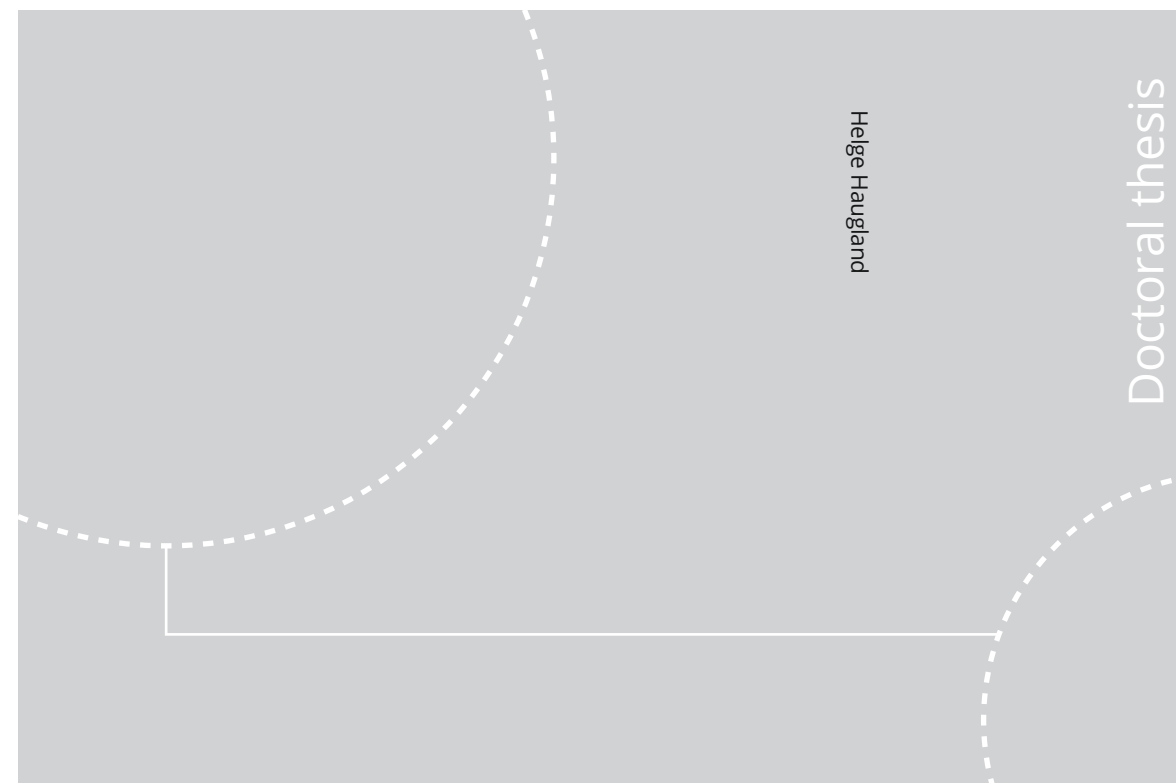


ISBN 978-82-326-4530-5 (printed ver.)  
ISBN 978-82-326-4531-2 (electronic ver.)  
ISSN 1503-8181



Doctoral theses at NTNU, 2020:89

Helge Haugland

## Quality in physician-staffed pre-hospital emergency medical services

– developing a method for continuous  
quality estimation

Helge Haugland

# **Quality in physician-staffed pre-hospital emergency medical services**

– developing a method for continuous quality  
estimation

Thesis for the Degree of Philosophiae Doctor

Trondheim, March 2020

Norwegian University of Science and Technology  
Faculty of Medicine and Health Sciences  
Department of Circulation and Medical Imaging



**NTNU**

Norwegian University of  
Science and Technology

**NTNU**

Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences

Department of Circulation and Medical Imaging

© Helge Haugland

ISBN 978-82-326-4530-5 (printed ver.)

ISBN 978-82-326-4531-2 (electronic ver.)

ISSN 1503-8181

Doctoral theses at NTNU, 2020:89

Printed by NTNU Grafisk senter

To my father

*"Quality is not an act. It is a habit."*

Aristotle





## 1 Table of contents

1	Table of contents .....	5
2	Acknowledgements.....	7
3	List of papers.....	9
4	Acronyms .....	11
5	Sammendrag .....	13
6	Summary .....	17
7	Introduction to study .....	19
8	Background .....	21
8.1	Pre-hospital care .....	21
8.2	Physician-staffed emergency medical services.....	22
8.3	What is quality? .....	25
8.4	Quality measurement in health care .....	26
8.5	Quality measurement in pre-hospital critical care .....	34
9	Motivation for the thesis .....	45
10	Aims .....	47
11	Materials and methods.....	49
11.1	Study setting and descriptions.....	49
11.2	Participants .....	50
11.3	Measurement and data collection.....	53
11.4	Statistical methods.....	56
11.5	Ethical considerations and approvals .....	57
11.6	Financial support.....	58
12	Results.....	59
12.1	Paper I .....	59
12.2	Paper II .....	59
12.3	Paper III .....	60

12.4	Paper IV .....	63
13	Discussion.....	67
13.1	The importance of a multidimensional quality measurement .....	67
13.2	Precision in dispatch is necessary to improve quality .....	68
13.3	Response time (Time to arrival of P-EMS). ....	74
13.4	Mortality in P-EMS .....	76
13.5	Limitations.....	78
14	Future perspectives .....	81
15	Conclusions .....	85
16	References .....	87
17	Appendix A .....	93
18	Papers I-IV .....	95

## 2 Acknowledgements

To my main supervisor **Andreas Krüger**; thanks for introducing me to research. It was your idea to develop a method for quality measurement in physician-staffed emergency medical services, and through this topic you inspired me to become a PhD student. Your genuine motivation for the project, your energy and above all – your presence as a supervisor has been a tremendous help for me. Thanks for being available and friendly whenever I call you with small and big questions. I am deeply grateful for your guidance.

**Marius Rehn**, thanks for your systematic approach whenever I need your help. Thanks for responding so quickly and thoroughly to my requests. I am especially grateful for your competent, diplomatic and effective hosting of the consensus meeting in paper I. Your respectful and polite way of approaching everyone involved was pivotal to make the meeting successful. I would also like to thank you for your rigorous and patient supervision of paper II.

**Pål Klepstad**, thanks for precise, critical and direct supervision. You are not afraid to use the correction pen when you think it is needed and your input is always improving my work. Most of all, however, your feedback has given me a feeling of reassurance and safety. I am impressed by your ability to separate the important from the unimportant. This combination of pragmatism and high knowledge makes you a credible and effective supervisor both in research and clinical work.

Thanks to **The Norwegian Air Ambulance Foundation** and all its donors for financing my research. Being a part of The Norwegian Air Ambulance Foundation has been a true pleasure; your easy-going and professional staff has given me lots of inspiration and help during these years. Thanks for providing me with necessary infrastructure and for introducing me to national and international networks. To all air ambulance crew members and employees in The Norwegian Air Ambulance Foundation working at Rosten, Trondheim; thanks for inspiring and nice chats at the lunch table and in the corridors. **Oddvar Uleberg**; thanks a lot for your efforts in paper II, for your practical advices in the final phase of my PhD project and for making me laugh. A particular thank you to the always helpful **Elin Staveli** – you are a rock.

To all participating air ambulance services in paper III and IV; thank you for your efforts – this PhD could not have been possible without your help. **Anna Olkinuora**, **Leif Rognås** and **David Ohlén**; thank you for running The Nordic Quality Measurement Study in your countries – I am deeply grateful for all the work you put into this study.

Thanks to my bosses at St. Olav's Hospital; **Sigurd Fasting** and **Jostein Dale** for giving me the trust and freedom I needed in the years as a PhD student. I have not taken this for granted and I thank you both for your goodwill.

To all my former colleagues at Førde Sentralsjukehus; thank you for raising me as an anaesthesiologist. I am particularly grateful to my supervisors **Svein Ove Vingsnes** and **Rune Larsen**, whose caring and competent guidance I will always appreciate. Finally, I am extremely grateful for the opportunity to work at The Norwegian Air Ambulance base in Førde; the colleagues, the professional challenges and the beautiful scenery made me feel like coming home.

To all my good friends and colleagues at the 330 squadron, Ørland. I am truly grateful for working with all of you and I am looking forward to every shift at the Sea King Rescue Helicopter. Not only because of the joy of solving meaningful tasks together, but also because I appreciate spending time with you.

My good friend **Peter Bakkeid**; thanks for your unselfish help and support whenever my IT knowledge comes short. Thank you for stimulating discussions about quality management in business; this has certainly expanded my view on quality measurement in health care.

To my parents **Anne Grete Nystad** and **Sverre Haugland**; thank you for raising me to the person I am. Sadly, my father died way too young; this thesis is for him. I would also like to express my gratitude to my sister **Gunhild Nystad Haugland** for being such a good companion in both childhood and adulthood.

Finally, but most importantly I would like to thank my own family. **Heidi**, **Helle**, **Sverre**, **Vemund** and **Tarjei**; loving you and being loved by you is the greatest gift of my life.

### 3 List of papers

**Paper I:** Haugland H, Uleberg O, Klepstad P, Krüger A, Rehn M. Quality measurement in physician-staffed emergency medical services: a systematic literature review. *Int J Qual Health Care*. 2019 Feb 1;31(1):2-10. doi: 10.1093/intqhc/mzy106.

**Paper II:** Haugland H, Rehn M, Klepstad P, Krüger A, EQUIPE-collaboration group. Developing quality indicators for physician-staffed emergency medical services: a consensus process. *Scand J Trauma Resusc Emerg Med*. 2017 Feb 15;25(1):14. doi: 10.1186/s13049-017-0362-4.

**Paper III:** Haugland H, Olkinuora A, Rognås L, Ohlén D, Krüger A. Testing quality indicators and proposing benchmarks for physician-staffed emergency medical services: A prospective Nordic multi-centre study. *BMJ Open*. 2019;9:e030626. doi:10.1136/bmjopen-2019-030626.

**Paper IV:** Haugland H, Olkinuora A, Rognås L, Ohlén D, Krüger A. Exploring the relationship between mortality and quality scores in Nordic physician-staffed emergency medical services

Submitted to *Acta Anaesthesiologica Scandinavica* November 2019



## 4 Acronyms

ACS-COT The American College of Surgeons Committee on Trauma

B.C. Before Christ

ED Emergency Department

EMCC Emergency Medical Coordination Centre

EMS Emergency Medical Services

ETI Endotracheal Intubation

HEMS Helicopter Emergency Medical Services

IOM The Institute of Medicine

PCI Percutaneous Coronary Intervention

PDSA Plan Do Study Act

P-EMS Physician-staffed Emergency Medical Services

QI Quality Indicator

SAR Search and Rescue

STEMI ST-segment Elevation Myocardial Infarction

TQS Total Quality Score

WHO World Health Organization





## 5 Sammendrag

Norsk tittel: Kvalitet i legebemannede utrykningstjenester – utvikling av en metode for kontinuerlig kvalitetsmåling.

Pre-hospitale utrykningstjenester er tradisjonelt bemannet av paramedics, ambulansearbeidere og sykepleiere og er en viktig del av helsetjenesten i de fleste land. I noen høyinntektsland finnes i tillegg et høyere nivå av pre-hospitale utrykningstjenester; de legebemannede utrykningstjenester. I Norge er denne tjenesten primært organisert i den landsdekkende og statsfinansierte luftambulansetjenesten.

Kvalitetsmåling i legebemannede utrykningstjenester er nødvendig for å forbedre disse tjenestenes kvalitet. Til tross for at kvalitetsmåling er blitt utbredt i helsetjenesten, er det lite kunnskap og konsensus om dette emnet i legebemannede utrykningstjenester. Dette gjør kvalitetsmåling av legebemannede utrykningstjenester til et prioritert forskningsområde. Målet med doktorgradsprosjektet var å utvikle en modell for kontinuerlig kvalitetsmåling i legebemannede utrykningstjenester.

Studie 1 var en systematisk litteraturgjennomgang. Målet med denne studien var å identifisere, beskrive og evaluere studier som omhandler kvalitetsmåling i legebemannede utrykningstjenester. Vi søkte etter artikler som beskriver bruken av en eller flere kvalitetsindikatorer i legebemannede utrykningstjenester. Dette litteratursøket ble gjort i databasene MEDLINE, Embase og Scopus og identifiserte 4 699 artikler. Inklusjonskriteriene ble oppfylt av 27 artikler. De mest brukte kvalitetsindikatorene var «Etterlevelse av medisinske retningslinjer», «Utførelse av avanserte intervensjoner», «Responstid» og «Uønskede hendelser». Studien påviste at det var liten felles forståelse for hvilke kvalitetsindikatorer man bør bruke i legebemannede utrykningstjenester. Vi observerte også at litteraturen var dominert av artikler som bare brukte én kvalitetsindikator, noe som øker risikoen for at perspektivet i kvalitetsmålingen blir for smalt. Vi konkluderte med at fremtidig kvalitetsmåling i legebemannede utrykningstjenester bør baseres på et sett med konsensus-baserte kvalitetsindikatorer for å sikre en tilnærming til kvalitetsmåling som er omfattende nok.

