n = 9$)	Md	=	11.50	[10.50, 13.50]
unsupported professionals ($n = 8$)	Md	=	11.75	[9.25, 12.50]
Schuirmann one-sided test at 5 % significance with a 1 item margin		p	=	.03$*$

Group	supported non-professionals	unsupported professionals
Mean	11.8333	10.6875
Standard Error (Mean)	0.5892	0.9398
Geometric Mean	11.7135	10.3237
Median	11.5000	11.7500
Standard Deviation	1.7677	2.6583
Variance	3.1250	7.0669
Min	9.0000	5.5000
Max	14.0000	13.0000
Range	5.0000	7.5000
n	9	8
Common standard deviation σ	2.22998	

Tab. 3.1: "Descriptive statistics of the groups of supported non-professionals and unsupported professionals [...]. Common standard deviation was estimated by pooling standard deviations of both groups." The unit is items correctly performed except for n which counts the teams in each group. Table and caption from Landgraf et al. [31].

The common standard deviation (see table 3.1) is calculated as

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}$$

where s_1^2 and s_2^2 are sample variances from samples with n_1 and n_2 observations. The $(1-\alpha)$-confidence interval for the difference of means $\mu_1 - \mu_2$ is defined as

$$P\left(\overline{X}_1 - \overline{X}_2 - t_{\alpha/2,df}\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + t_{\alpha/2,df}\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}}\right) = 1-\alpha$$

with $\overline{X}_1$ and $\overline{X}_2$ denoting arithmetic means of samples and $t_{\alpha/2,df}$ the quantile of the t-distribution with $df = n_1 + n_2 - 2$ degrees of freedom.

In one-sided equivalence testing, the $(1-2\alpha)$-confidence interval is to be used [35]. In other words, we may calculate the confidence interval for $\alpha = 10\,\%$ or $1-\alpha = 90\,\%$.

The corresponding 90 %-confidence limits for the difference of means in our example are therefore given as

$$\overline{X}_1 - \overline{X}_2 - t_{\alpha/2,df}\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}} = -.7522 \quad \text{(lower limit)}$$

and

$$\overline{X}_1 - \overline{X}_2 + t_{\alpha/2,df}\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}} = 3.0438 \quad \text{(upper limit)}$$

The lower confidence limit is above the non-inferiority margin ($\delta = -1$), therefore one-sided non-inferiority has been directly proved by confidence interval inclusion. Moreover, Schuirmann's one-sided test rejects the null-hypothesis of non-equivalence with $p = .0330$.

b. "Supported outperformed unsupported participants. This intervention effect was statistically significant among non-professionals but not among professionals."

unsupported non-professionals ($n = 9$)	Md	=	4.50	[3.50, 7.50]
supported non-professionals ($n = 9$)	Md	=	11.50	[10.50, 13.50]
Two-sided exact Mann-Whitney U test for independent measures	d	=	3.13,	p < .01∗
unsupported professionals ($n = 8$)	Md	=	11.75	[9.25, 12.50]
supported professionals ($n = 9$)	Md	=	13.00	[12.00, 14.00]
Two-sided exact Mann-Whitney U test for independent measures	d	=	.86,	p = .11

c. "Performance differences due to medical expertise is statistically significant in unsupported groups; professionals correctly performed twice as many items as non-professionals."

unsupported non-professionals ($n = 9$)	Md	=	4.50	[3.50, 7.50]
unsupported professionals ($n = 8$)	Md	=	11.75	[9.25, 12.50]
Two-sided exact Mann-Whitney U test for independent measures	d	=	2.25,	p < .01∗

d. “With telemedical support, this difference was not statistically significant.”

supported non-professionals ($n = 9$) $Md = 11.50$ $[10.50, 13.50]$

supported professionals ($n = 9$) $Md = 13.00$ $[12.00, 14.00]$

Two-sided exact Mann-Whitney U test for independent measures $d = .54$, $p = .29$

e. “No statistical evidence was found that unsupported professionals outperformed supported non-professionals.”

supported non-professionals ($n = 9$) $Md = 11.50$ $[10.50, 13.50]$

unsupported professionals ($n = 8$) $Md = 11.75$ $[9.25, 12.50]$

Two-sided exact Mann-Whitney U test for independent measures $d = .38$, $p = .46$

f. “Participants who were supported required significantly more time compared with unsupported participants.”

unsupported non-professionals ($n = 9$) $Md = 6.55$ $[4.38, 8.55]$

supported non-professionals ($n = 9$) $Md = 11.87$ $[9.25, 13.05]$

Two-sided exact Mann-Whitney U test for independent measures $d = 1.95$, $p < .01*$

unsupported professionals ($n = 8$) $Md = 4.38$ $[3.92, 7.08]$

supported professionals ($n = 9$) $Md = 7.62$ $[7.05, 9.50]$

Two-sided exact Mann-Whitney U test for independent measures $d = 1.62$, $p < .01*$

g. “Without telemedical support, no statistical evidence was found for either group being faster.”

unsupported non-professionals ($n = 9$) $Md = 6.55$ $[4.38, 8.55]$

unsupported professionals ($n = 8$) $Md = 4.38$ $[3.92, 7.08]$

Two-sided exact Mann-Whitney U test for independent measures $d = .75$, $p = .16$

h. “In supported groups, non-professionals took more time.”

supported non-professionals ($n = 9$) $Md = 11.87$ $[9.25, 13.05]$

supported professionals ($n = 9$) $Md = 7.62$ $[7.05, 9.50]$

Two-sided exact Mann-Whitney U test for independent measures $d = 1.40$, $p = .01*$

i. “Supported non-professionals required more time than unsupported professionals.”

supported non-professionals ($n = 9$) $Md = 11.87$ $[9.25, 13.05]$

unsupported professionals ($n = 8$) $Md = 4.38$ $[3.92, 7.08]$

Two-sided exact Mann-Whitney U test for independent measures $d = 3.10$, $p < .01*$

“Non-professionals without support likely surveyed consciousness (item 1), pulse rate (8), symptoms (12) and environment (17) correctly (see figure 5 [in Landgraf et al. [1]]). Professionals in the intervention group more often correctly surveyed respiration (4-6) and food intake (16) but less often manually surveyed radial pulse (7-8)."

The sample sizes were not a result of statistical power calculations, as "knowledge of the feasibility and acceptability of the intervention, design, procedures, measurability of effects and estimation of precision" [1] was not yet gained. Instead, limited by the SOS project's resource budget, it was chosen as large as possible. In the post-hoc power analysis for the non-inferiority test, we determined a power of 22.26 % for the one-sided, two-group t-test of equivalence in means, assuming an expected mean difference of 0, a one-sided significance level of 5 %, a lower limit equivalence margin of -1 and a common standard deviation of 2.23.

"Inter-rater reliability was estimated using average intraclass correlation coefficients in a two-way random-effects model for absolute agreement ($ICC(2{,}2)$), 95 % confidence intervals are reported" [1] in brackets as [lower limit, upper limit]. The reliability of scores per item ($ICC_{\mathrm{ppi}}(2{,}2) = .66\,[.52, .74]$) was lowest, followed by that of scores per team ($ICC_{\mathrm{ppt}}(2{,}2) = .71\,[-.10, .90]$) with that of required time ($ICC_{\mathrm{tpt}}(2{,}2) = .99\,[.98, .99]$) being the highest.

Essential to this positive effect was the feasibility of the support concept and technology application.

4 Additional Details of the Discussion

4.1 Findings

In summary, "[s]upported non-professionals performance was non-inferior to that by unsupported professionals [...] [T]he telemedically supported non-professionals performed more items correctly than unsupported non-professionals [, they][...] performed slightly more items correctly than unsupported professionals.[...] The performance of additional items required additional time [which][...] does not necessarily delay rescue." [1]

This is a positive effect of telemedical support on the correctness and completeness of information retrieval and a negative effect on time required. However, "more accurate first response could expedite triggering the appropriate rescue chain, which is independently associated with greater survival [36] and plausibly reduce on-scene time [37]. During a support gap pertaining to the remoteness and accessibility of offshore wind farms, even slow information collection may significantly improve outcome if it is reliable and available to alert and involve relevant specialists especially before arrival of a rescue team [38–40]" [1].

Due to the reference count limit of this work, comparisons of these results with the literature remain only in Landgraf et al. [1]. In summary, no qualitative discrepancies were found.

Due to resource constraints on the SOS project, the sample size per group was less than twelve, the rule of thumb for pilot studies [41]. The post-hoc power considerations were published alongside descriptive statistics [31] for future investigators' reference.

The high inter-rater reliability on duration suggested there being no systematic error. A detailed analysis of outliers in observed proportion of positive agreement between the raters [32] yielded findings on internal and external validity with implications for increasing internal and external validity in future study designs.

4.2 Construct Validity

As described in Landgraf et al. [1], efforts were made to validate the clinical setting and facilities dimensions. The patient dimension was constrained by technology availability, similar in most simulation based research and education [26] and seems to be secondary in relevance [25].

The high score by medical professionals suggests the scenario was common to them. It was used to scale the metrics to resolve group differences and substantiate construct validity. The small, albeit potentially life-saving, effect the intervention had among the benchmark cohort of medical professionals confirms that telemedical support can further empower even trained professionals.

For medical non-professionals, a medical emergency scenario should be uncommon. Construct validity was probed in stakeholder and participant interviews and maximized by introducing the scenario designers thoroughly to the real environment and real scenario to be modeled [1].

4.3 Internal Validity

Controlling for team performance undue to the intervention was attempted by randomization by lot. A pre-post scoring without intervention could have added to confidence in the effect of this randomization but was not within the scope of the SOS project. Instead, simply the largest sample size available within tight resource constraints on the project was included. The intervention effect was therefore possibly superposed with non-intervention effects, the standard deviations high, the post-hoc power low. However, our results were statistically significant. These superposed non-intervention effects, therefore, did not obfuscate intervention effects.

Findings on inter-rater reliability for performance per team and per item suggest that disagreement stems from “an emerging frame of reference in which the raters’ interpretations became consistent, [...] differences in raters’ prioritization of accuracy of deductions versus precision of adherence to instructions or guidelines, [...] differences in raters’ understanding of item dependencies” [32]. These hidden variables in rater experience, expertise and interpretation should either be charted for isolation or homogenized. The homogenization might be achievable through repetition as a non-linear learning curve seemed to be visible among the raters but was not investigated due to the small sample size. A discussion among raters may homogenize but it is neither clear which biases would thereby spread, nor does this seem easily measurable.

4.4 External Validity

The same findings on inter-rater reliability for performance per team and per item suggest that disagreement also stems from “simulation artifacts causing participants to behave differently in an artificial setting than they might in reality” [32]. In detail, one of thirty five teams clearly behaved unrealistically. While this is no evidence for the realism of the behavior of all other teams, this outlier’s singularity may point to variability in the participants being the primary and insufficient simulation fidelity being the secondary cause.

Confirming immersion of participants in the simulation would corroborate the chosen fidelity but was not within the scope of the SOS project. In previous unpublished projects, immersion proved difficult to reliably measure so that a confirmation thereby seemed to be an uncertain prospect.

The higher scores among the benchmark cohort of medical professionals suggests a transferability of skills and experience to the simulation. Pending evaluation in a multicenter trial [25], this applicability of skills and experience to the simulation may plausibly hold vice versa and for intervention effects.

4.5 Perspectives for Applications and Research

Currently, quality of healthcare gains priority over its constituents[6] even in public discourse[42]. In shifting the priority from infrastructure and utilization toward availability, transparency and avoidance of outcome irrelevant utilization, this reprioritization is in progress.

Telemedical concepts use high speed networks to decouple availability of care from that of healthcare infrastructure and by design increase transparency, information quality and manageability, optimizing many subsequent processes [12].

In Landgraf et al. [1] we reported that medical non-professionals, through a telemedical concept, were non-inferior to medical professionals at performing the first step of emergency response. Distinguishingly, the non-professionals in this scenario being closer to the scene, their response could begin with less delay although it might take longer. Thus, this telemedical support exemplifies quality of information retrieval improvement achieved without additional healthcare infrastructure or local personnel.