Studie 2 var en konsensusprosess hvor vi benyttet en modifisert nominell gruppeteknikk. Målet var å utvikle kvalitetsindikatorer for internasjonal bruk i legebemannede utrykningstjenester. Et panel

bestående av 18 eksperter og representanter for interessegrupper fikk i oppgave å foreslå, rangere og velge ut et sett med kvalitetsindikatorer. Alle medlemmene av ekspertpanelet var på ulike vis vurdert å inneha ekspertise om legebemannede utrykningstjenester eller samarbeidende tjenester. Medlemmene kom fra åtte ulike nasjoner. Ekspertpanelet oppnådde konsensus om 15 responsspesifikke og 11 systemspesifikke kvalitetsindikatorer for legebemannede utrykningstjenester. Alle de seks kvalitetsdimensjonene som Institute of Medicine har definert var representert og kvalitetsindikatorene bestod både av struktur-, prosess- og resultatindikatorer. Dette settet med 26 kvalitetsindikatorer ble utviklet for å kunne gi en bred og omfattende nok tilnærming til kvalitetsmåling i legebemannede utrykningstjenester, som i tillegg kunne gjøre det mulig med fremtidig kvalitetsmåling på tvers av tjenester.

Studie 3 var en prospektiv observasjonell studie som involverte 16 legebemannede luftambulansetjenester i Finland, Sverige, Danmark og Norge. De involverte tjenestene registrerte nødvendige data for å skåre de 15 responsspesifikke kvalitetsindikatorene fra studie 2. Videre ble det innhentet 30-dagers mortalitetsdata for de inkluderte pasientene. Resultatene fra denne studien presenteres i to ulike artikler; artikkel III og IV.

Målet med artikkel III var å teste viktige egenskaper ved kvalitetsindikatorene i en reell klinisk hverdag. Disse egenskapene var gjennomførbarhet, rangeringsevne, variabilitet, handlingsmulighet og dokumentasjon. Videre hadde vi et mål om å etablere referansepunkt for fremtidig kvalitetsmåling i legebemannede utrykningstjenester. Datasettet bestod av 5 638 henvendelser til tjenestene som var med i studien. 2 814 av henvendelsene resulterte i pasientkontakt. Det var gjennomførbart å innhente data til alle kvalitetsindikatorene. Variabiliteten var adekvat for 14 av 15 kvalitetsindikatorer. Rangeringsevnen var adekvat for alle kvalitetsindikatorene. Tjenestenes mulighet til å kunne påvirke kvalitetsindikatoren score i positiv retning ble vurdert som adekvat for ti av kvalitetsindikatorene. Dokumentasjonen var adekvat for 14 kvalitetsindikatorer. Vi presenterte referansepunkter for alle kvalitetsindikatorer og vi presenterte i tillegg en aggregert skår som samler prestasjonen for alle 15 kvalitetsindikatorer i én skår; den totale kvalitetsskåren (Total Quality Score (TQS)). Både referansepunktene og den totale kvalitetsskåren kan være nyttige verktøy i fremtidig kvalitetsmåling.

Målet med artikkel IV var å studere om det finnes en sammenheng mellom de 15 responsspesifikke kvalitetsindikatorene som ble utviklet i studie 2 og 30-dagers mortalitet. Vi observerte at andelen pasienter som overlevde de 30 første dager etter å ha blitt tatt hånd om av utrykningstjenestene, varierte

signifikant. Den respektive 30-dagers overlevelsen var 76.1 % (Finland), 83,5 % (Danmark), 84,1 % (Norge) and 89,0 % (Sverige). For 9 av 15 kvalitetsindikatorer fant vi en korrelasjon mellom kvalitetsindikatorskår og 30-dagers mortalitet. Det var imidlertid ikke noe som tilsa at en presumptiv bedre prestasjon bedømt ut fra en kvalitetsindikator førte til lavere mortalitet enn dårlige prestasjoner – og vice versa. En korrelasjonsanalyse av TQS og 30-dagers mortalitet viste ingen signifikant korrelasjon. Vi konkluderte med at måling av mortalitet og prosesskvalitet er komplementære størrelser og de er begge sentrale for å identifisere den totale kvaliteten som et system oppnår.

Til sammen bidrar studiene med økt kunnskap om kvalitetsmåling i legebemannede utrykningstjenester. De bidrar også med verktøy for fremtidig kvalitetsmåling i disse tjenestene. Kvalitetsmåling er en forutsetning for kvalitetsforbedring. Derfor anser vi utviklingen av disse verktøyene som et nyttig første steg for å igangsette kvalitetsforbedringsprosjekter i legebemannede utrykningstjenester.

Kandidat: Helge Haugland

Stiftelsen Norsk Luftambulans

Institutt for sirkulasjon og bildediagnostikk

Fakultet for Medisin og Helsevitenskap, NTNU

Hovedveileder: Andreas Krüger

Biveiledere: Professor Pål Klepstad og Marius Rehn

Finansieringskilde: Stiftelsen Norsk Luftambulans

*Overnevnte avhandling er funnet verdig til å forsvares offentlig for graden Ph.d. i medisin.  
Disputas finner sted i KBA på Kvinne-Barn-Senteret, St. Olavs Hospital  
Torsdag 5. mars 2020, kl 1215*

## 6 Summary

Traditionally, Emergency Medical Services (EMS) are staffed by paramedics, emergency medical technicians, or nurses, and are an important part of health service in most countries. In some high-income countries, a second tier of pre-hospital services exists: the physician-staffed EMS (P-EMS). In Norway, P-EMS is organised primarily in the government-funded National Air Ambulance Service. Quality measurement of P-EMS is necessary to improve service quality. Despite widespread quality measurement initiatives in health care, knowledge of and consensus on this topic are scarce in P-EMS, making the quality measurement of these services a high-priority research area. The aim of this PhD project was to develop a model for continuous quality measurement in physician-staffed pre-hospital emergency services.

Study 1 was a systematic literature review. The aim of this review was to identify, describe, and evaluate studies of quality measurement in P-EMS. The MEDLINE, Embase, and Scopus databases were searched for papers describing the use of one or more quality indicators in P-EMS. The literature search identified 4,699 papers. Twenty-seven papers met the eligibility criteria and were included in the data extraction. The most widely used QIs were “Adherence to medical protocols”, “Provision of advanced interventions”, “Response time”, and “Adverse events”. However, the review demonstrated a lack of shared understanding of which QIs to use in P-EMS. Moreover, papers using only one QI dominated the literature, thus increasing the risk of a narrow perspective in quality measurement. We concluded that future quality measurement in P-EMS should rely on a set of consensus-based QIs, ensuring a comprehensive approach to quality measurement.

Study 2 was a consensus process using the modified nominal group technique. The aim was to develop quality indicators for international P-EMS. A panel of 18 P-EMS experts and stakeholders was given the task of proposing, ranking, and selecting a set of quality indicators. All panel members were, in different ways, considered experts in P-EMS or in collaborating services of P-EMS and practiced in eight different countries. The expert panel reached consensus on 15 response-specific and 11 system-specific quality indicators for P-EMS. All six quality dimensions stated by the Institute of Medicine were covered, and the quality indicators represented structure, process, and outcome indicators. This large set of 26 quality indicators was developed to represent a broad and comprehensive approach to quality

measurement in P-EMS, allowing future quality measurement to be comparable across different P-EMS systems.

Study 3 was a prospective observational study including 16 physician-staffed helicopter emergency services in Finland, Sweden, Denmark, and Norway. The involved P-EMSs registered necessary data to score the 15 response-specific QIs developed in study 2. Moreover, 30-days mortality for the included patients was collected. Study 3 resulted in two papers: papers III and IV.

The aim of paper III was to test the QIs for important characteristics in a real clinical setting. These characteristics were feasibility, rankability, variability, actionability, and documentation. We further aimed to propose benchmarks for future quality measurements in P-EMS. The dataset consisted of

5,638 requests to the participating P-EMSs; 2,814 requests resulted in patient contact. All QIs were feasible to obtain. The variability of 14 out of 15 QIs was adequate. Rankability was adequate for all QIs. Actionability was assessed as adequate for 10 QIs. Documentation was adequate for 14 QIs. Benchmarks for all QIs were proposed. We also presented an aggregated score made up of the performances for all 15 QIs: the Total Quality Score (TQS). Both the benchmarking and the TQS can be useful tools for future quality measurement.

The aim of paper IV was to assess whether there was a relationship between the 15 response-specific quality indicators developed in study 2 and 30-days mortality. We observed that the proportion of patients surviving until 30 days after the actual P-EMS response varied significantly; the figures were 76.1% (Finland), 83.5% (Denmark), 84.1% (Norway), and 89.0% (Sweden), respectively. For 9 out of 15 QIs, we found a correlation between QI score and 30-days mortality. However, for these nine QIs, there was no logical consistent pattern showing that presumptive good QI performances lead to lower mortality than do poor performances and vice versa. A correlation analysis of TQS and 30-days mortality revealed no significant correlation. We concluded that our findings indicate that measurements of mortality and process quality are complementary and that both are central to identifying the total quality achieved in a system.

Together, these studies provide increased knowledge about quality measurement in P-EMS. More importantly, they develop tools for future quality measurement in these services. Quality measurement is a pre-requisite for quality improvement. Therefore, providing these tools is a useful step for launching quality improvement initiatives in P-EMS.

## 7 Introduction to study

*“The ultimate goal is to manage quality. But you cannot manage it until you have a way to measure it, and you cannot measure it until you can monitor it”.*

Florence Nightingale (1820-1910)

As early as the 19<sup>th</sup> century, Florence Nightingale pointed out that measurements are imperative to quality improvement. Today, the importance of quality measurement in health care is widely recognised [1-4]. Because it is difficult to measure quality directly, quality is typically measured by using quality indicators (QI), which are instrumental in helping clinicians, organisations, health care managers, and societies achieve improvements in health care quality [5]. Thus, defining quality indicators (QI) for traditional EMS and P-EMS has been identified as a high-priority research area [6-8].

In Norway, selected data from P-EMS activity are stored in databases. Attempts to extract information about the quality of the service have been limited primarily to data on time-variables. Response time is important for some patients but not equally important for all patients. In selected situations, too much emphasis on these variables can be misleading or even incorrect with respect to what really represents quality for the patient. In the United Kingdom, paramedics criticised the use of a time target structure measure (eight-minute response time for 75% of category A or emergency calls) as the main performance indicator in EMS. They argued that it was a poor quality indicator because it was “too simplistic and narrow”, and that it even put patients and ambulance crews at risk [9]. The existing documentation systems of Norwegian P-EMS do not include information about patient outcomes or allow for clinical quality assessments because no quality indicators are systematically or continuously registered in P-EMS. Rehn and Krüger argue that a model for quality measurement is necessary to achieve appropriate governance, quality assurance, and even quality improvement of P-EMS [10]. Such a model should provide continuous monitoring, as this is a pre-requisite for quality improvement of a system, a process, or an organisation [11].

Patients taken care of by Nordic P-EMS constitute a very heterogenic group: neonates and elderly people, medical and surgical diagnoses, patients rescued from open water, and patients transferred from one ICU to another [12, 13]. Additionally, the challenge of isolating the pre-hospital care effect from that of the emergency department (ED) and hospital care further increases the complexity of



measuring quality in P-EMS. With this complexity of cases, treatments, and operational contexts, we argue that it is unfeasible to establish a one-dimensional quality indicator. A model for continuous quality estimation should reflect the diversity of diagnoses in patients taken care of by P-EMS, as well as the heterogeneity of P-EMS missions. What is considered beneficial for patients in P-EMS missions can be highly variable, and the model should reflect this. Thus, the quality estimation should be multi-factorial rather than based exclusively on data such as time variables or mortality.

Quality estimation of P-EMS is not straightforward. As described in this thesis, the quality dimensions are many, and the diversity of diagnoses among the patient population is large. So, how can we assess the quality of our work without measurement? And how can we improve something we do not measure? This thesis is an attempt to develop a tool for continuous quality estimation in P-EMS. By measuring the quality of P-EMS responses, we hope to identify the best performers so that we can all learn from them. By benchmarking the performance in P-EMS, we hope to stimulate and enlighten the services in their quality improvement work. A man dancing in the dark might get the feeling that he is a very good dancer. However, once the light is turned on and he sees the other dancers, he can compare his performance to those of the others. Turning the light on is, thus, a reality check. In P-EMS, the dancer is us caring for the patient. The turned-on light is the benchmarking of this care.

Furthermore, by identifying what is considered good quality in P-EMS and by giving tools to P-EMS to improve this quality, we hope to contribute to even better utilisation of P-EMS given its limited availability. We hope this will help ensure that the patients in most need of P-EMS assistance are provided with this care.

My oldest daughter came by when I was writing the final part of this thesis. She looked at my laptop and said soberly, *"If I was a doctor, I would never read what you are writing there, Dad"*. *"Most doctors won't either,"* I replied. *"But then it is totally useless!"* she said and continued with her own to-dos. When I started this PhD project, I was certainly aware of the fact that only a handful of people would read my thesis. However, my hope and motivation have been that my work will at least draw the attention of colleagues in pre-hospital medicine who have a desire to improve the quality of our work – and to give them the necessary tools for this task.

## 8 Background

### 8.1 Pre-hospital care

Pre-hospital severe illness or injury is a major cause of death and morbidity. The European Resuscitation Council has identified five critical conditions that require immediate pre-hospital management: cardiac arrest, severe respiratory failure, severe trauma, chest pain, and stroke. Four of these conditions are among the leading causes of death in the European Union [14]. In an observational study of Scandinavian physician-staffed emergency medical services from 2013, the pre-hospital incidence of severe illness or injury was estimated to be 25-30 per 10,000 person-years [12]. Many of these conditions benefit from interventions that rapidly correct deranged physiology and improve tissue oxygen delivery [15]. Beyond its impact on mortality, pre-hospital severe illness or injury can cause disability, sometimes lifelong. In 1982, Donald Trunkey published that in traumas, two patients remain permanently disabled for every death [16]. Today, 30 years later, more than three patients are permanently disabled for every death caused by road traffic accidents [17].

Services delivering pre-hospital critical care remain a critical link in the chain of survival for several frequent and life-threatening conditions. Therefore, Emergency Medical Services (EMS) are established in most countries [18-25]. Ground ambulances usually conduct transport. However, some EMS systems also use boats, planes, or helicopters, depending on the mission type, demography, and weather conditions. Traditionally, paramedics, emergency medical technicians, or nurses staff ground ambulances. In Helicopter EMS (HEMS), the crew differs between services. In some countries, the medical part of the crew consists of nurses or paramedics, while in other countries, the helicopters are staffed by physicians [20, 21, 26]. HEMS is a more expensive service than ground-based EMS systems and is found mainly in high-income countries.

In the Nordic countries, EMS systems serve a mixed rural and urban population. The population of approximately 27 million inhabitants is spread over an area of 1,369,000 km<sup>2</sup>, which creates a rather low population density (19.7 inhabitants/km<sup>2</sup>). In the most rural areas, patients travel long distances. For instance, if you live in Kautokeino, Norway, you will have a four-hour drive to the nearest hospital, which is located in Hammerfest. Hammerfest hospital is a smaller hospital, which means the journey will be even longer for patients who need a university hospital's facilities, such as percutaneous coronary

intervention, a neonatal intensive care unit, thoracic surgery, or neurosurgery. Furthermore, the sub-arctic climate – especially snow and wind during the winter season – can be rough in the Nordic countries. For EMS workers, this creates additional challenges, such as poor driving conditions or even closed roads, as well as ferries, helicopters, or fixed-wing planes that cannot operate due to the weather conditions.

## 8.2 Physician-staffed emergency medical services



Figure 1: Scene from Norwegian physician-staffed emergency medical service. Photo: Fredrik Naumann/Felix Features (with permission).

We define physician-staffed emergency medical service (P-EMS) as EMS staffed by physicians trained in emergency care exceeding the competency of a general practitioner on call. Depending on the country,

these physicians are often anaesthesiologists, surgeons, internists, or emergency physicians [18, 21]. P-EMS may be based on a ground ambulance platform, a helicopter platform, or a combination of platforms.

The delivery of emergency medical services in pre-hospital settings can be broadly categorised into Franco-German or Anglo-American models according to the philosophy of pre-hospital care delivery [27]. The Franco-German model is usually run by physicians who have extensive experience usually in anaesthesiology, emergency medicine, or critical care. The physicians have a wide scope of therapeutic measures, often including resuscitative thoracotomy, the induction of general anaesthesia, advanced diagnostic interventions, and haemodynamic monitoring and treatment. The philosophy is to bring the hospital to the patient; therefore, a “stay and play” approach is often seen in these services. The Anglo-American model is usually run by paramedics. The philosophy of this model has usually been a “load and go” strategy whose aim is to bring the patient to the hospital as quickly as possible. Pre-hospital interventions are more limited in this model. The Anglo-American model is found primarily in the USA and Canada. The Franco-German model is widely implemented in Europe, including the UK, Poland, Hungary, Italy, France, Germany, Greece, Austria, Switzerland, Denmark, Sweden, Finland, and Norway. In addition, Japan and Australia, as well as Russia and China, have implemented this model, at least partly. Thus, millions of people depend on P-EMS if they suffer severe illness or injury.

The first physician-staffed emergency medical service was established in Heidelberg, Germany in 1938 by the surgeon Professor Martin Kirschner. He stated that “the doctor should come to the patient, and not the patient to the doctor” [28]. He also called for the use of aircraft in the transportation of critically ill patients between hospitals. In Norway, P-EMS is well-established and the Norwegian civilian air ambulance service has been operating since 1978 – eight years after the first civilian HEMS was established in Germany [29]. It was actually German and Swiss physician-staffed HEMS that served as the model for the Norwegian HEMS [30]. The foundation of the Norwegian air ambulance service was led by Dr. Jens Moe, who established “Bård Østgårds Stiftelse”, later known as The Norwegian Air Ambulance Foundation. In 1988, 10 years after the first air ambulance service was established by this ideal foundation, The National Air Ambulance Service was formed. Today, the air ambulance service is government-funded with bases across the country, as depicted in figure 2. The majority of the P-EMS units are helicopters – some air ambulances and some search-and-rescue helicopters. Moreover, every HEMS base utilises a rapid response car for missions in the local catchment area, or for missions in which

a helicopter is not an option due to weather or technical conditions. The physicians staffing Norwegian P-EMS are anaesthesiologists, most of whom work in both P-EMS and hospitals [20, 13].

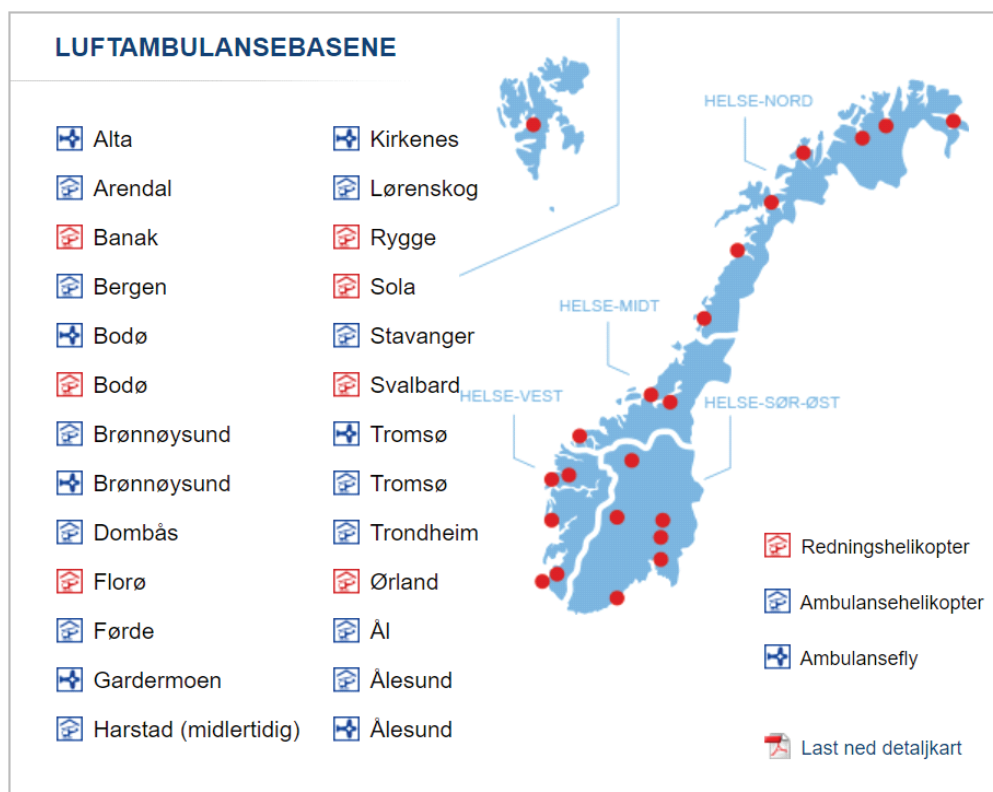


Figure 2: Air ambulance bases in Norway. (Reprinted from [www.luftambulansetjenesten.no](http://www.luftambulansetjenesten.no) with permission from Luftambulansetjenesten HF)

Although P-EMS is widely established in many countries, the benefit of this service, as opposed to paramedical service, is not clearly established. However, there is some evidence that physicians, in selected patient groups, improve outcomes or at least physiological variables [14, 31]. In a systematic review of the potential beneficial effect of physicians in pre-hospital care, Bøtker et al. demonstrated increased survival with physician treatment in trauma and, based on more limited evidence, cardiac arrest [32]. For respiratory diseases and acute myocardial infarction, there were indications of increased

survival. Despite the reports indicating a better outcome for defined patient groups treated by P-EMS, the studies include only a selection of the medical conditions regularly met by P-EMS; studies of P-EMS are few and often lack the scientific design needed to assess cause-and-effect relationships.

### 8.3 What is quality?

The word “quality” originates from the Latin “qualis”, which means “how” or “which”, and describes the significant characteristic of someone or something. However, the word is often used independently in a positive way, about things of high quality.

Quality is defined differently in business, health care, the service industry, etc. However, a common feature is that quality is often defined as how expectations are met. The expectations might be those of a customer, a patient, a provider, or other stakeholders. This definition of quality (Q), as a comparison between the expectations of a service (E) and the perceived performance (P), has given rise to the equation [33]:

$$Q=P/E$$

The control or quality assurance of a customer or controller is central in this equation. However, a more ambitious definition of quality belongs to the American industrialist Henry Ford:

*“Quality means doing it right when no one is looking”. – Henry Ford (1863-1947)*

As Henry Ford states here, quality should be a proper way of doing things every day, all the time – regardless of whether or not someone controls you. This view has a lot in common with that of Aristotle, who said:

*“Quality is not an act. It is a habit. Men acquire a particular quality by constantly acting a particular way ... you become just by performing just actions, temperate by performing temperate actions, brave by performing brave actions”. – Aristotle (384-322 B.C.)*

Based on the quotes of Henry Ford and Aristotle, quality is not a time-limited project in a service. It is a continuous job that should involve the whole organisation, from the top leader to the worker on the floor.

## 8.4 Quality measurement in health care

### 8.4.1 History of quality measurement in health care

In recent years, there has been an increased focus on minimizing errors and improving patient care. Quality measurement in health care has, thus, become widely recognised [1-4]. However, the beginning of quality measurement in health care took place in the 1850s, when Florence Nightingale demonstrated that basic hygiene standards reduced mortality among wounded soldiers in the Crimean War [34]. Nightingale emphasised the role of quality measurement in quality improvement:

*“The ultimate goal is to manage quality. But you cannot manage it until you have a way to measure it, and you cannot measure it until you can monitor it”.*

At the beginning of the 20<sup>th</sup> century, Ernest Amory Codman, a surgeon from Harvard Medical School and the Massachusetts General Hospital, kept track of his patients via “End Result Cards”, which contained information about the patients’ diagnoses, treatments, and outcomes. He believed that the key to improving care for future patients was to understand why treatments had been unsuccessful in the past.

In 1966, Avedis Donabedian introduced a concept and a terminology of categorizing quality measures into three groups: structure, process, and outcome measures [35]. His framework is still used in today’s quality measurement.

In 1999, the central report “To Err is Human: Building a Safer Health System” was published by the Institute of Medicine (IOM) [36]. This report identified large safety gaps in U.S. health care, concluding that approximately 90,000 patients die every year in U.S. hospitals due to medical errors. Another report by the IOM was published in 2001: “Crossing the Quality Chasm: A New Health System for the 21<sup>st</sup> Century” [37]. The report further described the lack of high-quality care for all patients in U.S. health care. Together, the two IOM reports have led to an increased public demand for quality measurement and improvement in health care.

### 8.4.2 Definitions of quality in health care

Quality in health care has been defined somewhat differently. The World Health Organisation (WHO) states that “quality of care is the level of attainment of health systems’ intrinsic goals for health improvement and responsiveness to legitimate expectations of the population” [38]. A more intuitive

definition comes from the National Health Service in the United Kingdom, which states that high-quality health care should be “clinically effective, personal and safe” [39]. However, one of the most-cited quality definitions is the definition formulated by IOM: “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge”. IOM further described six dimensions of quality care: a care that is safe, effective, patient-centered, timely, efficient, and equitable [37], c.f. figure 3:

**Person-centeredness:** providing care that is responsive to individual personal preferences, needs, and values and assuring that patient values guide all clinical decisions;

**Safety:** avoiding injuries to patients from health care that is intended to help them;

**Effectiveness:** providing services based on scientific knowledge;

**Efficiency:** avoiding waste, including the waste of equipment, supplies, ideas, and energy;

**Equity:** providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, or socio-economic status; and

**Timeliness:** reducing waits and sometimes-harmful delays for both those who receive care and those who give care.

These six dimensions of quality should, as far as possible, be measured when one is assessing quality in health care.



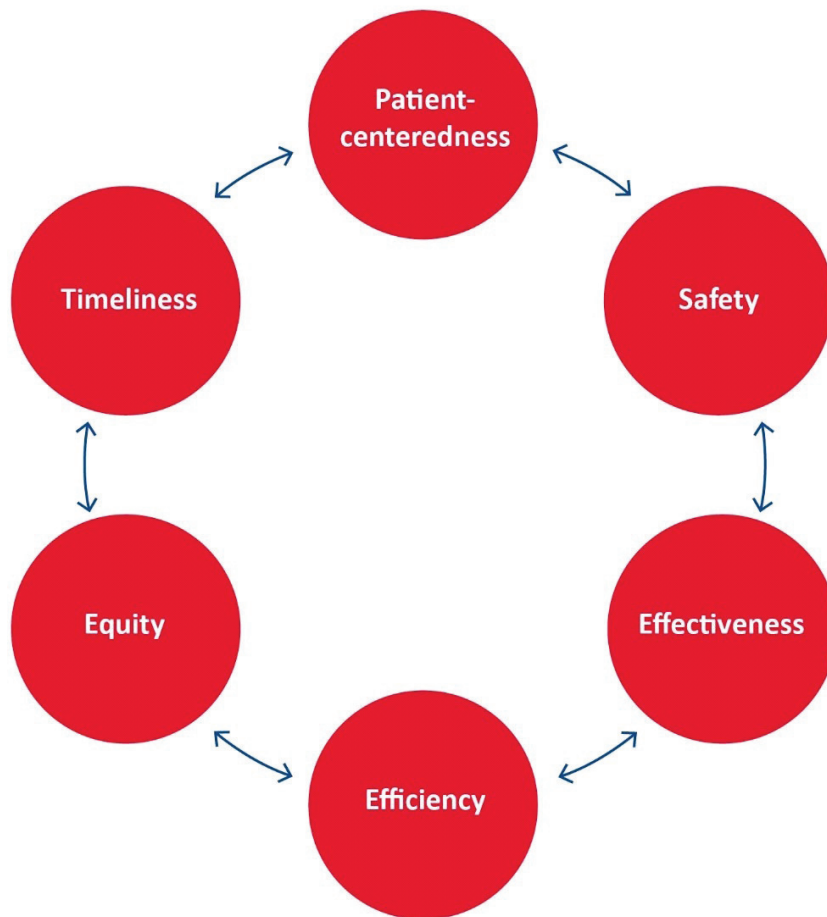


Figure 3: The six quality dimensions in health care.