For our special use case, after insubstantial technological improvements, this telemedical support seems applicable and may improve emergency response.

Interesting examples of subsequent investigations would be:

- How do these cohorts perform at different tasks if they are telemedically supported? In other words: What is the range of scenarios where medical non-professionals are non-inferior to medical professionals?
- How do these cohorts perform compared to different cohorts?
- What are limitations and challenges to transferability of these findings for real situations and how should we improve simulation fidelity?
- What are limitations and challenges to implementability of telemedical support concepts? Particularly, what are legal requirements for remote treatment[7] through bystanders?

In short, this concept needs to be tested under less extreme and less artificial boundary conditions, for different cohorts and scenarios. Additional aspects, implications and implementability of this concept require investigation. Findings therefrom could shape telemedical concepts for rural and otherwise remote settings. In a far vision, these concepts could build more confidently on supported non-professionals covering the last mile or first link of the chain of survival and perhaps even gaps in general and specialist healthcare everywhere.

---

[6] Depending on the framework of quality, time to treatment, access to expertise, regain of life skills and many more constitute a summative quality.

[7] Legalization of exceptional direct remote diagnosis and treatment was consented by the board of the German Medical Association (translation of "Vorstand Bundesärztekammer") in March 2018.

5 Conclusion

This synopsis is meant to supply a reader of Landgraf et al. [1] with additional perspectives on telemedicine and information retrieval, additional details of methods, findings and their applications and prospects.

"The investigated telemedical support of offshore engineers during the initial survey of a simulated emergency benefited guideline adherence and required additional time." [1] The performance by supported medical non-professionals was not inferior to that by unsupported medical professionals. In other words, emergency information retrieval and emergency information quality were improved by telemedical support.

Therefore, "first response to medical emergencies during construction and maintenance of offshore wind farms with substantially delayed professional care may be improved by telemedical support." [1]

Further research questions concern the generalizability and transfer of methods and findings.

Together with the original publication and its supplements, this work commences the establishment of simulator-based homogenization of intervention boundary conditions and scalable, video-based quantification of performance and inter-rater reliability among the supervisors' research groups.

Acknowledgments

I deeply appreciate and am grateful to Univ.-Prof. Dr. med. Claudia Spies, Prof. Dr. rer. nat. habil. Klaus-Dieter Wernecke and Dr. med. Torsten Schröder for invaluably supporting and guiding me throughout the publication, this thesis and my career. My special thanks are owed to Dr. Irmgard Landgraf, Dr. Luise Gaede and Dr. Brian Hemby for supporting and guiding me beyond this work.

References

1. Landgraf, P, Spies, C, Lawatscheck, R, Luz, M, Wernecke, KD, and Schröder, T. Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study. *BMJ Open* 2019;9(8):e027563. doi:`10.1136/bmjopen-2018-027563`.
2. Hanefeld, J, Powell-Jackson, T, and Balabanova, D. Understanding and measuring quality of care: dealing with complexity. *Bulletin of the World Health Organization* 2017;95(5):368. doi:`10.2471/BLT.16.179309`.
3. Pitschas, R. Medizinische Versorgung zwischen Recht und Ökonomie–Verfassungsrechtliche Grenzen der Priorisierung und Rationierung von Gesundheitsleistungen. In: *Privatrecht, Wirtschaftsrecht, Verfassungsrecht*. Nomos Verlagsgesellschaft mbH & Co. KG; 2015:930–939. doi:`10.5771/9783845264783-930`.
4. Stahmeyer, JT, Geyer, S, Epping, J, Tetzlaff, J, and Eberhard, S. Gesundheitsausgabenentwicklung und der Einfluss des demografischen Wandels. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz* 2018;61(4):432–441. doi:`10.1007/s00103-018-2713-3`.
5. Köbberling, J. Economic Pressure in Hospitals: Profound Disagreement over Priorities Between Doctors and Managers. *Deutsches Ärzteblatt International* 2017;114(47):795. doi:`10.3238/arztebl.2017.0795`.
6. Kopetsch, T. Dem deutschen Gesundheitswesen gehen die Ärzte aus. *Studie zur Altersstruktur und Arztzahlentwicklung* 2010;5:1–147.
7. Stuhr, M, Dethleff, D, Weinrich, N, Nielse n, M, Hory, D, Kowald, B, Seide, K, Kerner, T, Nau, C, and Jürgens, C. Notfallmedizinische versorgung in offshore-windparks. *Der Anaesthesist* 2016;65(5):369–379. doi:`10.1007/s00101-016-0154-7`.
8. Durstewitz, M and Lange, B. Meer–Wind–Strom: Forschung am ersten deutschen Offshore-Windpark alpha ventus. Springer-Verlag; 2016. doi:`10.1007/978-3-658-09783-7`.
9. Perednia, DA and Allen, A. Telemedicine technology and clinical applications. *Jama* 1995;273(6):483–488. doi:`10.1001/jama.1995.03520300057037`.
10. Hudson, HE and Parker, EB. Medical communication in Alaska by satellite. *New England Journal of Medicine* 1973;289(25):1351–1356. doi:`10.1056/NEJM197312202892506`.
11. Zachrison, KS, Boggs, KM, M Hayden, E, Espinola, JA, and Camargo, CA. A national survey of telemedicine use by US emergency departments. *Journal of telemedicine and telecare* 2018;:1357633X18816112doi:`10.1177/1357633X18816112`.
12. Culmer, N, Smith, T, Stager, C, Meyer, H, Quick, S, and Grimm, K. Evaluation of the triple aim of medicine in prehospital telemedicine: A systematic literature review. *Journal of telemedicine and telecare* 2019;:1357633X19853461doi:`10.1177/1357633X19853461`.
13. Jacobs, I, Callanan, V, Nichol, G, Valenzuela, T, Mason, P, Jaffe, AS, Landau, W, and Vetter, N. The chain of survival. *Annals of emergency medicine* 2001;37(4):S5–S16. doi:`10.1067/mem.2001.114176`.
14. Mafimoradi, S and Kabiri, N. Challenges and Opportunities of Telemedicine: A Narrative Review Study. *Health Information Management* 2019;15(6):294–299.
15. Batsis, JA, DiMilia, PR, Seo, LM, Fortuna, KL, Kennedy, MA, Blunt, HB, Bagley, PJ, Brooks, J, Brooks, E, Kim, SY, Masutani, RK, Bruce, ML, and Bartels, SJ. Effectiveness of Ambulatory Telemedicine Care in Older Adults: A Systematic Review. *Journal of the American Geriatrics Society* 2019;67(8):1737–1749. doi:`10.1111/jgs.15959`.
16. du Toit, M, Malau-Aduli, B, Vangaveti, V, Sabesan, S, and Ray, RA. Use of telehealth in the management of non-critical emergencies in rural or remote emergency departments: a systematic review. *Journal of telemedicine and telecare* 2019;25(1):3–16. doi:`10.1177/1357633X17734239`.
17. Perkins, GD, Olasveengen, TM, Maconochie, I, Soar, J, Wyllie, J, Greif, R, Lockey, A, Semeraro, F, Van de Voorde, P, Lott, C, Monsieurs, K, and Nolan, J. European resuscitation council guidelines for resuscitation: 2017 update. *Resuscitation* 2018;123:43–50. doi:`10.1016/j.resuscitation.2017.12.007`.
18. Stiell, A, Forster, AJ, Stiell, IG, and van Walraven, C. Prevalence of information gaps in the emergency department and the effect on patient outcomes. *CMAJ* 2003;169(10):1023–1028.
19. Mashoufi, M, Ayatollahi, H, and Khorasani-Zavareh, D. A review of data quality assessment in emergency medical services. *The open medical informatics journal* 2018;12:19. doi:`10.2174/1874431101812010019`.
20. Everson, J, Kocher, KE, and Adler-Milstein, J. Health information exchange associated with improved emergency department care through faster accessing of patient information from outside organizations. *Journal of the American Medical Informatics Association* 2016;24(e1):e103–e110. doi:`10.1093/jamia/ocw116`.
21. Janakiraman, R, Park, E, Demirezen, E, and Kumar, S. The effects of health information exchange access on healthcare quality and efficiency: An empirical investigation. *Mays Business School Research Paper* 2017;doi:`10.2139/ssrn.2915190`.
22. Menachemi, N, Rahurkar, S, Harle, CA, and Vest, JR. The benefits of health information exchange: an updated systematic review. *Journal of the American Medical Informatics Association* 2018;25(9):1259–1265. doi:`10.1093/jamia/ocy035`.
23. Bakke, HK, Steinvik, T, Eidissen, SI, Gilbert, M, and Wisborg, T. Bystander first aid in trauma–prevalence and quality: a prospective observational study. *Acta Anaesthesiologica Scandinavica* 2015;59(9):1187–1193. doi:`10.1111/aas.12561`.
24. Latifi, R, Hadeed, GJ, Rhee, P, O'Keeffe, T, Friese, RS, Wynne, JL, Ziemba, ML, and Judkins, D. Initial experiences and outcomes of telepresence in the management of trauma and emergency surgical patients. *The American Journal of Surgery* 2009;198(6):905–910. doi:`10.1016/j.amjsurg.2009.08.011`.
25. Lamé, G and Dixon-Woods, M. Using clinical simulation to study how to improve quality and safety in healthcare. *BMJ Simulation and Technology Enhanced Learning* 2018;:bmjstel–2018doi:`10.1136/bmjstel-2018-000370`.
26. Cook, DA, Hatala, R, Brydges, R, Zendejas, B, Szostek, JH, Wang, AT, Erwin, PJ, and Hamstra, SJ. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *Jama* 2011;306(9):978–988. doi:`10.1001/jama.2011.1234`.
27. Slakey, DP, Simms, ER, Rennie, KV, Garstka, ME, and Korndorffer Jr, JR. Using simulation to improve root cause analysis of adverse surgical outcomes. *International journal for quality in health care* 2014;26(2):144–150. doi:`10.1093/intqhc/mzu011`.
28. Dieckmann, P, Phero, JC, Issenberg, SB, Kardong-Edgren, S, Østergaard, D, and Ringsted, C. The first research consensus summit of the society for simulation in healthcare: conduction and a synthesis of the results. *Simulation in Healthcare* 2011;6(7):S1–S9. doi:`10.1097/SIH.0b013e31822238fc`.
29. Pucher, PH, Tamblyn, R, Boorman, D, Dixon-Woods, M, Donaldson, L, Draycott, T, Forster, A, Nadkarni, V, Power, C, Sevdalis, N, and Aggarwal, R. Simulation research to enhance patient safety and outcomes: recommendations of the simnovate patient safety domain group. *BMJ Simulation and Technology Enhanced Learning* 2017;3(Suppl 1):S3–S7. doi:`10.1136/bmjstel-2016-000173`.
30. Tun, JK, Alinier, G, Tang, J, and Kneebone, RL. Redefining simulation fidelity for healthcare education. *Simulation & Gaming* 2015;46(2):159–174. doi:`10.1177/1046878115576103`.
31. Landgraf, P, Spies, C, Lawatscheck, R, Luz, M, Wernecke, KD, and Schröder, T. Supplement A to: Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study. *BMJ Open* 2019;doi:`10.1136/bmjopen-2018-027563`.

32. Landgraf, P, Spies, C, Lawatscheck, R, Luz, M, Wernecke, KD, and Schröder, T. Supplement B to: Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study. *BMJ Open* 2019;doi:`10.1136/bmjopen-2018-027563`.
33. Chow, SC and Liu, JP. On assessment of bioequivalence under a higher-order crossover design. *Journal of biopharmaceutical statistics* 1992;2(2):239–256. doi:`10.1080/10543409208835042`.
34. Liu, JP and Weng, CS. Evaluation of parametric and nonparametric two one-sided tests procedures for assessing bioequivalence of average bioavailability. *Journal of biopharmaceutical statistics* 1993;3(1):85–102. doi:`10.1080/10543409308835050`.
35. Wellek, S. Testing statistical hypotheses of equivalence and noninferiority. Chapman and Hall/CRC; 2010.
36. Nichol, G, Cobb, LA, Yin, L, Maynard, C, Olsufka, M, Larsen, J, McCoy, AM, and Sayre, MR. Briefer activation time is associated with better outcomes after out-of-hospital cardiac arrest. *Resuscitation* 2016;107:139–144. doi:`10.1016/j.resuscitation.2016.06.040`.
37. Østerås, Ø, Heltne, JK, Vikenes, BC, Assmus, J, and Brattebø, G. Factors influencing on-scene time in a rural norwegian helicopter emergency medical service: a retrospective observational study. *Scandinavian journal of trauma, resuscitation and emergency medicine* 2017;25(1):97. doi:`10.1186/s13049-017-0442-5`.
38. Sejersten, M, Sillesen, M, Hansen, P, Nielsen, S, Nielsen, H, Trautner, S, Hampton, D, Wagner, G, and Clemmensen, P. Effect on Treatment Delay of Prehospital Teletransmission of 12-Lead Electrocardiogram to a Cardiologist for Immediate Triage and Direct Referral of Patients With ST-Segment Elevation Acute Myocardial Infarction to Primary Percutaneous Coronary Intervention. *Am J Cardiol* 2008;101(7):941–946. doi:`10.1016/j.amjcard.2007.11.038`.
39. Demaerschalk, B. Telemedicine or telephone consultation in patients with acute stroke. *Curr Neurol Neurosci* 2011;11(1):42–51. doi:`10.1007/s11910-010-0147-x`.
40. Sanchez-Ross, M, Oghlakian, G, Maher, J, Patel, B, Mazza, V, Hom, D, Dhruva, V, Langley, D, Palmaro, J, Ahmed, S, Kaluski, E, and Klapholz, M. The STAT-MI (ST-Segment Analysis Using Wireless Technology in Acute Myocardial Infarction) Trial Improves Outcomes. *JACC: Cardiovascular Interventions* 2011;4(2):222–227. doi:`10.1016/j.jcin.2010.11.007`.
41. Julious, SA. Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 2005;4(4):287–291. doi:`10.1002/pst.185`.
42. Loos, S, Albrecht, M, and Zich, K. Zukunftsfähige Krankenhausversorgung - Simulation und Analyse einer Neustrukturierung der Krankenhausversorgung am Beispiel einer Versorgungsregion in Nordrhein-Westfalen. *Bertelsmann Stiftung* 2019;doi:`10.11586/2019042`.

Appendix

Eidesstattliche Versicherung

Ich, Philipp Landgraf, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Effects of Telemedical Support on Quality of Emergency Information Retrieval Considering Offshore Wind Power Infrastructure“ (übersetzt „Effekte von telemedizinischer Unterstützung auf die Qualität von Notfallinformationsgewinnung unter Berücksichtigung von Offshore Windkraftinfrastruktur“) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE - www.icmje.org) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM und werden von mir verantwortet.

Mein Anteil an etwaigen Publiikationen zu dieser Dissertation entspricht dem, der in der gemeinsamen Erklärung mit dem/der Betreuer/in angegeben ist. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM und werden von mir verantwortet. Ich erkläre ferner, dass mir die Satzung der Charité Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.

Unterschrift des Doktoranden,
Berlin, 10/02/19

Ausführliche Anteilserklärung

Philipp Landgraf hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Landgraf, P, Spies, C, Lawatscheck, R, Luz, M, Wernecke, KD, and Schröder, T. Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study. BMJ Open 2019;9(8):e027563. doi:10.1136/bmjopen-2018-027563
Beitrag im Einzelnen: Der Doktorand hat die Initiative ergriffen, die Idee selbstständig weiterentwickelt und in Diskussion mit seinen Betreuern das Konzept entwickelt. Er hat die Datenverarbeitung und Moderation vollständig selbst, die Akkumulation, Exploration und Validierung mit sehr wenig und die biometrische Analyse mit wenig Unterstützung durch seine Koautoren geleistet. Er hat die Diskussion mit Koautoren und Peers auf Kongressen überwiegend selbst geführt und den Peer Review Prozess überwiegend selbstständig durchlaufen. Er hat alle Visualisierungen und Tabellen selbst erstellt, Feedback darüber eingeholt und entsprechende Weiterentwicklungen selbst vorgenommen. Er hat alle Versionen der Manuskripte vollständig selbst erstellt und im kritischen Diskurs weiterentwickelt.

Unterschrift des Doktoranden,
Berlin, 10/02/19

Unterschrift und Stempel der
erstbetreuenden Hochschullehrerin,
Berlin, 10/02/19

Journal Ranking

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI
Selected Categories: **“MEDICINE, GENERAL and INTERNAL”**
Selected Category Scheme: WoS
Gesamtanzahl: 160 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NEW ENGLAND JOURNAL OF MEDICINE	344,581	70.670	0.686700
2	LANCET	247,292	59.102	0.427870
3	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	156,350	51.273	0.300810
4	Nature Reviews Disease Primers	4,339	32.274	0.019740
5	BMJ-British Medical Journal	112,901	27.604	0.152760
6	JAMA Internal Medicine	15,215	20.768	0.095580
7	ANNALS OF INTERNAL MEDICINE	57,057	19.315	0.096020
8	PLOS MEDICINE	30,689	11.048	0.071200
9	Journal of Cachexia Sarcopenia and Muscle	2,799	10.754	0.005870
10	BMC Medicine	13,630	8.285	0.045220
11	Cochrane Database of Systematic Reviews	67,607	7.755	0.158690
12	MAYO CLINIC PROCEEDINGS	14,695	7.091	0.025750
13	CANADIAN MEDICAL ASSOCIATION JOURNAL	15,351	6.938	0.016500
14	JOURNAL OF INTERNAL MEDICINE	10,547	6.051	0.015700
15	Journal of Clinical Medicine	2,315	5.688	0.007210
16	MEDICAL JOURNAL OF AUSTRALIA	11,134	5.332	0.012600
17	PALLIATIVE MEDICINE	5,682	4.956	0.009860
18	AMYLOID-JOURNAL OF PROTEIN FOLDING DISORDERS	1,335	4.919	0.003270

1

Selected JCR Year: 2018; Selected Categories: “MEDICINE, GENERAL und INTERNAL”

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
19	Translational Research	3,669	4.915	0.008530
20	AMERICAN JOURNAL OF MEDICINE	25,051	4.760	0.026650
21	JOURNAL OF GENERAL INTERNAL MEDICINE	19,431	4.606	0.028130
22	Deutsches Arzteblatt International	4,331	4.469	0.007630
23	AMERICAN JOURNAL OF PREVENTIVE MEDICINE	22,339	4.435	0.041750
24	BRITISH JOURNAL OF GENERAL PRACTICE	6,489	4.434	0.009370
25	ANNALS OF FAMILY MEDICINE	5,314	4.185	0.010880
26	JOURNAL OF TRAVEL MEDICINE	2,229	4.155	0.003410
27	European Journal of Internal Medicine	4,559	3.660	0.009420
28	JOURNAL OF THE ROYAL SOCIETY OF MEDICINE	4,051	3.538	0.003080
29	AMERICAN JOURNAL OF CHINESE MEDICINE	3,152	3.510	0.002950
30	PREVENTIVE MEDICINE	16,004	3.449	0.029820
31	JOURNAL OF PAIN AND SYMPTOM MANAGEMENT	11,229	3.378	0.015750
32	Frontiers in Medicine	1,598	3.113	0.005060
33	ANNALS OF MEDICINE	4,437	3.049	0.005440
34	Polish Archives of Internal Medicine-Polskie Archiwum Medycyny Wewnetrznej	1,287	2.882	0.002000
35	JOURNAL OF THE FORMOSAN MEDICAL ASSOCIATION	3,274	2.844	0.004420
36	BRITISH MEDICAL BULLETIN	4,435	2.804	0.003670
37	EUROPEAN JOURNAL OF CLINICAL INVESTIGATION	5,946	2.784	0.006180

Selected JCR Year: 2018; Selected Categories: “MEDICINE, GENERAL und INTERNAL”

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
38	PAIN MEDICINE	6,988	2.758	0.013020
39	UPSALA JOURNAL OF MEDICAL SCIENCES	1,109	2.747	0.001980
40	MEDICAL CLINICS OF NORTH AMERICA	2,990	2.716	0.004280
41	KOREAN JOURNAL OF INTERNAL MEDICINE	1,887	2.714	0.003600
42	QJM-AN INTERNATIONAL JOURNAL OF MEDICINE	5,711	2.649	0.004350
43	INTERNATIONAL JOURNAL OF CLINICAL PRACTICE	5,439	2.613	0.005980
44	AMERICAN FAMILY PHYSICIAN	6,860	2.580	0.005110
45	Journal of the American Board of Family Medicine	3,654	2.511	0.006540
46	Diagnostics	533	2.489	0.001350
47	MINERVA MEDICA	817	2.475	0.000950
48	BMC Family Practice	4,209	2.431	0.009370
49	Archives of Medical Science	2,581	2.380	0.005110
> 50	BMJ Open	26,298	2.376	0.108600
51	CURRENT MEDICAL RESEARCH AND OPINION	7,181	2.345	0.010810
52	Internal and Emergency Medicine	1,978	2.335	0.004390
53	International Journal of Medical Sciences	3,358	2.333	0.006330
54	Journal of Hospital Medicine	3,068	2.276	0.009200
55	POSTGRADUATE MEDICINE	2,311	2.237	0.003640
56	CANADIAN FAMILY PHYSICIAN	3,761	2.186	0.005200
56	PANMINERVA MEDICA	653	2.186	0.000620

Selected JCR Year: 2018; Selected Categories: “MEDICINE, GENERAL und INTERNAL”

Originalbeitrag [1]

Open access Research

BMJ Open

Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study

Philipp Landgraf,[1] Claudia Spies,[1] Robert Lawatscheck,[1,2] Maria Luz,[3]
Klaus-Dieter Wernecke,[4] Torsten Schröder[1]

To cite: Landgraf P, Spies C, Lawatscheck R, *et al*. Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study. *BMJ Open* 2019;**9**:e027563. doi:10.1136/bmjopen-2018-027563

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-027563).

Other parts of the results of the 'SOS - Sea and Offshore Safety' project funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), funding code BMU-41V6169, have been presented at: RAVE Konsortial meeting 2012 Hamburg (RL), RAVE Conference 2012 Bremerhaven (RL), renewable UK 2013 Sheffield (RL), Rechts symposium 2013 Hamburg (TS), DIVI Workshop 2013 Hamburg (RL). This workin progress has been discussed at GMDS annual conference 2017 (PL) and conhIT2018 (PL).

Received 13 November 2018
Revised 24 May 2019
Accepted 17 July 2019

For numbered affiliations see end of article.

Correspondence to
Dr Torsten Schröder;
torsten.schroeder@charite.de

ABSTRACT

Objective To investigate, in a simulator-based prospective study, whether telemedical support improves quality of emergency first response (performance) by medical non-professionals to being non-inferior to medical professionals.
Setting In a simulated offshore wind power plant, duos (teams) of offshore engineers and teams of paramedics conducted the primary survey of a simulated patient.
Participants 38 offshore engineers and 34 paramedics were recruited by the general email invitation.
Intervention Teams (randomised by lot) were supported by transmission technology and a remote emergency physician in Berlin.
Outcome measures From video recordings, performance (17 item checklist) and required time (up to 15 min) were quantified by expert rating for analysis. Differences were analysed using two-sided exact Mann-Whitney U tests for independent measures, non-inferiority was analysed using Schuirmann one-sided test. The significance level of 5 % was Holm-Bonferroni adjusted in each family of pairwise comparisons.
Results Nine teams of engineers with, nine without, nine teams of paramedics with and eight without support completed the task. Two experts quantified endpoints, insights into rater dependence were gained. Supported engineers outperformed unsupported engineers ($p<0.01$), insufficient evidence was found for paramedics ($p=0.11$). Without support, paramedics outperformed engineers ($p<0.01$). Supported engineers' performance was non-inferior (at one item margin) to that by unsupported paramedics ($p=0.03$). Supported groups were slower than unsupported groups ($p<0.01$).
Conclusions First response to medical emergencies in offshore wind farms with substantially delayed professional care may be improved by telemedical support. Future work should test our result during additional scenarios and explore interdisciplinary and ecosystem aspects of this support.
Trial registration number DRKS00014372

Strengths and limitations of this study

- This telemedical intervention applied in a simulated offshore emergency scenario was feasible.
- The chosen methods were adequate in resolving the hypothesised effects. Despite improved validity and reliability, generalisability of our findings may be limited.
- Telemedical support improved layman performance to being non-inferior to that by paramedics with relevance for current medical service challenges.