As this thesis stated earlier, the number of reports, papers, and projects pertaining to quality is increasing. The augmented use of the word “quality” in these settings may, of course, be due to the fact that more attention is being paid to this topic. However, another possible explanation might be that the term “quality” now has a wider embrace than it did before. Thus, almost every aspect of health care can be regarded as related to quality. A third possibility is that quality has been reduced to a fashionable term that is uncritically used by those who try to boost the legitimacy of their projects.

A separate term that is often used together with “quality of care” is “value of care”. What is, then, the difference between quality and value in health care? The ultimate goal in health care, including P-EMS, should be the achievement of the highest possible value for the patient, with value defined as the health outcome divided by costs [1]. Costs include the actual use of resources involved, including personnel, facilities, and supplies. In principle, value in health care should be measured by the outcomes achieved, not by the volume of services delivered. As depicted in figure 4, the value equation illustrates that increasing quality while maintaining resource expenditure (cost) will result in increased value. Conversely, reducing costs without regard to the outcomes can create “false savings” and limit the effectiveness of the care provided.

A red rectangular box containing the title "The Healthcare Value Equation" at the top. Below the title is the equation: Value = Quality / Cost. The words "Value", "Quality", and "Cost" are in a large, bold, white sans-serif font. The equals sign and the division line are in a smaller white font. The entire equation is centered within the box.
$$\text{Value} = \frac{\text{Quality}}{\text{Cost}}$$

Figure 4: Defining value in health care.

#### 8.4.3 Conceptual framework

For purposes of this thesis, we used the recognised framework described by Donabedian, which groups QIs into three broad categories: the structure, process, or outcome of health care [35, 40]. Structure indicators describe the infrastructure of a health care system, such as the competence of the staff, the available equipment, and deployment and response times. Process indicators evaluate the care provided to the patient, whereas outcome indicators address the change in the patient’s health status as a result of the provided care. Each type of QI will not provide a complete description of the quality of care but, rather, addresses a component of the care. Thus, different types of QIs should be combined when one is

estimating the quality of a service [5]. Moreover, as stated by the Institute of Medicine, all six quality dimensions should be covered when one is measuring the quality of a health care service. This is necessary to obtain a comprehensive assessment of the total quality of care. When the Donabedian framework is combined with the six quality dimensions of the Institute of Medicine, the result can be depicted as in figure 5. A quality measure can be categorised into one of 18 sections in this figure. This should be interpreted as a description of the conceptual framework, not as a necessary recipe for quality measurement. Even in a thorough quality measurement, it is not likely that all 18 sections will be represented by a quality measure.

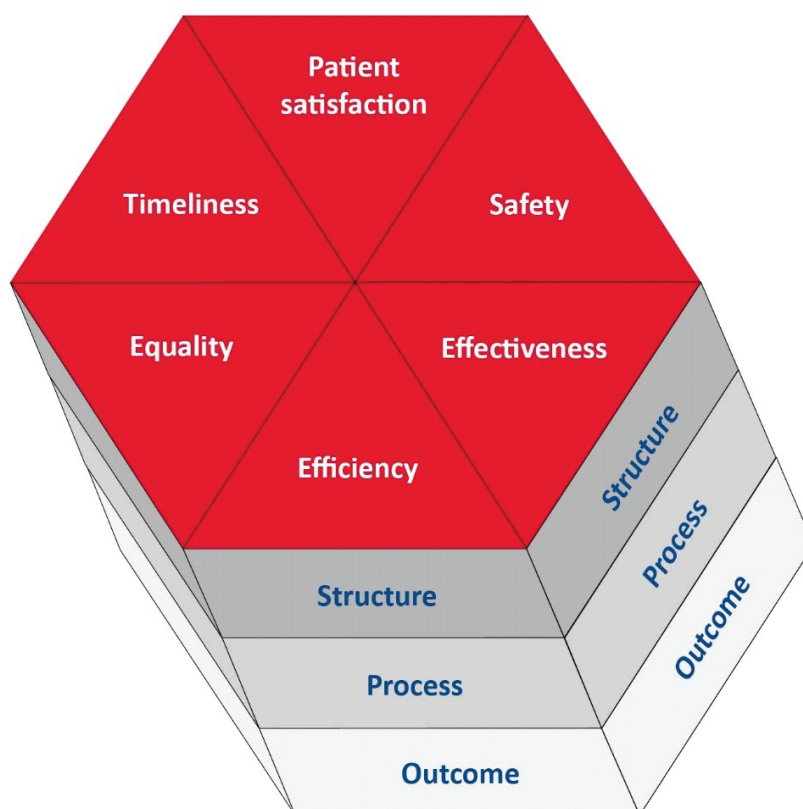


Figure 5: Conceptual framework for multidimensional quality measurement in P-EMS.

Quality improvement is typically done in so-called improvement cycles, also known as Plan-Do-Study-Act (PDSA) cycles (figure 6) [41]. An improvement cycle represents a framework for deciding what to test (Plan), carrying out the test (Do), observing and learning from the consequences (Study), and determining what changes should be made to the test (Act). Although it is possible to implement the quality improvement cycle once, a single cycle improvement isn't quality improvement in the purest sense. This is due to the lack of continuous attention on the evaluative "study" step, which is critical for successful quality improvement. Therefore, proper quality improvement requires constant measurement and evaluation.

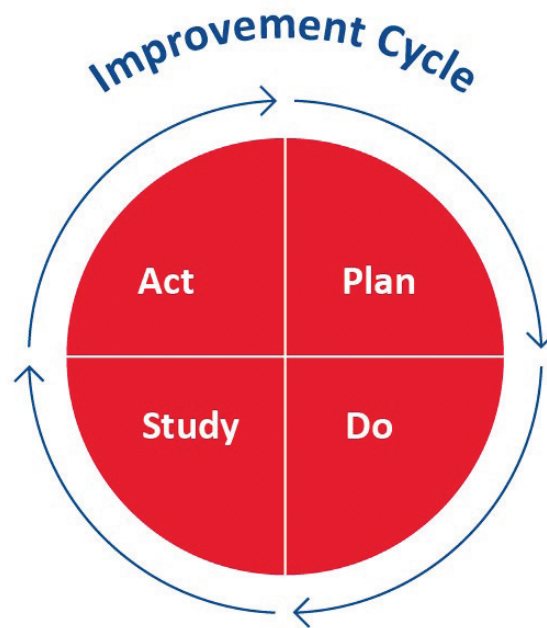


Figure 6: Plan-Do-Study-Act improvement cycle.

(Based on the presentation of the PDSA cycle in "The Improvement Guide" by Langley et al. [41].)

#### 8.4.4 What is a quality indicator?

A quality indicator is a tool for quality measurement. Quality measurement can be defined as measuring the extent to which set targets are achieved [5]. A quality indicator (QI) is used to measure performance against these targets, which normally are a recognised standard of care. Standard of care is defined as care that is delivered in accordance with clinical practice guidelines or other evidence-based care protocols [42]. Quality indicators are expressed as numbers, rates, or averages and are instrumental in helping clinicians, organisations, health care managers, and societies achieve improvements in health care quality [5]. They can be measures of structure, process, and outcome, either as generic measures relevant for all diseases or as disease-specific measures that describe the quality of patient care related to a specific diagnosis. Further, QIs should integrate the best research evidence with clinical expertise and patient values [43] and allow for the measurement of health care quality by creating a quantitative basis that indicates performance.

Ideally, all QIs should be based on rigorous scientific evidence of their relevance and importance [5]. However, this is not always achievable in health care, including P-EMS, where evidence is often absent. This necessitates the use of other methods for the development of QIs.

To identify potential QIs, a widely used method is a combination of a systematic review of current literature and some sort of systematic collection of expert opinions, e.g., the use of an expert panel in a consensus method – for instance, a Delphi technique. Thus, available evidence is combined with expert opinions. In the development of QIs, the expert panel should consist of people who are considered experts in the appropriate area and who have credibility among the target audience [44]. Clinical expertise is typically represented by clinicians, scientific expertise by researchers, and user expertise by patients. The experts and stakeholders evaluate the evidence for possible QIs and eventually define the final QIs [42].

As described, ensuring that a QI is properly developed is important. Furthermore, the number of QIs chosen for quality measurement in a service is important. In short, there must be enough QIs. If too few QIs are used, the quality measurement will be narrow and not comprehensive enough for assessing the total quality of the actual service. On the other hand, registration fatigue is a risk if one chooses too many QIs. This may eventually lead to poor data quality and meaningless quality measurement. Therefore, a golden rule is to use as few QIs as you dare but as many as you need.

#### 8.4.5 Challenges in quality measurement in health care

Quality measurement is something other than comparing a new blood pressure drug with a placebo. It is not very accurate, objective, or controlled. Hence, the answers in quality measurement can be only estimates but they are important estimates because they answer important questions. A number of specific challenges arise in quality measurement. A brief presentation of the most important ones follows.

As discussed more thoroughly elsewhere in this thesis, hard clinical endpoints matter the most but are not necessarily the optimal quality indicators. This is primarily due to the exposure to many care units as well as a mortality rate that may be too low in some patient groups. The latter illustrates an important challenge, i.e., one must choose quality indicators that measure events occurring with a certain frequency. The success rate for establishing a surgical airway, for instance, would be a poor quality indicator for P-EMS because it happens so rarely. Further, quality measurement must be so simple that it can be included as part of everyday registration, yet, at the same time, so detailed that it can tell us something about the quality of the work.

There is always a methodological trade-off when one is launching a quality improvement project: On the one hand, you should aim for systematic development and the selection of quality indicators to ensure that you get reliable and valid measures. On the other hand, you should not let such a systematic approach hinder all projects that do not meet this standard [45]. To start a quality improvement project is more important than to perfectly select quality indicators for the project. When selecting your quality indicators, seek usefulness rather than perfection [46]. By completing quality improvement projects, one identifies what should be done differently in the next project. Therefore, it is advisable to launch a quality improvement project while also recognizing that it has limitations and might be further developed.

When one is aiming for quality improvement, a necessary first step is to measure quality. It is easier to improve something that you can measure. However, Dr. W. Edwards Deming once said: *“It is wrong to suppose that if you can’t measure it, you can’t manage it – a costly myth”*. Thus, even though quality measurement is a highly valuable tool for quality improvement, it is still possible to achieve quality improvement in those aspects of a service that are difficult to measure. The fact that the quality of an activity in your service is difficult to measure is no excuse for not trying to improve the quality of that activity.

## 8.5 Quality measurement in pre-hospital critical care

The literature on QIs in pre-hospital critical care is scarce [7, 47] and there is no international agreement on a conceptual framework or choice of QIs for P-EMS. Appropriate QIs are needed to identify both high-quality care and areas with room for improvement in care.

### 8.5.1 Different perspectives on quality in health care

The quality of P-EMS is obviously related to the achieved quality for the patients involved. Moreover, the quality of the service is related to benefits for systems – nationally, regionally, and locally, as illustrated in figure 7. Different stakeholders have different perspectives on what represents quality in health care; therefore, various QIs are possible [48, 49]. It is plausible that this would also be the case for P-EMS specifically.

Examples of quality for patients might be pain relief, early diagnosis, reduced mortality or morbidity, and other outcome variables. Quality is a result of different factors depending on the patient's condition and the related needs in the specific situation. In several P-EMS missions, the potential quality could be due to the presence of a qualified physician. These could be missions involving advanced medical procedures like thoracostomy or endotracheal intubation (ETI). The presence of a qualified physician is also necessary for the administration of individualised therapy beyond the instructions given in guidelines, or for avoiding unnecessary or futile treatment.