INTRODUCTION

Context

During construction and maintenance of offshore wind farms, medical support has to be available to teams as small as three maintenance engineers.[1] Emergency rescue was performed in 70 of 319 medical support cases in four German offshore wind farms between 2008 and 2012. Most rescues were triggered by accidents and respiratory or intestinal illness.[2] Medical response teams may take more than 90 min to arrive due to distance (up to 200 km), weather and structural inaccessibility.[1,3] Delay of necessary treatment or rescue due to misjudgement may plausibly result in lasting damage or loss of life for patients and psychological trauma for first responders or unnecessary risk and costs for the rescue teams.[4] Furthermore, in both scenarios 'rescue required and arriving quickly' and 'low urgency, relief under way', immediate assessment and treatment may be warranted[2,5] and reduce on-scene time of medical services.[6]

Telemedicine

Various teleconsultations between medical professionals were reported in recent years to benefit quality of care.[7–11]

Telemedical support may already be provided by the medical professionals to non-professionals off shore,[12] but the effects on quality of care by non-professionals are only clear for cardiopulmonary resuscitation.[13 14] We hypothesised that real-time teleconsultations for non-professional first responders in an emergency scenario improves guideline adherence and time to treatment surrogative of quality of care. Under the assumption of this scenario being commonplace for professionals, we used this group's performance as a benchmark.

METHODS AND MATERIALS

Summary

In a simulated scenario, two participants were to survey vital data of a simulated polytraumatised colleague. The intervention groups (half of the non-professional and professional teams) had a telemedically enabled vital data monitoring unit (corpuls-3 with corpuls.web, GS Elektromedizinische Geräte G. Stemple, Kaufering, Germany) and were supported by a remote physician by means of a communication device (Frontline Communicator, MDAI mobile solutions, Wolfratshausen, Germany). Guideline adherence in terms of the number of correctly performed checklist items and, as indicative of time prior to treatment, required time for the initial survey were adjudicated using video recordings of the scenario.

Patient and public involvement

Offshore engineers were involved in the study design, scenario validation, recruited as participants, debriefed and interviewed about their participation and for feedback on the study design and telemedical intervention.

Subject pool

To compare the effect of the intervention on groups with different medical expertise, after receiving positive ethics committee vote (Ethikkommission—Ethikausschuss 1 am Campus Charité Mitte, EA1/181/13), offshore maintenance engineers working with the industrial partner (EWE Energie AG, Oldenburg, Germany) and paramedics working with the Berlin Fire Department were recruited through a general email invitation. Responders were eligible for participation if they gave personal informed consent to participate, confirmed they were healthy, fit, proficient in German, understood that they would be videotaped and evaluated. The engineers were paired up into teams with at least one member with offshore emergency response officer (50 hours) or higher medical training. These teams were randomised by lot into intervention and control groups. The paramedics were likewise paired up into teams and the teams likewise randomised by lot into intervention and control groups. All participants were familiarised with devices (particularly simulators) to be used ahead of time.

Sample size and power calculations

Without knowledge of the feasibility and acceptability of the intervention, design, procedures, measurability of effects and estimation of precision, the calculation of necessary sample size was not possible. The number of participants was, therefore, generally justified by practical needs of the project, acknowledging the interindividual and intraindividual variability as well as psychosociological processes under described circumstances. From the data this exploration yielded, the post hoc power was calculated (see online supplementary A) to advise future research.

Experimental set-up

The paramedics participated in a seminar room at Charité in Berlin and the engineers in a seminar room in Oldenburg.

A standard emergency case[15] was available within the simulation area. It contained among other items gloves, stethoscope, blood pressure cuff and diagnostic penlight.

As shown in figures 1 and 2, supported teams were additionally given a monitoring device to survey noninvasive blood pressure, blood oxygen saturation and ECG (corpuls-3) and a two-way audio and one-way video transmission device (Frontline Communicator). Both devices connected wirelessly to the remote physician in an office at Charité. The remote physician had at his disposal a checklist containing: the items in table 1, reminders to advise the teams to use gloves and advice on finding non-invasive blood pressure cuff and stethoscope inside the emergency case.

Five anaesthesiologists with expertise in prehospital emergency medicine and in conducting simulation workshops volunteered as remote physicians at Charité. They telemedically supported intervention groups.

A clothed human patient simulator (Emergency Care Simulator, CAE Healthcare, Mainz, Germany) lay on the floor inside the simulation area. Two of the authors took turns operating the simulator, using a microphone to speak as though from the simulator's mouth. This simulator had controllable eye lids, interfaced with arbitrary peripheral oxygen saturation (SpO_2) clips, emulated 5 channel ECG, pupillary light reflex, radial and carotid pulses, chest excursion and breathing sounds. A phantom fracture was fixed to its arm and a phantom burn to its lower leg.

Two perspectives on the simulation area and one on the remote physician were video recorded for adjudication by two experts in prehospital emergency medicine.

Scenario: Retrieve vital data

The team was described to be on maintenance mission in a specific offshore wind park, inside a power plant marked by the simulation area. A colleague among the team of three (the human patient simulator) had fallen from 4 m height. To increase immersion, the simulator kept moaning and calling for pain relief.

Each team of two participants was given the task of only collecting vital signs, patient history and 'anything that might help a medic'.

Scenario validation

The representative scenario designers (two of the authors) conducted unstructured literature reviews on most common emergencies and causes of death in offshore settings, unstructuredly interviewed offshore maintenance

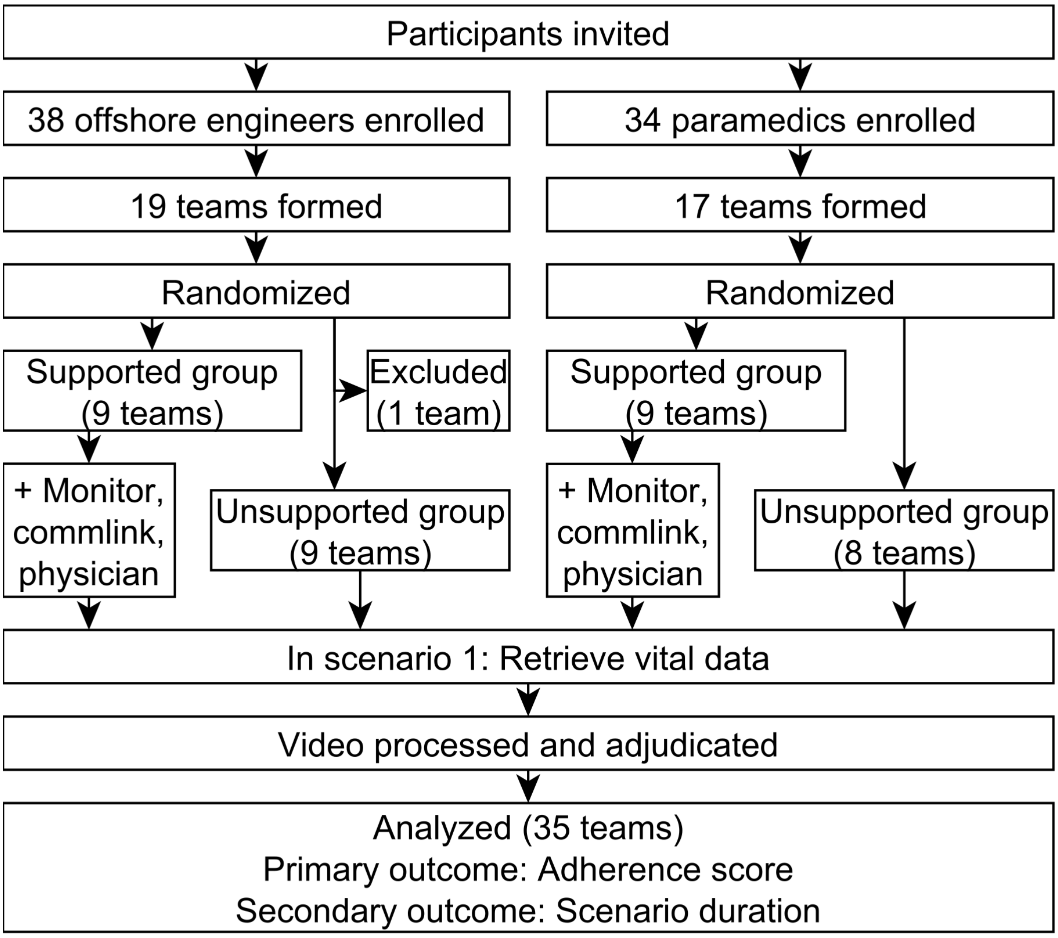

Figure 1 Flow diagram of the simulator study. Randomisation into groups, half of them supported by a remote physician via a telemedically enabled monitoring unit and bidirectional commlink. One team dropped out due to illness.

engineers and wind power plant operators and discussed typical emergencies,[2] response and treatment with the typical rescuers. These typical rescuers were paramedics working at the offshore rescue headquarters and physicians working at the receiving hospital department. Moreover, both of them visited an onshore wind power plant, attained helicopter hoist certification and survival at sea certification and visited an offshore wind power plant by helicopter hoist.

All offshore wind power plant construction company representatives at the consortium meetings informally confirmed the validity of the experimental set-up.

Supervision

Every participant and remote physician was given a briefing in writing and a 2-hour introduction to the technology to be used. Additionally, a moderator recounted for each team the scenario and task using scripted sentences at the beginning of the scenario. Any side-tracking due to problems with the telemedical equipment was dealt with in technical timeouts. Any action beyond the given task was halted to be performed in a later scenario.

Data acquisition

A checklist (see table 1) of correct–incorrect–unclear items was preliminarily designed to operationalise adherence to European Resuscitation Council (ERC) guidelines[16] by two anaesthesiologists (two of the authors). Notably, the Airway, Breathing, Circulation, Disability, Exposure approach therein is applicable for all emergencies. A total of 14 out of 17 items (82 %) were asked directly after a team claimed to have collected all relevant information. Supported teams were to confer with their remote physician, the combined knowledge being the endpoint. Perspective-synchronised recordings of the scenario, remote physician and questioning constituted the data for review. One of the authors and an independent peer adjudicated all scenarios, classifying items as correctly/incorrectly performed or unclear and measured scenario duration without timeouts.

Metrics validation

In operationalising the ERC guidelines, all applicable Good ResearArch for Comparative Effectiveness principles (grace-principles.org) were considered. The checklist reliability

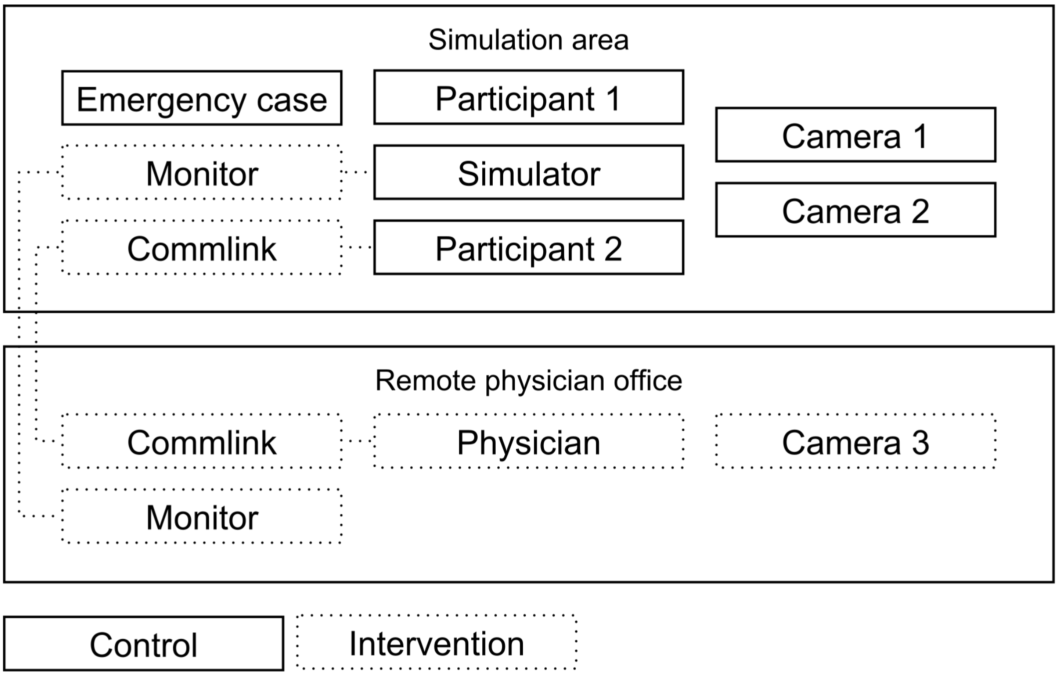

Figure 2 Schematic experimental set-up for control (continuous line) and intervention groups (dotted line): participants and simulator in the simulation area were video recorded from two angles, the remote physician from one angle. All groups had a standard emergency case but intervention groups additionally had audio–video communication (commlink) with and vital data transmission (monitor) to a remote physician.

and validity were probed in two test simulations. No efforts were undertaken to validate the duration metric.

Under the assumption that professional first responders adhere to guidelines well and deliver high quality of care, the checklist item scoring range is calibrated to resolve group differences. Differences in required time are expected to be clearly resolved and will not be scaled.

Data validation

Inter-rater reliability was estimated using average intraclass correlation coefficients (ICC) in a two-way random-effects model for absolute agreement (ICC (2,2)), 95 % CIs are reported ([·,·]). Quantile-quantile plots and minima in proportionate positive inter-rater agreement were explored. Please see online supplementary B for details.

Data analysis

Average rater rating of number of items correctly performed per team (itemised performance) was analysed with two-sided exact Mann-Whitney U test for independent measures (2sU) comparing

1. Supported against unsupported engineers and supported against unsupported paramedics.
2. Unsupported engineers against unsupported paramedics.
3. Supported engineers against supported paramedics.
4. Supported engineers against unsupported paramedics.

Additionally,

5. Non-inferiority of the performance by supported engineers compared with unsupported paramedics was analysed using Schuirmann one-sided test at 5 % significance with a one-item margin (SOST).

Average rater measurement of scenario duration was evaluated using 2sU comparing

6. Supported against unsupported engineers and supported against unsupported paramedics.
7. Unsupported engineers against unsupported paramedics.
8. Supported engineers against supported paramedics.
9. Supported engineers against unsupported paramedics.

Finally,

10. The number of teams per group who, by average rating, correctly performed each item was analysed descriptively.

Medians (Md), 95 % CIs of Md ([·,·]), absolute z-value (z), 1-/2-sided exact significance (p) and effect size (d) are reported.

The significance level of 5 % was Holm-Bonferroni adjusted in each family of pairwise comparisons: Performance, non-inferiority and duration. Adjusted significance is denoted (*).

All tests were conducted using SPSS V.24 (IBM) where possible and EquivTest (StatCon, Witzenhausen, Germany) otherwise.

RESULTS

After drop-outs due to illness, 36 off-shore maintenance engineers (medical non-professionals, 32 male and 4

Table 1 Checklist of itemised performance quality attributes as used by the remote physician and for evaluation per team and scenario

Index	Performance item	Description of correct result
1	Patient conscious	Yes
2	Patient orientated	At least one information retrieved: name, location, time
3	Glasgow Coma Scale	14–15
4	Does the patient breathe normally?	No
5	Respiratory rate	18–22 breaths per minute
6*	Respiratory rate correctly surveyed	Counted for 20 s or using monitoring device
7	Radial and if not carotid pulse	Both pulses perceptible
8*	Radial pulse correctly surveyed	By palpation
9*	Blood pressure correctly measured	Unsupported: Cuff placed correctly, stethoscope used in Riva-Rocci fashion Supported: Cuff placed correctly
10	Blood pressure	Unsupported: Korotkoff sound inaudible Supported: 110/60 mm Hg using monitoring device
11	Blood pressure normal	Unsupported: Unclear evidence Supported: Yes
12	Symptoms	Patient was asked
13	Allergies	Patient was asked
14	Medication	Patient was asked
15	Medical history	Patient was asked
16	Last meal	Patient was asked
17	Environment/course of accident	Fall from 4 m height after burning lower leg

They are ordered according to the sequence the authors would perform them in.
*Item not asked at the end of the scenario but observed in the video material.

female, aged Md = 41, IQR = [32, 47]) and 34 paramedics (medical professionals, 30 male and 4 female, aged Md = 36, IQR = [27, 45]) participated in the study. All engineers had taken at least first aid (16 hours) training refreshed Md = 12, IQR = [9, 16] months prior. All paramedics had at least 2 years of work experience and regular first aid training. Figure 1 illustrates the study design with final sample sizes.

Inter-rater reliability was estimated for performance per team $ICC_{ppt}(2,2) = 0.71$ [−0.10, 0.90], required time per team $ICC_{tpt}(2,2) = 0.99$ [0.98, 0.99] and performance per item $ICC_{ppi}(2,2) = 0.66$ [0.52, 0.74].

Scaling the item–score relationship to 100 % (a difference of one item translates to a difference by one performance score) resolves group differences sufficiently.

Effects of both experimental conditions on performance rating and time in minutes required to complete the scenario (average across raters) are condensed in figure 3 and shown in figure 4.

a. Supported outperformed unsupported participants. This intervention effect was statistically significant among non-professionals but not among professionals.
b. Performance differences due to medical expertise is statistically significant in unsupported groups; professionals correctly performed twice as many items as non-professionals.
c. With telemedical support, this difference was not statistically significant.
d. No statistical evidence was found that unsupported professionals outperformed supported non-professionals.
e. Moreover, SOST rejects the null hypothesis of non-equivalence. Thus, non-inferiority of supported non-professionals' to unsupported professionals' performance within the specified equivalence bounds (difference of means less than or equal one item correctly performed) can be claimed.
f. Participants who were supported required significantly more time compared with unsupported participants.
g. Without telemedical support, no statistical evidence was found for either groupbeing faster.
h. In supported groups, non-professionals took more time.
i. Supported non-professionals required more time than unsupported professionals.
j. Non-professionals without support likely surveyed consciousness (item 1), pulse rate (8), symptoms (12) and environment (17) correctly (see figure 5). Professionals in the intervention group more often correctly surveyed respiration (4–6) and food intake (16) but less often manually surveyed radial pulse (7-8).

	Ref.	unsupported		supported		Test	Statistic	Effect Size	p-value
		non- ($n = 9$)	prof. ($n = 8$)	non- ($n = 9$)	prof. ($n = 9$)				
Md		4.50 [3.50, 7.50]	11.75 [9.25, 12.50]	11.50 [10.50, 13.50]	13.00 [12.00, 14.00]				
Performance	a	X		X		2sU	3.59	3.13	< .01∗
			X		X	2sU	1.65	.86	.11
	b	X	X			2sU	3.09	2.25	< .01∗
	c			X	X	2sU	1.11	.54	.29
	d		X	X		2sU	.78	.38	.46
	e		X	X		SOST	1.98	(a)	.03∗
Md		6.55 [4.38, 8.55]	4.38 [3.92, 7.08]	11.87 [9.25, 13.05]	7.62 [7.05, 9.50]				
Required time	f	X		X		2sU	2.96	1.95	< .01∗
			X		X	2sU	2.60	1.62	< .01∗
	g	X	X			2sU	1.44	.75	.16
	h			X	X	2sU	2.43	1.40	.01∗
	i		X	X		2sU	3.47	3.10	< .01∗

Figure 3 Median (Md) performance ratings and required time in minutes with 95 % CIs of Md with results for performance rating tests A–E and required time tests F–I rounded to two decimal places. X indicates tested pairs of groups (size indicated by n) of unsupported (control) and supported (intervention) engineers (medical non-professionals) and paramedics (medical professionals). Tests are two-sided exact Mann-Whitney U test for independent measures (2sU) or Schuirmann one-sided test at 5 % significance with a one-item margin (SOST). Typical denotation of the statistic of 2sU tests is *z*, that of SOST is t and typical denotation of effect size is d. Adjusted significance (p value) is indicated by ∗. (a) Note that SOST tests for an effect being smaller than the margin.

DISCUSSION

In a crucial information collection phase of emergency care, the telemedically supported non-professionals performed more items correctly than unsupported non-professionals (see figure 4). The authors hypothesise that, because the scenario was subjectively immersive and the live video feed subjectively very helpful, this effect would also occur if real medical emergencies were similarly supported. Supported non-professionals' performance was non-inferior to that by unsupported professionals. As can be seen in figure 4, supported non-professionals performed slightly more items correctly than unsupported professionals. This could be life-saving in environments of no or substantially delayed professional medical response.[17]

Among professionals, less impact on performance was observed (see figure 4). This confirms our assumption of this phase of emergency care being commonplace.

The performance of additional items required additional time (see figure 4). While in onshore scenarios (eg, cardiac arrest in King Country, Washington, USA excluding Seattle) professional help would probably have arrived during the scenario duration,[18] helicopter emergency medical services in rural scenarios would not (~30 min to arrival).[19 20] In remote cases, the required time does not necessarily delay rescue. On the contrary, more accurate first response could expedite triggering the appropriate rescue chain, which is independently associated with greater survival[21] and plausibly reduce on-scene time.[6] During a support gap pertaining to the remoteness and accessibility of offshore wind farms, even slow information collection may significantly improve outcome if it is reliable and available to alert and involve relevant specialists especially before arrival of a rescue team.[8 22 23] Non-professionals need to and, as our results

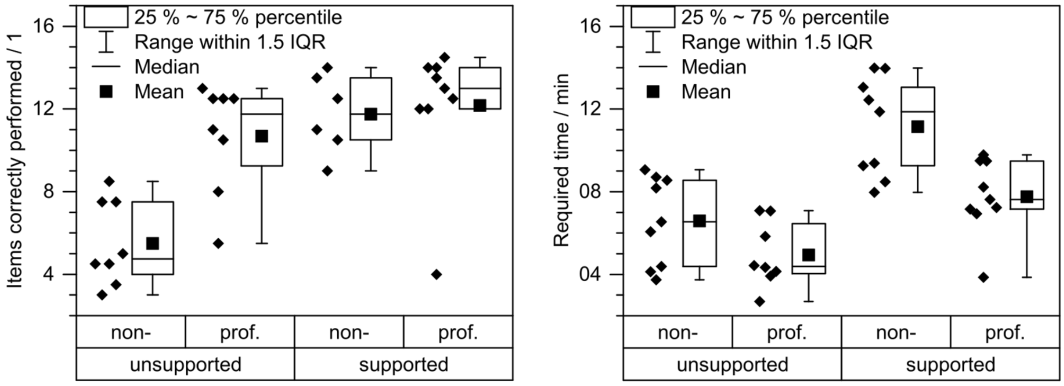

Figure 4 Itemised performance rating and time required by a team to finish the scenario: unsupported (control) and telemedically supported (intervention) groups with breakdown into medical non-professionals (non-) and professionals (prof.).

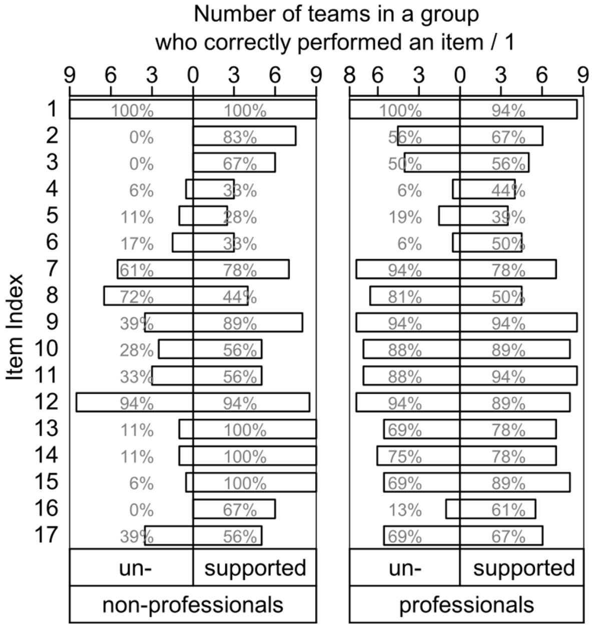

Figure 5 Distribution of number of teams who correctly performed an item (description in table 1) averaged across raters. Medical non-professionals versus professionals with breakdown into unsupported (control) and telemedically supported (intervention) groups.

show, could be able to bridge this gap with telemedical support.