In other types of missions, the potential quality might be the result of factors other than the presence of the physician. For instance, these might be missions in which the medical treatment per se does not require the competence and skills of a physician but in which significantly reduced transport time, due to helicopter transport, represents quality to the patient. Bringing patients with ST-segment elevation myocardial infarction (STEMI) from rural areas to timely percutaneous coronary intervention (PCI) would be an example of this. Another example would be the transportation of a patient who is suffering a stroke and who is a potential candidate for in-hospital thrombolytic therapy. Patients suffering STEMI or stroke are cared for on a regular basis by the traditional EMS, and, normally, the pre-hospital treatment does not require the competence of a physician. However, these patients require time-critical advanced hospital treatment, which is crucial to a favourable outcome. A mode of transportation that significantly reduces the time to treatment will, therefore, be of importance. In terms of reducing the time to treatment even more for stroke patients, an ambulance equipped with computer tomography and staffed with an anaesthesiologist has been demonstrated as being feasible and effective in Norway [50,

51]. Also, for trauma patients, rapid transportation to trauma hospitals is achieved by P-EMS in many cases, especially in rural areas. In fact, P-EMS is often the only way to give patients from rural areas in Scandinavia timely access to high-level trauma care, PCI, etc. Sometimes, potential quality for the patient might be related to the operational capacities of a helicopter, such as missions in which lifting the patient out of inaccessible terrain or a hostile environment (by a long-line rescue technique) is necessary. As such, the technical rescue skills of a HEMS-crew member could, indeed, produce quality for the patient in selected cases [52].

In addition to the quality produced for single patients, P-EMS can produce quality for systems. Quality for local systems is linked primarily to the EMS systems and hospitals. Today, many treatment options are centralised in university clinics and leading centers. P-EMS might contribute to better logistics and economy by transporting patients to these centers. Therefore, faster transport to specialised centers produces quality not only for the patient but also for EMS systems, which achieve faster turnaround and greater availability for transportation needs. In addition, hospitals can benefit from the flexibility of HEMS-transportation; an interfacility transport (i.e., secondary transport) can relieve pressure on an overcrowded intensive care unit or provide early diversion to an appropriate facility. Moreover, P-EMS could be beneficial in unusual and sometimes extreme situations, in which the helicopter's flexibility and capabilities make it a special resource. For instance, the air ambulance can provide quick and flexible transport when a medical expert (other than the accompanying physician) is needed in the pre-hospital setting [53]. In situations in which avalanches or mudslides have blocked roads and isolated local communities, helicopters may be one of few options when patients require transport to a hospital.

Quality for regional systems might be in the form of the involvement of P-EMS-physicians in the leadership and education of traditional EMS and Emergency Medical Coordination Centers (EMCC) as well as their involvement in research. An operational quality might be the presence of P-EMS in major incidents, during which additional competence and capacity are brought to the trauma scene [54-56]. During the Underground bombings of London in 2005, London's Air Ambulance flew at least 25 missions in which the main task was to bring medical teams to the trauma scene. Because of the chaotic situation on the streets of London at the time, the helicopter's contribution was later judged to be an important part of the emergency response [57].

Contributing to the safety of the population, P-EMS also produces quality for national systems. Furthermore, a central principle of Norwegian health legislation is that citizens should have equal rights



to health care – a principle of equity of access [58]. In Norwegian emergency medicine, it is not realistic to achieve this goal, as there is a substantial difference in the health care provided in rural and central areas. However, utilisation of P-EMS can help reduce these differences. Thus, P-EMS can contribute to the reaching of defined political goals.



Figure 7: Examples of quality produced by P-EMS. The different boxes illustrate what could be considered achieved quality for the patient and for systems: local, regional, and national, respectively.

#### 8.5.2 The use of time variables

Pre-hospital services evolved in response to the need for acute health care for certain time-critical conditions (war injuries, cardiac arrest, major trauma) [8, 59]. Further, in 2003, a European project defined “The First Hour Quintet” [60]. This is a selection of five critical conditions (cardiac arrest, respiratory failure, trauma, acute coronary syndrome, and stroke) that are of great importance in pre-

hospital care. For these conditions, immediate and appropriate pre-hospital care and transport to a hospital are critical.

When goals are set to measure the quality of pre-hospital care, time variables are typically used [7, 59]. The importance of reducing the time to definitive treatment for stroke and myocardial infarction is well-documented [61, 62]. For trauma patients, on-scene time seems to increase mortality, though not in all settings and conditions [63]. For some patients, however, increased on-scene time is deliberately chosen to provide beneficial treatment and promote stabilisation of the patient [64, 65]. However, interventions should be limited to those that increase the likelihood of survival and that reduce morbidity [66]. Despite the possibility of more advanced interventions, the presence of a pre-hospital physician does not necessarily increase on-scene time [67-69].

Examples may illustrate the limitation of time variables as quality indicators in P-EMS. If a P-EMS unit is bringing blood products from a hospital to a trauma scene, this may prolong response time. However, in a certain situation, the total quality for the patient could be greater than if the P-EMS arrived some minutes earlier but without blood products. Another example: A mountaineer is traumatised with a spinal injury and neurogenic shock after suffering a fall. Packing the patient well to prevent further hypothermia and placement of an arterial line followed by haemodynamic stabilisation using fluid administration and vasoactive medications will benefit the patient, even though the time spent on the scene is increased compared to a "load and go/scoop and run" strategy. A further example: Performing an ultra-sound scan of the traumatised patient may slightly prolong the on-scene time but if it results in a change of treatment strategy, it can be well worth the extra time spent on scene [70]. Depending on the patient's condition and the mission's circumstances, other quality dimensions may be more relevant than time consumption. What really benefits the patient is not the fact that he is put in a helicopter or admitted to a hospital but that deranged physiology is corrected and that delivery of oxygen (DO<sub>2</sub>) to tissues is improved. Thus, the measures performed on scene, and their results, are often more relevant to the patient than is focusing only on time consumption. Moreover, due to the mixed patient population in Nordic HEMS, many patients do not necessarily suffer from highly time-critical conditions. This, again, means that for a certain proportion of our patients, shorter response or reaction times will not improve survival or other significant outcomes [64, 65]. A unilateral focus on time variables will not be a relevant quality measurement for these patients. However, time variables should

be included as some of the multiple quality measures when the quality of pre-hospital care should be assessed.

#### 8.5.3 The use of outcome variables

As described earlier in this thesis, modern quality improvement methodology often highlights three types of indicators: structure, process, and outcome. A simplified overview of the three categories, including examples of quality indicators as they can be used in an emergency medical service, is depicted in figure 8. Outcome indicators address the change in the patient's health status as a result of the provided care. The outcome can be seen as the sum of factors relating to the patient, the illness, the treatment, and the organisation (figure 9). Health care workers can influence only the last two factors. Therefore, it is necessary to adjust for factors pertaining to the patient and the illness (case-mix) if one is comparing the outcome of different health services. The most-used primary endpoint is functional survival to hospital discharge [57, 71]. This is the main endpoint in most HEMS studies. However, it can be a problem if patient numbers are small and if overall mortality in the HEMS population is low. Second, inference problems can occur because HEMS-triaged patients tend to have a higher severity of disease than do ground-transported patients. This must be adjusted for in comparative studies of HEMS and traditional EMS, or the groups must be matched.

Indicator type	Definitions	Examples of quality indicators
Structure	Characteristics of the different components of the system	<ul style="list-style-type: none"> <li>Facilities</li> <li>Equipment</li> <li>Staffing</li> <li>Knowledge base of providers</li> <li>Deployment</li> <li>Response times</li> </ul>
Process	Patient care intended to improve patient outcome	<ul style="list-style-type: none"> <li>Adherence to medical protocols</li> <li>Correct medication administration</li> <li>Transport to appropriate facility</li> </ul>
Outcome	Changes in health and well-being	<ul style="list-style-type: none"> <li>Out of hospital cardiac arrest survival</li> <li>Patient satisfaction</li> <li>Improvement in pain score</li> </ul>

Figure 8: Quality indicators in the quality measurement of EMS.

(The figure is based on the Structure-Process-Outcome Model for EMS systems [59].)

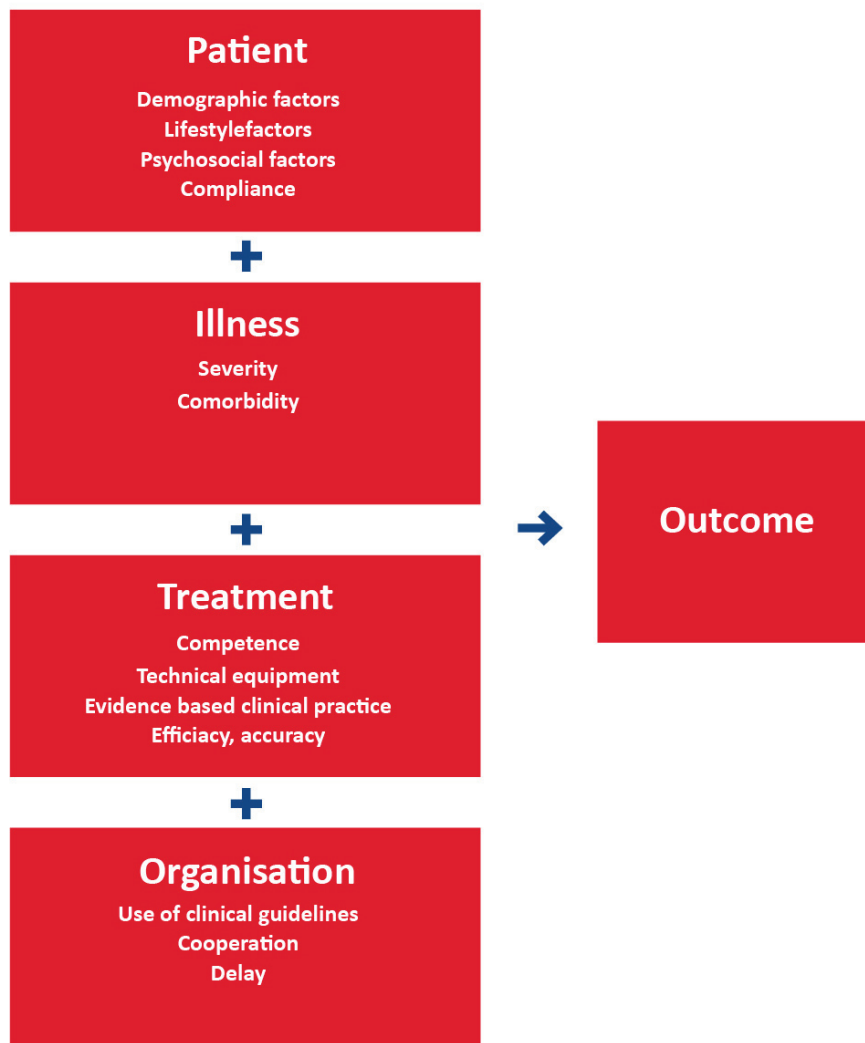


Figure 9: Conceptual framework for factors influencing the outcome of care [5].

A challenge pertaining to the outcome indicators is that the patient is exposed to many care units and a significant amount of treatment in the period from the P-EMS care until the time of discharge or death. All of these units influence the patient’s chances of survival. To isolate the contribution of P-EMS, or any other involved unit, during the patient’s treatment and convalescence is difficult [72]. This can be a challenge when one is trying to identify the outcome effect of care rendered in each care unit. A way to solve this problem is to conduct stepwise risk adjustment as the patient moves through different care units. This enables the identification and measurement of the effect of therapeutic interventions rendered in each care unit and is illustrated in figure 10. When care is given to a patient by different providers and in different places, eventually making up a chain of events, they are called complex interventions [73]. In pre-hospital care, complex interventions are common. The evaluation of complex interventions is difficult due to problems involved in developing, identifying, documenting, and reproducing the intervention. The number of clinical events is often high [74], leading to many possible outcomes. To capture this complex set of outcomes, a comprehensive approach is necessary for describing the quality of patient care. When one focuses on too few outcome measures in these contexts, important information may be lost.

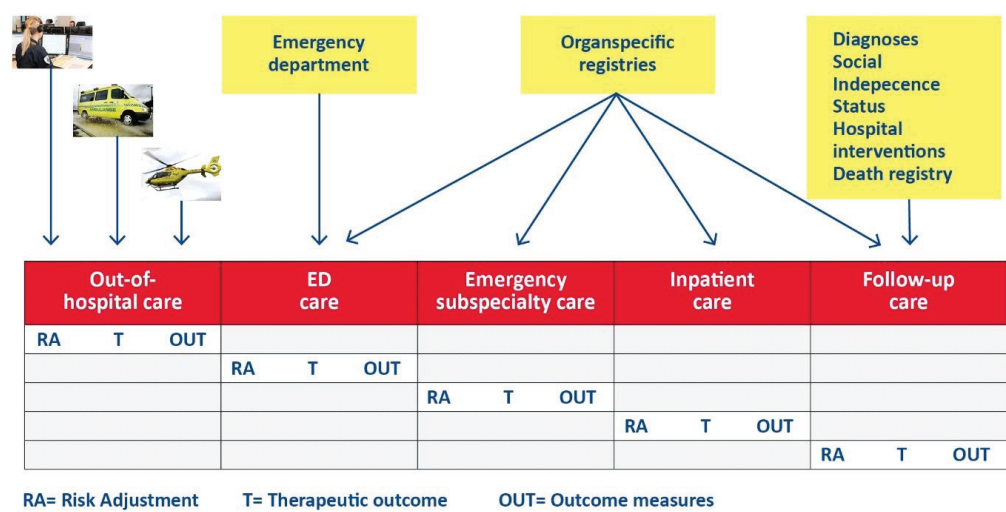


Figure 10: The episode of acute care. Based on a figure by Spaite et al. [72].