Neither participants nor remote physicians were experienced in telemedicine. The authors suppose, after reviewing frequent misunderstandings, that training and hands-on experience would optimise communication, reducing the increase in required time.

The observed benefit in information completeness (see figure 5) may be crucial.[24] In agreement, Rörtgen *et al*[25] found significantly increased completeness only of survey of symptoms, allergies and medication from the SAMPLE mnemonic among physician-paramedic teams (EMS teams) in supported groups. The same research group reported in Skorning *et al*[26] that EMS teams with telemedical support significantly outperformed those without support in surveying allergies from the SAMPLE mnemonic. Moreover, analgosedation prior to cardioversion, synchronicity of shocks and adequate medication for intubation were improved by telemedical support. Cathlabs were significantly more often and trauma centres more quickly informed.

Unsupported non-professionals correctly surveyed consciousness, heart rate, symptoms and environment correctly (see figure 5). This may plausibly be due to the first aid or emergency response officer course all but two non-professionals had taken at most 2 years (Md = 12, IQR = [9, 16] months) prior. With telemedical support, surveys by non-professionals and professionals were more complete excepting only the notably low score in breathing related items 4–6 and item 8 (radial pulse palpated). In agreement, real-time provider feedback alone during surgery seems to improve checklist compliance.[27]

The low scores in breathing related items 4–6 may in part be due to raters' different interpretation of deviation from actually simulated vital data and from guidelines. These different interpretations manifest in correlation coefficients for performance per item being lower than for performance or time per team and are therefore discussed in depth in online supplementary B. The authors hypothesise that remaining room for improvement is a compound need-for-training and remote physician effect requiring a larger sample to be explored.

Item 8 on radial pulse palpation having been rated less often correctly performed for supported groups seems to stem from only one rater judging the correct result (pulse rate) integral to the correct measurement (see online supplementary B). However, with support, most teams used the networked monitoring device to attain a pulse rate result.

Our results imply that telemedically supporting non-professionals could be more than a fallback option: It could viably expand the continuum of care beyond the reach of specialised staff.

Limitations

Like Branzetti *et al*,[28] we designed a checklist ourselves. Unavoidably, our own choice of scenario, metric and data validation efforts may be biassed.

Results concerning offshore engineers may not be representative for non-professionals in general; offshore engineers are highly trained and relatively proficient in emergency first response, following technical instructions and manual tasks.

We did not isolate the effect of variations in the intervention design (such as only monitoring device vs only communication device vs both, different remote physicians or different communication strategies) for lack of sample size. We know only that a visual reminder alone would be unlikely to achieve the observed effect on quality of care. The impact of visual reminders on guideline adherence (quality of care), displayed on a phone during cardiac arrest first response, was not statistically significant while the impact by medical expertise was.[29] An audio–video instruction on a phone or big screen seems to improve at least some aspects of resuscitation performance[30–32] and trained laymen phone support outside of emergencies seems to improve only some risk behaviour.[33]

Moreover, we resolved neither intrateam effects, nor team versus individual performance.

The non-inferiority margin of one item was chosen after the scenarios had been conducted but before scenario recordings had been processed based on medical expertise. It should have been selected prior to possible bias by trial experience, which had not been considered.

The simulator operator, facilitator, remote physician and video adjudicator roles were assigned based on availability, not randomisation. No one person could have taken on more than one of these roles at a time and bias in the results was minimised by averaging with an independent review (see online supplementary B on data validation). Random assignment of independent persons without any rotation into one of the other roles would have been preferable but was not feasible in our hospital context.

Remote physician unfamiliarity with an uncommon scenario would have been realistic.[34] On the other hand, our scenario was common (as the small impact of our intervention on medical professionals confirms). Moreover, detailed checklists for both common and less common scenarios would also be realistic.

In Skorning *et al*,[26] 'two faculty investigators using (…) scoring items (…) had to reach common decisions'. In contrast, we used ratings by one of the authors and an independent peer to probe reliability, learn about df and reduce bias (see online supplementary B). The quality of care under investigation, however, is distorted by the simulation setting, the scenario artificiality, the recording modality and the projection onto items. During the transformation from input (care) to output (scores) via video recording and adjudication, fit between video and item confounded the relationship between quality of care and score. The choice of experts with similar backgrounds plausibly reduced but failed to eliminate this factor fully (see online supplementary B). Training experts to a level of agreement first, as done by Branzetti *et al*,[28] seems to be a sensible effort at reduction of room for interpretation.

The scenario has been designed as a compromise between investigative goals and safety as is known to serve training purposes well.[28] The offshore boundary conditions restricted space, limited resources, fibre internet, medical devices and supplies and standard emergency case were reproduced based on first-hand observation by the scenario designers and are also reported by Stuhr *et al*.[2 3] Most imaginable offshore situational conditions, such as high stress levels, cumbersome clothing, extreme weather, blood everywhere and emotional involvement, were not or not entirely reproduced, as they were expected to add inhomogeneous effects requiring additional power for isolation. To confirm homogeneity of immersion, the perception of the simulation environment should have been rigorously investigated but was not, because simulation of the kind employed by us has been found to be a sufficiently valid surrogate for scenarios otherwise impossible to observe consistently.[35 36] Blinding was not feasible. Setting effects and expectations may have distorted the results. For the non-professionals who are luckily unlikely to have ever encountered a polytraumatised patient but also for the professionals who working in Berlin are unlikely to have first-hand experience of an offshore wind power plant.

Perspectives

The chosen offshore scenario is distinguished from more common scenarios, such as in urban environments, by magnitude of consequences of rescue decisions and of the supply gap. Performance difference in medically advanced scenarios and feasible equipment scope are relevant subsequent questions. Considering the prevalence of hypoxia prior to intubation by emergency medical services reported by Sunde *et al*,[37] the disagreement as well as the imperfect scores by supported teams in breathing-related items are important findings to be investigated closely. The unsupported versus supported decision-making process in the context of risk for the patient, risk for the rescue teams and of potential health economic outcomes are of great interest.[38 39] Particularly, first aid response rate among offshore personnel[40 41] may be increased. The technological implementation (Such as [12] or ghc-tech.de.) and business models (Such as gmn-bremen.de.) are topics of investigation. While support of medical professionals is gaining momentum, emerging onshore telematics infrastructure, assessment and communication best practices should be adapted to offshore use cases. Effective training and guidelines for remote physicians warrant comparative analyses. With increasingly complex and invasive medical tasks, differences in immediate and latent psychological responses with or without telemedical support gain importance.

CONCLUSION

The investigated telemedical support of offshore engineers during the initial survey of a simulated emergency benefited guideline adherence and required additional time. First response to medical emergencies during construction and maintenance of offshore wind farms with substantially delayed professional care may be improved by telemedical support.

Author affiliations
[1]Department of Anesthesiology and Intensive Care Medicine Campus Charité Mitte and Campus Virchow-Klinikum, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[2]Global Medical Affairs, Therapy Area Cardiovascular, Bayer Pharma AG, Berlin, Germany
[3]Faculty of Computer Science, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany
[4]Sophisticated Statistical Analysis GmbH, Berlin, Germany

Contributors PL, CS, RL, KDW and TS made substantial contributions to the conception and design of this work. PL, RL and TS made substantial contributions to the acquisition, PL, ML, KDW and TS to the analysis. PL, CS and TS to the interpretation of data.PL, ML and TS made substantial contributions to drafting the work. PL, CS, RL, ML, KDW and TS to revising it critically for important intellectual content. PL, CS, RL, ML, KDW and TS gave final approval of the version published. PL, CS, RL, ML, KDW and TS agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding statement The 'SOS—Sea and Offshore Safety' project was funded solely bythe German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), funding code 0325532.

Competing interests CS was leader of the 'SOS - Sea and Offshore Safety' project and received funding BMUB 0325532 for the project. All other authors declare that they have no financial or personal relationships with other people or organizations that biased this work. None of the authors has been remunerated for work on this publication.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

REFERENCES

1. Stuhr M, Kraus G, Weinrich N, *et al*. Erste Hilfe in Offshore-Windparks in deutschen Gewässern. *Der Notarzt* 2014;30:159–68.
2. Stuhr M, Dethleff D, Weinrich N, *et al*. Vorläufige Auswertung: Medizinische Ereignisse in Offshore-Windparks – Erste Informationen zu Unfallverletzungen und Erkrankungen. *Flug u Reisemed* 2015;22:14–19.
3. Stuhr M, Dethleff D, Weinrich N, *et al*. [Medical Emergency Preparedness in offshore wind farms : New challenges in the german north and baltic seas]. *Anaesthesist* 2016;65:369–79.
4. Hirshon JM, Galvagno SM, Comer A, *et al*. Maryland's helicopter emergency medical services experience from 2001 to 2011: system improvements and patients' outcomes. *Ann Emerg Med* 2016;67:332–40.
5. Spaite DW, Hu C, Bobrow BJ, *et al*. Association of out-of-hospital hypotension depth and duration with traumatic brain injury mortality. *Ann Emerg Med* 2017;70:522–30.
6. Østerås Øyvind, Heltne J-K, Vikenes B-C, *et al*. Factors influencing on-scene time in a rural Norwegian helicopter emergency medical service: a retrospective observational study. *Scand J Trauma Resusc Emerg Med* 2017;25:97.
7. Latifi R, Hadeed GJ, Rhee P, *et al*. Initial experiences and outcomes of telepresence in the management of trauma and emergency surgical patients. *Am J Surg* 2009;198:905–10.
8. Demaerschalk BM. Telemedicine or telephone consultation in patients with acute stroke. *Curr Neurol Neurosci Rep* 2011;11:42–51.
9. McBeth P, Crawford I, Tiruta C, *et al*. Help is in your pocket: the potential accuracy of smartphone- and laptop-based remotely guided resuscitative telesonography. *Telemed J E Health* 2013;19:924–30.
10. Brokmann JC, Rossaint R, Bergrath S, *et al*. [Potential and effectiveness of a telemedical rescue assistance system. Prospective observational study on implementation in emergency medicine]. *Anaesthesist* 2015;64:438–45.
11. Fang JL, Campbell MS, Weaver AL, *et al*. The impact of telemedicine on the quality of newborn resuscitation: a retrospective study. *Resuscitation* 2018;125:48–55.
12. Czaplik M, Bergrath S, Rossaint R, *et al*. Employment of telemedicine in emergency medicine. Clinical requirement analysis, system development and first test results. *Methods Inf Med* 2014;53:99-107.
13. Bohn A, Lang S, Wnent J, *et al*. Lay-rescuer-CPR and telephone-assisted-CPR shorten the resuscitation-free interval and thus improve long-time survival. An analysis of the German resuscitation registry. *Resuscitation* 2017;118
14. Wu Z, Panczyk M, Spaite DW, *et al*. Telephone cardiopulmonary resuscitation is independently associated with improved survival and improved functional outcome after out-of-hospital cardiac arrest. *Resuscitation* 2018;122:135–40.
15. Castan J, Paschen H-R, Wirtz S, *et al*. Massenanfall von Verletzten auf see in deutschen Gewässern. *Anaesthesist* 2012;61:618–24.
16. Perkins GD, Olasveengen TM, Maconochie I, *et al*. European resuscitation Council guidelines for resuscitation: 2017 update. *Resuscitation* 2018;123:43–50.
17. Ljunggren M, Castrén M, Nordberg M, *et al*. The association between vital signs and mortality in a retrospective cohort study of an unselected emergency department population. *Scand J Trauma Resusc Emerg Med* 2016;24:21.
18. Gold LS, Fahrenbruch CE, Rea TD, *et al*. The relationship between time to arrival of emergency medical services (EMS) and survival from out-of-hospital ventricular fibrillation cardiac arrest. *Resuscitation* 2010;81:622–5.
19. Chesters A, Harris T, Hodgetts TJ, *et al*. Survival to discharge after cardiac arrest attended by a Doctor-Paramedic helicopter emergency medical service: an Utstein-style Multiservice review of 1085 activations. *J Emerg Med* 2015;49:439–47.
20. Nolan B, Tien H, Sawadsky B, *et al*. Comparison of helicopter emergency medical services transport types and delays on patient outcomes at two level I trauma centers. *Prehosp Emerg Care* 2017;21:327–33.
21. Nichol G, Cobb LA, Yin L, *et al*. Briefer activation time is associated with better outcomes after out-of-hospital cardiac arrest. *Resuscitation* 2016;107:139–44.
22. Sejersten M, Sillesen M, Hansen PR, *et al*. Effect on treatment delay of prehospital Teletransmission of 12-lead electrocardiogram to a cardiologist for immediate triage and direct referral of patients with ST-segment elevation acute myocardial infarction to primary percutaneous coronary intervention. *Am J Cardiol* 2008;101:941–6.
23. Sanchez-Ross M, Oghlakian G, Maher J, *et al*. The STAT-MI (ST-segment analysis using wireless technology in acute myocardial infarction) trial improves outcomes. *JACC Cardiovasc Interv* 2011;4:222–7.
24. Nishijima DK, Gaona SD, Waechter T, *et al*. Out-of-Hospital Triage of Older Adults With Head Injury: A Retrospective Study of the Effect of Adding "Anticoagulation or Antiplatelet Medication Use" as a Criterion. *Ann Emerg Med* 2017;70:127–38.
25. Rörtgen D, Bergrath S, Rossaint R, *et al*. Comparison of physician staffed emergency teams with paramedic teams assisted by telemedicine--a randomized, controlled simulation study. *Resuscitation* 2013;84:85–92.
26. Skorning M, Bergrath S, Rörtgen D, *et al*. Teleconsultation in pre-hospital emergency medical services: real-time telemedical support in a prospective controlled simulation study. *Resuscitation* 2012;83:626–32.
27. Overdyk FJ, Dowling O, Newman S, *et al*. Remote video auditing with real-time feedback in an academic surgical suite improves safety and efficiency metrics: a cluster randomised study. *BMJ Qual Saf* 2016;25:947–53.
28. Branzetti JB, Adedipe AA, Gittinger MJ, *et al*. Randomised controlled trial to assess the effect of a Just-in-Time training on procedural performance: a proof-of-concept study to address procedural skill decay. *BMJ Qual Saf* 2017;26:881–91.
29. Zanner R, Wilhelm D, Feussner H, *et al*. Evaluation of M-AID, a first aid application for mobile phones. *Resuscitation* 2007;74:487–94.
30. Renshaw J, Eaton G, Gregory P, *et al*. The BHF PocketCPR smartphone application: 'Staying alive' with bystander CPR. *Resuscitation* 2017;118:e3–4.
31. Lin Y-Y, Chiang W-C, Hsieh M-J, *et al*. Quality of audio-assisted versus video-assisted dispatcher-instructed bystander cardiopulmonary resuscitation: a systematic review and meta-analysis. *Resuscitation* 2018;123:77–85.
32. Beskind DL, Stolz U, Thiede R, *et al*. Viewing an ultra-brief chest compression only video improves some measures of bystander CPR performance and responsiveness at a mass gathering event. *Resuscitation* 2017;118:96–100.
33. Salisbury C, O'Cathain A, Thomas C, *et al*. Telehealth for patients at high risk of cardiovascular disease: pragmatic randomised controlled trial. *BMJ* 2016;353.
34. Siu LW, Boet S, Borges BCR, *et al*. High-Fidelity simulation demonstrates the influence of anesthesiologists' age and years from residency on emergency cricothyroidotomy skills. *Anesth Analg* 2010;111:955-60.
35. Beom JH, Kim MJ, You JS, *et al*. Evaluation of the quality of cardiopulmonary resuscitation according to vehicle driving pattern, using a virtual reality ambulance driving system: a prospective, cross-over, randomised study. *BMJ Open* 2018;8:e023784.
36. Cooper S, Cant R, Porter J, *et al*. Rating medical emergency teamwork performance: development of the team emergency assessment measure (team). *Resuscitation* 2010;81:446–52.
37. Sunde GA, Sandberg M, Lyon R, *et al*. Hypoxia and hypotension in patients intubated by physician staffed helicopter emergency medical services - a prospective observational multi-centre study. *BMC Emerg Med* 2017;17:22.
38. Pelaccia T, Tardif J, Triby E, *et al*. From context comes expertise: how do expert emergency physicians use their know-who to make decisions? *Ann Emerg Med* 2016;67:747–51.
39. Oliver GJ, Walter DP, Redmond AD. Prehospital deaths from trauma: are injuries survivable and do bystanders help? *Injury* 2017;48:985–91.
40. Bray J, Case R, Cartledge S, *et al*. Barriers to bystander cardiopulmonary resuscitation (CPR) in a dispatcher CPR system: a qualitative review of emergency calls. *Resuscitation* 2017;118:90.
41. Pollack RA, Brown SP, Rea T, *et al*. Impact of bystander automated external defibrillator use on survival and functional outcomes in shockable observed public cardiac arrests. *Circulation* 2018;137:2104–13.

Supplement [31]

Supplement A: Post-Hoc Power

Post-hoc power was calculated for the non-inferiority test e using nQuery Version 8.4.0.0 (Stat. Solutions Ltd. & South Bank, Crosses Green, Cork, Ireland). Results are shown in table 2, input values are listed in table 3. Descriptive statistics (see table 3) were calculated using EquivTest (StatCon GmbH, Witzenhausen, Germany).
The nQuery output was:

> When the sample sizes in the groups are 9 and 8, a two group one-sided 0.05 significance level t-test will have 22.26 % power to reject the null hypothesis that the test [(performance of supported non-professionals)] is inferior to the standard [(performance of unsupported professionals)] in favor of the alternative hypothesis that the treatment [(support of non-professionals)] is non-inferior [(to no support of professionals)], assuming that the expected difference in means is 0, a non-inferiority margin of 1 and the common standard deviation is 2.23.

	realistic	optimistic
Test Significance Level, α (one-sided)	0.050	0.050
Non-Inferiority Limit Difference Δ_0	1.000	1.000
Expected Difference, Δ_1	0.000	0.000
Difference of Deltas, $\Delta_0 - \Delta_1$	1.000	1.000
Common Standard Deviation, σ	2.230	1.770
Effect Size, $\delta = \lvert\Delta_0 - \Delta_1\rvert / \sigma$	0.448	0.565
Power (%)	22.26 [80]	29.65 [80]
Group 1 Sample Size, n_1	9 [63]	9 [40]
Group 2 Sample Size, n_2	8 [63]	8 [40]
Sample Size Ratio, n_2/n_1	0.889 [1]	0.889 [1]
Total Sample Size, $N = n_1 + n_2$	17 [126]	17 [80]

Table 2: Post-hoc power calculation results for test e using model *MTE0U-1 / Two Group t-test of Non-Inferiority in Means Unequal n's*. The realistic power calculation assumes the common standard deviation σ found by pooling the two underlying standard deviations (see table 3). The optimistic power calculation assumes the smaller of the two underlying standard deviations (that of the group of supported non-professionals). To achieve a power of 80 % with $\sigma = 2.230$ and sample size ration $n_2/n_1 = 1$, a total sample size of 126 would have been required. With a $\sigma = 1.770$ and sample size ration $n_2/n_1 = 1$, a total sample size of 80 would have been required.

Group	supported non-professionals	unsupported professionals
Mean	11.8333	10.6875
Standard Error (Mean)	0.5892	0.9398
Geometric Mean	11.7135	10.3237
Median	11.5000	11.7500
Standard Deviation	1.7677	2.6583
Variance	3.1250	7.0669
Min	9.0000	5.5000
Max	14.0000	13.0000
Range	5.0000	7.5000
n	9	8
Common standard deviation σ	2.22998	

Table 3: Descriptive statistics of the groups of supported non-professionals and unsupported professionals also used in test e (see results in figure 3). Common standard deviation was estimated by pooling standard deviations of both groups.

1

Supplement [32]

Supplement B: Data validation

Programmatically, data size, format, consistency and range were validated, batch total and logic checks (an item can be either correctly, incorrectly or unclearly performed, item dependencies should correspond with their ratings) performed.

Quantile-quantile plots are shown in figure 7. No normal distribution is discernible, supporting the choice of non-parametric tests. Tests for distribution were deemed superfluous hereafter.

To stress the validity of the data used for analysis, the plausibility of rater disagreement on performance is investigated hereafter. The authors hope that this investigation may guide methodical improvements of subsequent studies.

Agreement Outlier Discussion

Because marginal totals for agreement on performance are imbalanced in the corresponding contingency table, Cohen's Kappa (sensitive to this imbalance to the point of paradoxical behavior) was not calculated. Uni-directional proportions of agreement (p_{pos}, p_{neg}), their sum weighted by proportions of total ratings (p_0) and prevalence- and bias-adjusted kappa (PABAK, $K = 2p_0 - 1$) were calculated per team and per item, instead (see figure 6).

Outlying teams: Minima in proportionate positive agreement on performance were identified for one supported and one unsupported team of non-professionals ($p_{pos} = .48$ and $p_{pos} = .40$) and for one team of unsupported professionals ($p_{pos} = .36$). No other proportionate positive agreement on team performance undercut $p_{pos} = .61$.

The team of unsupported non-professionals spent nearly eight minutes trying to survey blood pressure, pulse rate and pupillary light reflex, then decided they had finished the task. This leaves very few items for the raters to agree on, some of which were repeatedly performed but not consistently correctly. The team of unsupported professionals decided they had completed the task after less than three minutes, not actually performing most of the surveys. This misunderstanding by participants, whether and to what extent activities were expected to be performed in a simulation setting, opened room for rater disagreement. The team of supported non-professionals, while performing most items, left out arguably critical steps such as in asking for the last meal not asking for the exact time. As another example, in measuring the breathing rate, they counted for too short a time, straying from the correct value by an arguably relevant amount. This room for interpretation may have been particularly large during the first recordings to be rated, when clarity of the frame of reference was only beginning to evolve.

Items of disagreement: Among professionals, proportionate positive agreement on items 3-6, 8, 16 and 17 undercut $p_{pos} = .63$.

Among non-professionals, that on items 4, 6, 8, 10, 11, 16 and 17 undercut the same threshold.

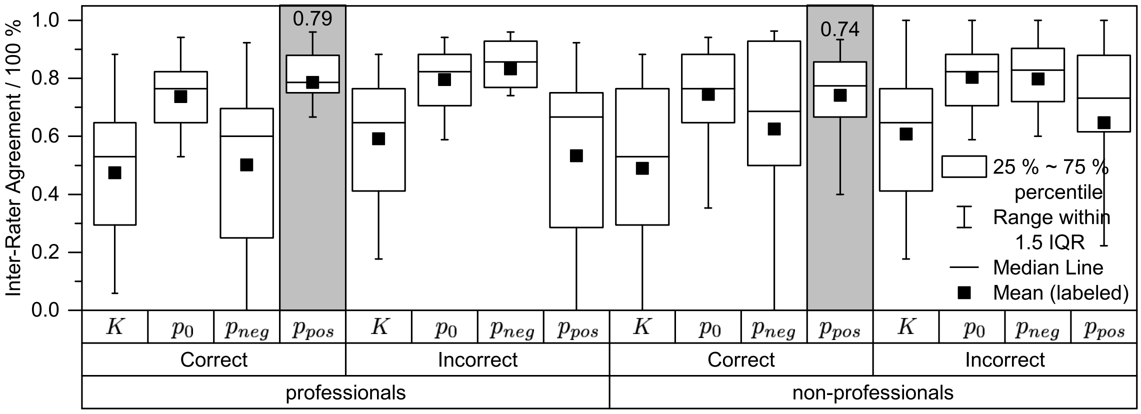

Figure 6: Overview of inter-rater agreement on skewed data per team. Inter-rater agreement on items rated correctly or incorrectly performed by professionals or non-professionals: prevalence- and bias-adjusted kappa (PABAK, K), the proportion of observed agreement p_0, which is the sum of observed proportion of positive agreement p_{pos} and observed proportion of negative agreement p_{neg} weighted by positive and negative ratings per reviewer and team. In a range of $[0, 1]$, higher means more inter-rater agreement. Highlighted and labeled agreement dimensions characterize score dimensions used for quantitative analysis.