Another approach to the challenge of several care units all influencing the chance of survival is to choose other quality indicators that might measure the actual process of care rather than the outcome of it (e.g., adverse events, the success rate of critical procedures, adherence to guidelines, etc.). If one is aware of the limitations of such quality indicators, and if one checks their validity and interprets the answers with care, these might be useful quality indicators. Moreover, it might be appropriate to evaluate the plausibility of good processes leading to good outcomes, as is believed in other high-risk environments like the oil industry, aviation, and nuclear power stations. In these industries, hard endpoints are often inappropriate quality measures. For instance, it would be a very poor measure of quality in a nuclear power station to record the number of melt-downs per year, as a melt-down should never happen. With that in mind, it is necessary to measure other everyday work aspects that are believed to predict a good outcome. In the tradition of medical research, hard clinical endpoints matter the most. However, they are not necessarily optimal quality indicators.

While mortality is the most widely used outcome measure, morbidity is another primary endpoint used in the literature. Quality-of-life outcomes and the Glasgow Neurological Outcome Score are relevant examples.

Also, secondary or surrogate outcome variables are in use. These variables can measure indicators known or assumed to indirectly reduce mortality. These variables are often physiologic parameters. For instance, the improved outcome of head-injured patients receiving HEMS airway management has been related to an improvement in physiological variables such as SpO<sub>2</sub> and end-tidal CO<sub>2</sub> [75, 76].

Relief of pain is an outcome traditionally reported as a secondary outcome. However, it has been suggested that this variable is of greater importance than previously recognised. Therefore, it has been considered that pain relief should be defined as a stand-alone (i.e., primary) outcome for pre-hospital care [77].

Another surrogate endpoint is earlier access to pre-hospital critical care. Although the association between earlier pre-hospital critical care and an improved outcome is not verified, many experts believe that the time to EMS intervention is important for optimizing treatment for severely ill or injured patients [57].

Finally, a surrogate endpoint is the level of critical care brought to the patients. HEMS crews can provide advanced interventions like thoracostomy and ETI. In the pre-hospital setting, this represents another

level of care compared to ordinary EMS. The knowledge and capabilities of air medical crews may also result in a higher level of care for small community hospitals treating critically ill or injured patients. Brown et al. concluded that for trauma patients in the USA, one of the major advantages of HEMS, indeed, is the higher level of care that this service provides [78].





## 9 Motivation for the thesis

In my first year as a HEMS doctor, the HEMS crew member and myself were dispatched to a road traffic accident quite close to our base. After only a few minutes' ride in our Rapid Response Car, we arrived at the scene, where a young couple had crashed into a rock wall. No other cars were involved. We quickly confirmed that the man was deceased. However, the girl was alive but unconscious. After a very short time, she went into cardiac arrest. With the help of skilled firemen, we evacuated her from the car wreck. Intubation, fluid resuscitation, and bilateral thoracostomies were done under ongoing CPR. After a while, she experienced a return of spontaneous circulation. She arrived at hospital alive but died within the first hour due to a cervical fracture with a serious spinal cord injury.

After this job, I asked myself whether it would be possible for me or anyone else to assess the quality of our on-scene efforts. At the time (and still so in many pre-hospital services), the most frequently used variables to assess quality were time variables and mortality. If our work was to be assessed in terms of response time, we had done a good job, as we had arrived after only a few minutes and had experienced no delays in our response. However, for the patient, the ultimate quality measure is survival. A dead patient finds no comfort in short response times. Carrying that, if we use mortality to assess the quality of our job, how would our efforts be evaluated? For the dichotome variable of survival, we ended up with a dismal result. Does that automatically mean that the quality of our efforts was poor? This was how I started thinking more systematically about the difference between quality and outcome. The query sparked my engagement in developing a method for quality estimation in P-EMS.

Most important is the fact that quality improvement exists for the benefit of the patient. By measuring and improving the quality of our work, we may offer our patients even better care than we can today. At the same time, we must ensure that we use the limited P-EMS resources in a way that maximises the achieved quality so that those in real need of P-EMS get the help they require, whilst those not in need of P-EMS competence and capacity are cared for by other skilled pre-hospital providers. Accordingly, quality improvement may ensure that our care is provided in a way that is just and that makes sense medically. Overtreatment is one of the biggest problems in modern medicine in high-income countries; we live in "the time of too much" [79]. Unnecessary management of patients with low risk or no benefit is defensive medicine and has high financial costs [80]. It certainly decreases the value of health care. By

increasing the quality of our services, we may oppose this development, as increased quality will increase the value of provided health care.

## 10 Aims

The overall aim of this thesis was to develop a method measuring quality in physician-staffed emergency medical services. The specific aims of the four papers were as follows:

### Study 1

To identify, describe, and evaluate studies of quality measurement in physician-staffed emergency medical services.

### Study 2

To develop a set of multidimensional quality indicators for physician-staffed emergency medical services.

### Study 3

To test the multidimensional quality indicators for important characteristics in a real clinical setting. We further aimed to propose benchmarks for future quality measurement in physician-staffed emergency medical services based on the data in this study.

### Study 4

To determine whether there is a relationship between the quality indicators developed for physician-staffed emergency medical services (in study 2) and 30-days mortality.



## 11 Materials and methods

To answer the questions posed, a systematic literature review was done in study 1. Study 2 was done using a modified nominal group technique. Study 3 (papers III and IV) was a prospective observational study. Studies 2 and 3 were international collaborations. No interventions were performed in any of the reported studies.

### 11.1 Study setting and descriptions

#### 11.1.1 Paper I

For purposes of this review in particular, and the PhD project in general, we defined P-EMS as emergency medical services staffed by physicians trained in critical care, exceeding the competency of an on-call general practitioner [81]. Moreover, we defined the term “pre-hospital” as relating to procedures administered or care provided prior to patient arrival at the hospital. These P-EMSs were the setting of interest in all studies. In study 1 we performed a literature search for studies on quality measurement in this setting. We did not limit the studies of interest to specific areas of the world. Nor did we limit the studies to specific languages. PRISMA guidelines were followed and the study was registered at PROSPERO prior to study instigation (<http://www.crd.york.ac.uk//prospero/>, registration number CRD42015024421).

#### 11.1.2 Paper II

In this study, named the EQUIPE project, we tasked an expert panel with developing QIs for P-EMS using the modified nominal group technique [44, 82]. P-EMS was defined as in study I. The QIs had to be feasible to collect during the pre-hospital time interval or in the ED at hand-over. Furthermore, the QIs had to, as much as possible, cover the six quality dimensions that define high-quality care, stated by IOM [37] and appreciated by WHO [3]. As described in chapter 8.4.2, these six quality dimensions are timeliness, safety, efficiency, equity, effectiveness, and patient-centeredness.

#### 11.1.3 Papers III and IV

In this prospective observational study, 16 physician-staffed helicopter emergency services in Finland, Sweden, Denmark, and Norway registered data for the EQUIPE quality indicators. Significant system

similarities had previously been documented in the P-EMS of the four participating countries, making them a suitable arena for multi-center studies [13]. All services engage in both primary responses, while the Swedish, Danish, and Norwegian services also engage in secondary responses. The former are defined as responses in which the patient is located outside a hospital, while the latter are inter-hospital transfers. Moreover, the Norwegian services engage in search-and-rescue responses (SAR responses). In addition, one Swedish (Karlstad) base, and all Finnish and Norwegian bases, employ a rapid response car for responses close to the base and for responses in poor weather conditions that prevent flight operations.

#### *11.1.3.1 Inclusion criteria for papers III and IV*

The response-specific quality indicators should be registered for all P-EMS responses. Thus, both primary and secondary responses should be included. For P-EMS bases utilizing both helicopter and a rapid response car, all responses should be included, regardless of the mode of transportation. Inquiries to dispatch the P-EMS unit but resulting in no response or no patient contact should also be included (except inquiries with *counselling as the only purpose*).

#### *11.1.3.2 Exclusion criteria for papers III and IV*

Inquiries made to the P-EMS crew with counselling as the only purpose of the contact should not be included in the study. However, every contact with the dispatch center with the purpose of dispatching the P-EMS-unit should be included.

## 11.2 Participants

### 11.2.1 Paper I

The papers included in the systematic review in paper I were all the identified articles describing quality estimation in P-EMS. Articles were included in the systematic literature review if they fulfilled all the following criteria: 1) literature describing methods of quality estimation in P-EMS, i.e., the use of one or more QIs based on quantitative methods, qualitative methods, or both; 2) original manuscripts; and 3) literature published after January 1<sup>st</sup>, 1968 and until the date of the literature search, October 5<sup>th</sup>, 2016. The rationale for including literature from January 1<sup>st</sup>, 1968 is the fact that the world's first civilian physician-staffed helicopter emergency medical service was established in Munich that year [83]. The search strategy identified a total of 4,699 articles. A total of 27 articles from the database search were

included for data extraction and quality appraisal. A review of the literature lists of included articles did not result in additional findings. Figure 11 depicts the flow of information through the different phases of the systematic review.

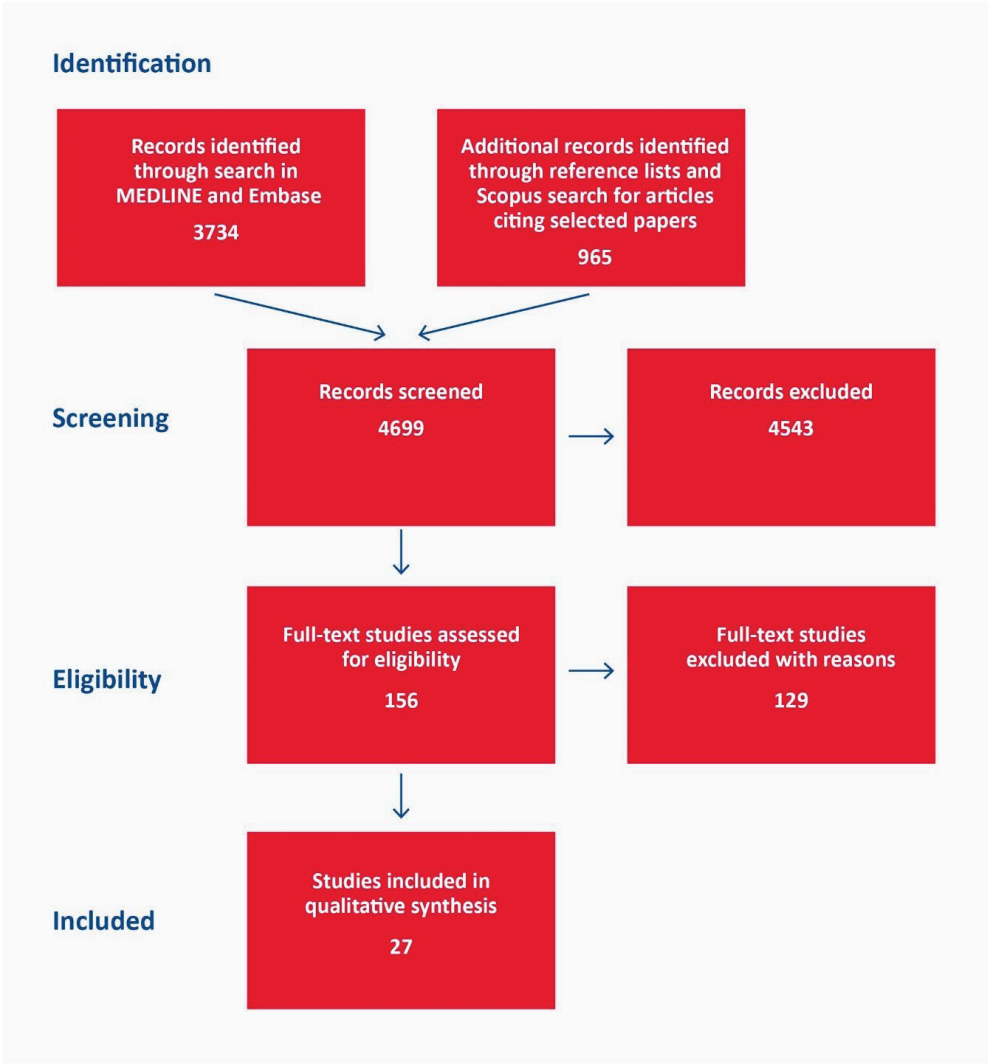


Figure 11 (Figure 1 in paper I): Information flow through the different phases of the systematic review.



### 11.2.2 Paper II

When developing QIs, the expert panel should consist of people who are considered to be experts in the appropriate area and who have credibility among the target audience [44]. Clinical expertise is represented by physicians, scientific expertise by researchers, and user expertise by patients.

Accordingly, this study's expert panel consisted of clinicians and researchers from different P-EMSs, as well as of stakeholders representing other perspectives in P-EMS. More specifically, the 18 members of the expert panel consisted of three general practitioners, two P-EMS medical directors, the director of a public health institute, a specialist in community medicine, a patient-organisation leader, and 10 physicians working in P-EMS. All panel members were, in different ways, considered experts in P-EMS or in collaborating services of P-EMS and practiced in Australia, Austria, Denmark, Finland, Norway, Scotland, the UK, and the USA.

### 11.2.3 Papers III and IV

We aimed to recruit both P-EMS services operating in rural areas as well as more urban services. The compliance and motivation of recruited P-EMS bases were considered very important by the study group.

In Sweden, the following P-EMS bases were recruited:

- Uppsala
- Karlstad

In Finland, the following P-EMS bases were recruited:

- Tampere
- Vantaa
- Turku
- Oulu
- Kuopio

In Denmark, the following P-EMS bases were recruited:

- Ringsted
- Skive
- Billund

In Norway, the following P-EMS bases were recruited:

- Trondheim
- Ørland
- Arendal
- Stavanger
- Lørenskog
- Rygge

### 11.3 Measurement and data collection

#### 11.3.1 Paper I

Prior to the literature search, the authors created templates for data extraction and quality appraisal. These templates were included in the PROSPERO registration before the literature search was conducted. The data extraction and quality appraisal variables were based on the authors' assumptions about what is important to report in quality measurement studies in P-EMS. However, these variables do not represent a reference standard, as, to the best of our knowledge, such a standard does not exist. As part of data extraction, fixed-system variables were included. Fixed-system variables relate to system characteristics concerning the organisation, staffing, and operational capacities of the service and are necessary for interpreting the results [81].

### 11.3.2 Paper II

In this study, the expert panel developed the QIs through a four-step modified nominal group technique. Stages 1, 2, and 4 consisted of e-mail correspondence. In stage 3, the expert panel gathered for a two-day consensus meeting in Oslo, Norway. Stages 1 and 2 were anonymous steps in the process, as identifying the proposals and rankings of the individual panel members was impossible for the rest of the expert panel. This information was available only to the project group.

Stage 1. The members of the expert panel were asked to propose QIs for P-EMS according to the following predefined instructions: QIs should be proposed for each of the three categories of QIs (structure, process, and outcome) and should be obtainable during the pre-hospital time interval.

The panel members returned their proposals to the project group via e-mail. The QIs within each category (structure, process, and outcome) were ranked according to the number of experts who had included each QI in their proposals.

Stage 2. The experts were asked to use the revised spreadsheet to rank the 10 most important QIs in each of the three categories (structure, process, and outcome). In each category, the quality indicator that was ranked in first place was given a point value of 10, while second place received a value of 9, third place received a value of 8, and so forth, until the tenth-place indicator was given 1 point. The point values from all panel members were added, and quality indicators with no ranking were removed from the list. The list with the remaining quality indicators, prioritised according to achieved point value, formed the basis for the consensus meeting.

Stage 3. The expert panel gathered for a two-day consensus meeting in Oslo, Norway. A moderator led the experts through discussions of the QIs in the spreadsheet developed during stage 2. The experts decided which QIs should be included in the final set. Further, preliminary definitions and limitations were established. All debates and discussions were plenary.

Stage 4. Based on the results of the consensus meeting, the project group prepared a document with the selected QIs, including definitions. This document was submitted to the panel members for comments. At this stage, no additional QIs were accepted. However, minor changes pertaining to definitions were allowed.

Consensus was defined as agreement on the proposed QIs during the meeting among the attending experts.

### 11.3.3 Papers III and IV

Data collection lasted for three months. The data necessary for obtaining the quality indicators were registered in a registration form designed for this specific study. In addition to the patient's personal ID number, the registration form consisted of the 15 response-specific quality indicators with the corresponding answer alternatives. In paper IV, the 30-days mortality was registered from the patient records by a local investigator. After this, the patient's complete data set was de-identified before being shared with the lead investigator. The national investigators were responsible for data registration in the corresponding participating countries.

In paper III, an assessment of the feasibility of the quality indicators was conducted based on comments from the recording physicians. In addition to feasibility, we assessed four other important characteristics of QIs: rankability, variability, actionability, and documentation [84, 85]. This was done according to the criteria for good quality indicators defined by the Organisation for Economic Cooperation and Development (OECD) and the Agency for Healthcare Research and Quality (AHRQ).

Rankability was assessed in terms of whether a QI had a clear direction of good and bad, i.e., the QI had good rankability if high values for a QI were always better than low values.

According to criteria for quality indicators, a good quality indicator must have enough variability to allow for improvement. To assess variability, we calculated the mean and median as well as the corresponding variance for each of the QIs. This illustrates both the average performance and the variation in the participating Nordic P-EMSs.

Actionability is the possibility of influencing the QI performance. For instance, a P-EMS has a limited opportunity to reduce the time to definitive care because this depends mainly on the distances that the P-EMS unit has to work with. In that case, actionability is rather low. The actionability of all QIs was assessed.

Furthermore, for a QI to be valid, the process or structure of defining the QI must have been documented to create a better outcome. The degree of such documentation was assessed for each QI.

In paper IV, the 30-days mortality data were collected by the national investigators in their respective nations by checking the updated National Population Registers to determine whether the patients were registered alive 30 days after the P-EMS response.

In both papers III and IV, we did not report which results belonged to the specific P-EMS bases simply because the aim of this study was to assess the characteristics of the QIs and not to compare the performance of the participating services.

## 11.4 Statistical methods

### 11.4.1 Paper I

No statistical analysis was done in paper I. Data from studies included in the systematic review were not suitable for meta-analysis.

### 11.4.2 Paper II

No statistical analysis was done in paper II.

### 11.4.3 Papers III and IV

Descriptive statistics were used in papers III and IV. The QI proportions were recorded for QIs that are categorical variables, while time was recorded in minutes for QIs that were continuous time variables. All quality indicators were reported by the mean and the corresponding 95% confidence interval as well as the median with the corresponding interquartile range (IQR).

In paper III, we also used data from the 16 P-EMS bases to propose benchmarks for all QIs. We set the benchmark at the lower end of the fourth quartile for QIs in which higher values reflect better performance. For QIs in which lower values reflect better performance, we set the benchmark at the highest end of the first quartile. We depicted the benchmarking graphically, so performances within the interquartile range (IQR) are shown in yellow (average performances). Performances better than the IQR level are green (high performances) and performances poorer than the IQR level are red (low performances), as depicted in figure 12.

In paper IV, we compared the QI means for the two groups “Alive after 30 days” and “Dead after 30 days” to determine whether there was an association between QI score and mortality. The QI means were compared using a Chi-Square test for the categorical variables and a Mann-Whitney U test for the continuous variables. The reason for the latter was that the continuous variables were not normally distributed; hence, a non-parametric test was chosen. P-values for all analyses were presented with a defined significance level of  $p < 0,05$ . Further, we conducted a linear correlation analysis between the

variables “30-days mortality” and “Total Quality Score” to determine whether there was an association between mortality and Total Quality Score.

## 11.5 Ethical considerations and approvals

### 11.5.1 Paper I

This paper did not require ethical approval. Consent to participate was obtained from all panel members. We did not approach patients to participate in the expert panel.

### 11.5.2 Paper II

This paper did not require ethical approval.

### 11.5.3 Papers III and IV

Medical research on severely ill or injured patients presents certain ethical dilemmas. However, we believe that studies like this are necessary to identify “standard care” and “best practice” so that P-EMS can be improved for the benefit of our patients.

This study includes data from clinical emergencies, meaning that obtaining informed consent from patients is difficult. The exemption from informed consent is justified within the WMA Declaration of Helsinki (59<sup>th</sup> WMA General Assembly, Seoul, October 2008) Part B § 25: *“There may be situations in which consent would be impossible or impractical to obtain for such research or would pose a threat to the validity of the research. In such situations, the research can be done only after consideration and approval of a research ethics committee”*. Prospective consent for research participants prior to enrollment in this study was not possible. Further, it would be unpractical – and sometimes impossible – to consult the next-of-kin, for instance, due to their absence. Moreover, a significant percentage of P-EMS patients die before arriving at hospital. Informed consent from patients who die during the pre-hospital phase would be impossible to obtain. The inclusion of all P-EMS patients was considered important for obtaining a comprehensive and meaningful quality measurement of P-EMS responses. Although it could have been possible to obtain permission from the next-to-kin to include the deceased’s health data, we argued that the exemption from informed consent is justified within the Helsinki Declaration. The potential risks and disadvantages for the patients were similar to those of routine clinical practice, as no interventions were planned.

The study was approved by the Committees for Medical and Health Research Ethics in Sweden (reference number: 2016/109) and Finland (reference number: R16031), respectively. In Denmark, application was waived by the Committee for Medical and Health Research Ethics due to the strictly descriptive nature of the study. The Norwegian Committee for Medical and Health Research Ethics defined the study as falling outside its legislation (reference number: 2016/371). This necessitated applications to the Norwegian Data Protection Authority (reference number: 16/01113-2/SBO), the Norwegian Directorate of Health (reference number: 16/14024-3), and the Data Protection Officers at the participating Norwegian health services, all of whom approved the study. According to the approvals from all four countries, the data were obtained without informed consent from patients or their next-of-kin.

#### 11.6 Financial support

During the studies in this thesis, no study participants, co-workers, or co-authors received any financial benefits or payments. The principal investigator (Helge Haugland) was employed as a PhD candidate in the Norwegian Air Ambulance Foundation and received a salary from this organisation. The consensus meeting in study II was also financed by the Norwegian Air Ambulance Foundation.

## 12 Results

### 12.1 Paper I

In the 27 papers included in the analysis, a common understanding of which QIs to use in P-EMS did not emerge. Twenty-four of the 27 papers use QIs that can be identified as process indicators. Structure indicators and outcome indicators are used less frequently, in two and seven papers, respectively. Fifteen papers used only a single QI, whereas 12 papers applied a set of quality indicators.

Twenty-five different QIs were identified, all of which were considered suitable for international use and transferable to other P-EMSs. The most widely used QIs were “Adherence to medical protocols”, “Provision of advanced interventions”, “Response time”, and “Adverse events”.

Pertaining to the internal validity of the papers, 10 of the 27 papers did not clearly explain the methodology for developing the QIs. None of the papers provided a complete report of fixed-system variables, which complicated the process of comparing involved P-EMS concepts.

### 12.2 Paper II

In stage 1, the expert panel proposed 358 QIs. Upon the completion of stage 2, 134 indicators were left to be discussed at the consensus meeting.

During the consensus meeting, the expert panel recommended that, for purposes of clarity, the QIs from stage 2 be classified into two different categories: response-specific QIs and system-specific QIs. The former is data from the pre-hospital time interval, measuring quality on the response level, and should be feasible to collect from any P-EMS response by the P-EMS physician. The latter should be administrative data describing fixed system characteristics and should be registered once a year for services using the set of quality indicators continuously or for study purposes. The expert panel argued that the combined information from response- and system-specific QIs allows for a more thorough quality measurement than relying exclusively on response-specific QIs.

Consensus was reached on 15 response- and 11 system-specific QIs. The expert panel allowed the project group to finalise the definitions of the indicators and propose them to all 18 experts in stage 4, during which the final result was agreed upon. The QIs were allocated into one of the six quality



dimensions as defined by the IOM. All six quality dimensions were covered by the QIs, and structure, process, and outcome indicators were all represented.

### 12.3 Paper III

The dataset consisted of 5,638 requests to the participating P-EMSs. A total of 2,814 requests resulted in completed responses with patient contact. All QIs were feasible to obtain. The variability of 14 out of 15 QIs was adequate. However, because of a low rate of adverse events, we deemed it unreasonable to analyse the QI “Adverse events” using a strictly quantitative approach. Rather, we recommended that this QI should be used to identify adverse events so that they can be analysed as sentinel events. Rankability was adequate for all QIs. Actionability was assessed as being adequate for 10 QIs. However, the actionability of the QIs “Able to respond immediately when alarmed”, “Time to arrival of P-EMS”, “Time to preferred destination”, “Provision of advanced treatment”, and “Significant logistical contribution” was assessed as being poor. Documentation was adequate for 14 QIs. Benchmarks for all QIs were proposed (figure 12). An illustration of the comparison between services using the proposed benchmarks is depicted in table 1. Additionally, we presented a Total Quality Score, a variable adding up the performance of all 15 quality indicators. Both the benchmarks and the Total Quality Score are intended to aid in future quality measurement initiatives.

Quality indicator	Unit of QI	Performance assessment
Able to respond immediately when alarmed	%	<p>75 80 85 90 95 100</p>
Time to arrival of P-EMS	minutes	<p>0 10 20 30 40 50</p>
On scene time	minutes	<p>0 10 20 30 40</p>
Time to preferred destination	minutes	<p>0 20 40 60 80</p>
Patients arriving hospital alive	%	<p>80 85 90 95 100</p>
Debriefed responses	%	<p>0 20 40 60 80 100</p>
Adverse events	%	<p>0 2 4 5 6 10</p>
Complete documentation	%	<p>0 20 40 60 80 100</p>
Guidelines for actual medical problem	%	<p>0 20 40 60 80 100</p>

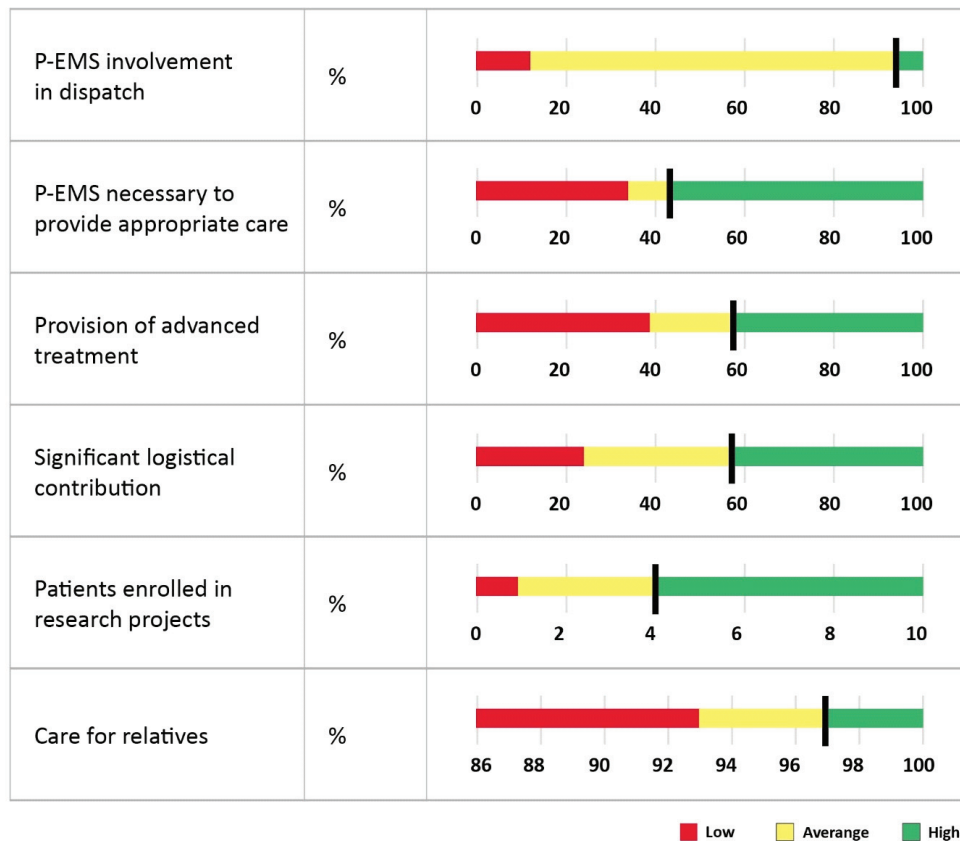


Figure 12 (Figure 2 in paper III): Benchmarking of quality indicators. Green zone: high performance. Yellow zone: average performance. Red zone: low performance. The benchmark is set at the transition between the green and yellow zones and is marked with a black and bold vertical line.