Item 3 concerns the Glasgow-Coma-Scale. No team of unsupported non-professionals but all teams of unsupported professionals at least attempted to survey this item. In contrast, all supported teams attempted to survey it. Thus, the room for disagreement on correct performance was reduced for non-professionals. The interpretation of correct performance seems to have ranged from accepting an intuitive close-enough estimation to accepting only accurately deduced scores, hence the low agreement.
Item 4 concerns normal breathing. One rater rated normal breathing and rate (item 5) independently with the rate being more often correctly surveyed:

- Items 4 and 5 were differently rated in 7 out of 15 teams who correctly performed either.
- In 6 of these 7, item 5 was rated as correctly performed without item 4 being rated as correctly performed.

The other rater rated "normal" as more often correctly performed:

- Both items were rated differently in 4 out of 6 teams who correctly performed either.
- Only in 1 out of these 4 teams was item 5 rated as correctly performed without item 4 having been rated so.

Conclusions should not be drawn from this tiny sample, a difference in interpretation of "normal" among raters seems to be hinted at.
Item 5 concerns the breathing rate. Raters agreed less on teams of professionals, of whom nearly twice as many surveyed the item at all compared with non-professionals. Among this higher number of potentially correctly performed items, the aforementioned difference in interpretation of arguably relevant deviation from the simulated rate may be dominant. Closely related is *item 6 concerning correctness of respiratory rate measurement*: One rater rated the measurement correctly performed in 14, the other in only 5 teams. The authors hypothesize that the raters prioritized minimal measurement duration differently.
In item 8 on correctly surveying radial pulse, like in item 6, one rater rated the survey correctly performed in 32, the other in only 11 teams. The authors hypothesize that the correct result of the measurement was for only one rater integral to the item. However, this result might not have been attained once the rate was available from the ECG in intervention groups. Supported were only 2 out of the 11 teams who correctly performed palpation. Of all 18 supported teams, the same rater rated 13 as having tested for radial and carotid pulse (item 7). If the result from palpation were not integral to the correct survey, 11 supported teams of non-professionals and professionals would have been rated as not having correctly performed palpation.
On items 10 and 11 concerning blood pressure and whether it is normal, proportionate positive rater agreement on non-professionals was low. Particularly, raters disagreed on correct performance by 6 supported and 3 unsupported teams. In 5 of the supported teams, the monitoring device measured a blood pressure lower than the simulator was programmed to output and the threshold criterion for rating the item correctly performed. One rater therefore did not mark these items correctly performed. The remaining supported team never notified the remote physician of the blood pressure, which one rater therefore deemed not measured, not knowing that the remote physician had received the measurement result from the monitoring device. Among the unsupported teams, one simulator malfunction, one supposition on normal blood pressure without measurement and one case of incorrectly performed item 9 (correctness of measurement) created uncertainty for the raters.
Item 16 concerns asking about the last food intake. One rater rated it as incorrectly performed 10 times more often than the other, each time because the exact time of the last meal had not been inquired after. Only the general time of day had been asked after, which the other rater deemed sufficient.
Item 17 concerning the course of accident was rated differently for 24 teams. The rater who rated the item these 24 times incorrectly performed noted down for 22 of these teams that the exact height of the fall of the patient had not been explicitly asked after. In 8 of these 24 teams, the burn that caused the fall had been left out of the course of accident. The other rater rated less accurate courses of accident as correct item performance in all of these 24 teams.
In conclusion, the rater disagreement seems to stem from

1. simulation artifacts causing participants to behave differently in an artificial setting than they might in reality,
2. an emerging frame of reference in which the

raters' interpretations became consistent,

3. differences in raters' prioritization of accuracy of deductions versus precision of adherence to instructions or guidelines,
4. differences in raters' understanding of item dependencies.

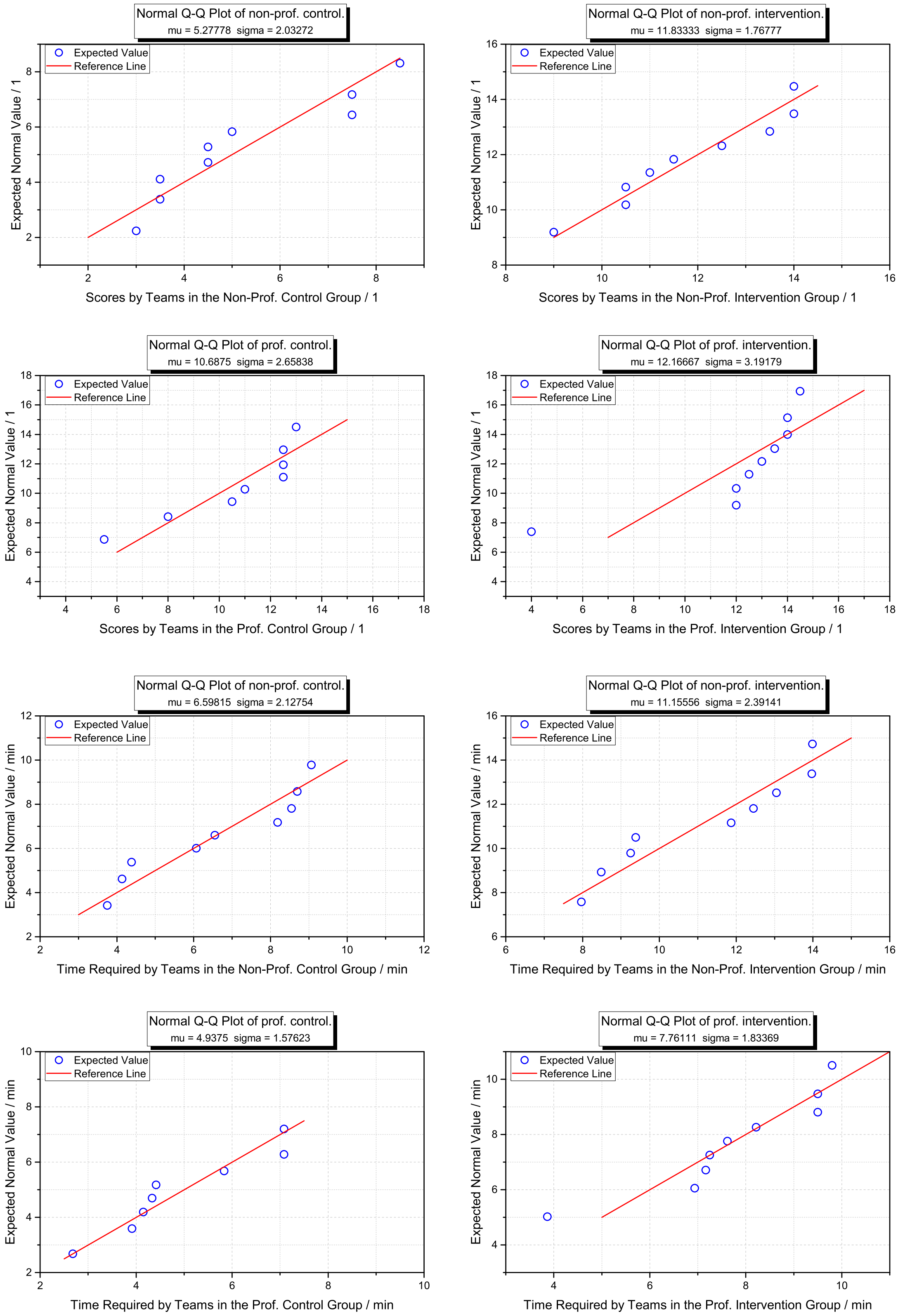

Figure 7: Quantile-quantile plots for every investigated group.

Mein Lebenslauf wird in dieser elektronischen Version nicht veröffentlicht.

Philipp Landgraf

Publikationsliste

Originalarbeiten (Peer Reviewed)

2019. Landgraf, P, Spies, C, Lawatscheck, R, Luz, M, Wernecke, KD, and Schröder, T. Does Telemedical Support of First Responders Improve Guideline Adherence in an Offshore Emergency Scenario? A Simulator-Based Prospective Study. *BMJ Open* 2019;9(8):e027563. doi:`10.1136/bmjopen-2018-027563`.

Abstracts

2018. Roethe, AL, Landgraf, P, Schröder, T, Vajkoczy, P, and Picht, T. Monitor-based exoscopic neurosurgical interventions: a task-based preparatory evaluation of 3D4k surgery. *69 Jahrestagung der Deutschen Gesellschaft für Neurochirurgie (DGNC)* 2018;doi:`10.3205/18dgnc387`.

Vorträge

2019a. Schröder, T, Lütz, A, Landgraf, P, Mülich, M, Hartog, C, Michalsen, A, and Gerritsen, R. Masterclass: Protect the mind: Delirium management, the answer to all your pain. LIVES 2019: 32nd Annual Congress, Berlin; 2019a. European Society of Intensive Care Medicine.

2019b. Landgraf, P. The Unkillable Patient. imagINe surgery - Surgical Education and Tool Training in the Digital Age, Berlin; 2019b. Cluster of Excellence An Interdisciplinary Laboratory, Charité and Rice University in Houston, Texas.

2018a. Landgraf, P, Dußmann, P, Habicher, M, and Jammer, I. Crisis resource management and emergency training, Masterclass: When ICU enhances postoperative recovery. LIVES 2018: 31st Annual Congress, Paris; 2018a. European Society of Intensive Care Medicine.

2018b. Landgraf, P. Vital Sign Retrieval: Telemedical Support for Offshore Engineers. conhIT - Connecting Healthcare IT, Berlin; 2018b. Bundesverband Gesundheits-IT - bvitg e.V.

2017. Landgraf, P. Vital Sign Retrieval: Telemedical Support for Offshore Engineers. 62. GMDS Jahrestagung 2017, Oldenburg; 2017. Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e.V.

2016. Landgraf, P, Kohn, S, and Boger, D. Pre-Course 1: Tricks und Kniffe in der Simulation – Low Cost vs. High End Simulation. 14. Interdisziplinäres Symposium für Simulation in der Medizin - Simulationstraining für Patientensicherheit, Dresden; 2016. InPASS - Institut für Patientensicherheit und Teamtraining GmbH.

2015. Kohn, S, Landgraf, P, Gehring, R, Stricker, E, and Lebeau, C. Pre-Course 1: Simulationstechnik selbst gestalten - Was Enthusiasten der Simulation gebaut haben. 13. Interdisziplinäres Symposium für Simulation in der Medizin - Simulationstraining für Patientensicherheit, München; 2015. InPASS - Institut für Patientensicherheit und Teamtraining GmbH.

2014. Schröder, T, Habicher, M, Landgraf, P, and Sander, M. Simulationstraining. Der Springende Punkt - 11. Kieler Symposium für Kardiovaskuläre Intensivmedizin und -pflege, Kiel; 2014. Klinik für Herz- und Gefäßchirurgie, UKSH Campus Kiel.

Monographien

2016. Landgraf, P. Implementability of Interoperability in Anesthesia Workstations; 2016. Masterarbeit Systems Architecting Prof. Hansen, Prof. Preim, Note: 1,7.

2013. Landgraf, P. Interferometric Investigation of Nanomechanical Gold-coated Silicon Nitride Resonators; 2013. Bachelorarbeit Nanomechanik Prof. Weig, Note: 1,3.

2010. Landgraf, P and Totel, M. Leermodulrecycling und Öffentlichkeitsarbeit im Rahmen des Projektes „Sammeldrache“ der Goethe-Oberschule Berlin; 2010. 5. Prüfungskomponente Abitur, Note: 14 Punkte.