Quality indicator	Unit of QI	P-EMS base	
		Base 1	Base 2
Able to respond immediately when alarmed	%	● 96	● 83
Time to arrival of P-EMS	minutes	● 27	● 24
On scene time	minutes	● 12	● 14
Time to preferred destination	minutes	● 62	● 62
Patients arriving hospital alive	%	● 92	● 95
Debriefed responses	%	● 88	● 66
Adverse events	%	● 3	● 1
Complete documentation	%	● 27	● 34
Guidelines for actual medical problem	%	● 15	● 41
P-EMS involvement in dispatch	%	● 48	● 95
P-EMS necessary to provide appropriate care	%	● 35	● 43
Provision of advanced treatment	%	● 33	● 37
Significant logistical contribution	%	● 50	● 52
Patients enrolled in research projects	%	● 0	● 10
Care for relatives	%	● 100	● 96
Total quality score	Points (Scale:-15,15)	-1	1

Table 1 (Table 3 in paper III): Illustration of comparison between services using the proposed benchmarks. Time variables are presented as medians as they are not normally distributed. The remaining QIs are presented as means of proportions.

#### 12.4 Paper IV

In paper IV, the figures for “Survival to patient handover” and “30-days survival” were documented for all four participating nations (figure 13). Survival to patient handover was defined as survival until the patient was handed over in the hospital or as survival until handover to EMS when transported by an entity other than P-EMS. Survival to patient handover was 93.2% (Denmark), 87.3% (Finland), 93.0% (Norway), and 95.5% (Sweden). The proportion of patients surviving until 30 days after the actual P-EMS response was 83.5% (Denmark), 76.1% (Finland), 84.1% (Norway), and 89.0% (Sweden), respectively.

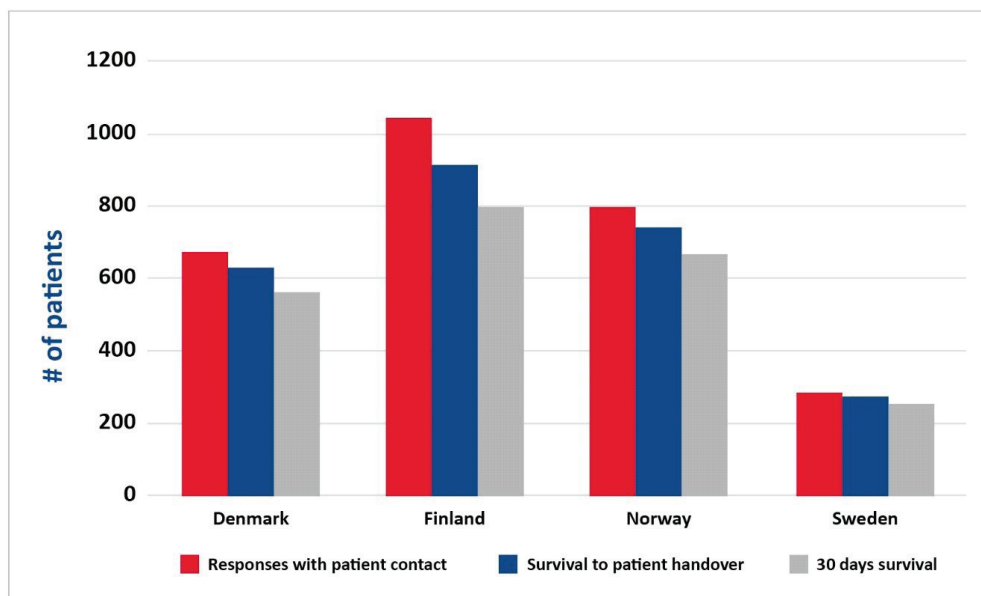


Figure 13 (Figure 3 in paper IV): Survival of patients cared for by Nordic P-EMS services.

When comparing the means of every QI for the groups “Alive after 30 days” and “Dead after 30 days”, we found a significant difference in mean values between the two groups for 8 out of 14 QIs. This indicated a correlation between QI score and 30-days mortality. However, for these eight QIs, there was no logical and consistent pattern showing that presumptive good QI performances lead to lower mortality than do poor performances and vice versa.

In study III, we presented a composite variable adding up the performance of all 15 QIs, named Total Quality Score. In study IV, we conducted a correlation analysis between 30-days mortality and Total Quality Score to determine whether a high TQS resulted in lower 30-days mortality and vice versa (figure 14). The Pearson  $r$  coefficient was 0.125, indicating no significant correlation. This is supported by the high  $p$ -value of 0.644.

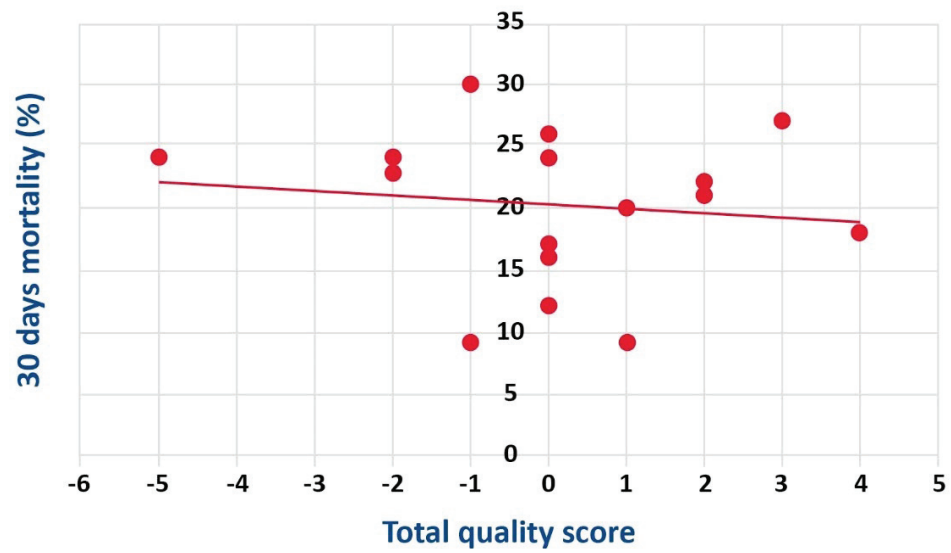


Figure 14 (Figure 4 in paper IV): The correlation between Total Quality Score and 30-days mortality.



## 13 Discussion

### 13.1 The importance of a multidimensional quality measurement

One of the main topics discussed in this thesis is that quality consists of different dimensions. When assessing the total quality of a service, one should measure the quality of different dimensions of this service and, finally, put together these assessments as a whole; this is well-described for health care by IOM [37]. This principle is, in fact, quite intuitive. An analogy from our everyday life could be the reviewing of restaurants. Imagine that the quality of two Copenhagen restaurants should be measured: Noma and McDonald's. Noma has been proclaimed the best restaurant in the world. Would we accept that McDonald's was proclaimed the better restaurant? Obviously not. But that would be the case if the restaurants were measured by only one quality indicator: namely, the time until dinner is served. We would all object to such a conclusion because it's obviously wrong; Noma is the better restaurant. This method of measuring quality makes no provision for interest in the product itself; it assesses only the time until the meal is served. However, food, service, atmosphere, value for money, etc. should be reviewed together and, based on all these assessments, it would be possible to assess the overall quality of the restaurants. This principle of reviewing multiple quality dimensions can be considered a rule to follow when measuring quality. However, unfortunately, these are often ignored when administrations and governments measure quality in pre-hospital services. In these situations, the results may be misleading or at least incomplete. Moreover, when one is measuring the overall quality of a health service, it is important to not only review multiple quality dimensions but also include a comprehensive selection of patients with different diagnoses as demonstrated by McGlynn et al. [86]. This issue is also addressed by Turner et al., who recently developed a set of QIs for EMS [87]. These QIs should reflect the preferences of both services and users. Furthermore, they should secure a relevant quality measurement for a wider range of conditions than those time variables are designed for, i.e., cardiac arrest, MI, stroke, etc.

It is unlikely that anyone would even consider evaluating restaurants in the manner described above, and we should not measure the quality of P-EMS this way either. For instance, by exclusively using time variables to measure P-EMS quality, we measure the logistical aspect of these services, not the level of care. This gives an impression of a lack of interest in measuring the quality of the care as such, or at least of assuming that all pre-hospital care is identical and of a high standard. The message is not to raise



doubts about the importance of time variables but simply to state that we must measure multiple quality dimensions to get an impression of the total quality of a system. Moreover, if P-EMS quality measurement focuses on only one quality dimension, the result may be an undesired attention shift, resulting in decreased quality for quality dimensions not measured. This undesired attention shift is not specific for some quality indicators but, rather, is a universal challenge in quality measurement when the used quality indicators are few.

### 13.2 Precision in dispatch is necessary to improve quality

#### 13.2.1 Economy and equity of care

Precision in dispatch is a prerequisite for obtaining a good value of P-EMS. If P-EM is unnecessarily dispatched to many patients who could be cared for by others, the value equation in figure 4 would be unfavourable because the obtained quality per cost unit spent would be low. The economy of P-EMS is an issue of interest when one is assessing the value of this service. The basic cost of helicopter-based P-EMS is the dominating extra cost, whereas costs per response are minor. In a study from 2003 comparing the costs of a German physician-staffed EMS with those of a British paramedic-staffed EMS, the additional cost of an experienced doctor on board was equal to € 4.49 per inhabitant per year [88]. Ringburg et al. found a high willingness to pay for lives saved by HEMS in a Dutch population [89]. However, more important than the economic costs related to HEMS are the human costs, i.e., the number of casualties from HEMS accidents.

In addition, precision in dispatch is important for securing equity of care, which is one of the six quality dimensions of the IOM. If a P-EMS unit is occupied by a patient who could be equitably cared for by EMS when the EMCC receives an alarm about a patient in more need of P-EMS competence, the result may be an unavailable or delayed P-EMS unit. This should, in itself, stimulate an effort to obtain the highest possible precision in dispatch.

#### 13.2.2 The dilemma of overtriage vs undertriage

Due to increased regionalisation, as well as increases in the population, the number of annual emergency calls, and the number of annual emergency ambulance missions throughout the past years, one might expect that the number of P-EMS responses would increase as well. Even “new” patient

groups may substantiate this expectation, e.g., an increased number of stroke patients in P-EMS responses. However, Østerås et al. demonstrated that the annual number of dispatches to primary P-EMS responses did not change from 2004 to 2013 [90]. This stability in dispatch numbers might indicate that the number of patients in need of P-EMS is quite stable. In 2013, Krüger et al. found that the pre-hospital incidence of severe illness or injury was estimated to be 25-30 per 10,000 person-years [12].

The optimal use of P-EMS would be that all patients in need of P-EMS competence are cared for by P-EMS (no P-EMS undertriage), whilst all patients not in need of P-EMS competence are cared for by others – for instance, traditional EMS (no P-EMS overtriage). The undertriage rate is often defined to capture the proportion of patients receiving suboptimal care, thus increasing the risk of mortality or other adverse events. The overtriage rate is defined to capture the proportion of unnecessary use of a resource – for instance, the P-EMS resource. The American College of Surgeons Committee on Trauma (ACS-COT) considers an undertriage rate of <5% and an overtriage rate of <35% as acceptable for the initial in-hospital care of trauma patients [91]. Trying to maximise specificity (undertriage) and sensitivity (overtriage) is a challenge in any health care setting. It might be even more challenging to achieve in acute care, where lack of information often dominates the initial presentation of a potential critically ill patient. Moreover, the consequences of undertriage can be fatal for some patients. However, the importance of selecting the appropriate patients is widely acknowledged and undertriage and overtriage rates are important quality indicators for trauma systems [92]. It seems reasonable that undertriage and overtriage rates could also be useful quality indicators in P-EMS.

Figure 15 describes the relationship between sensitivity and specificity for a diagnostic test. Setting the cut-off for disease at position B gives the best balance between false positives and false negatives. Moving the cut-off to the left increases the sensitivity but lowers the specificity. This leads to more false positives. On the other hand, moving the cut-off to the right lowers the sensitivity but increases the specificity, resulting in more false negatives. This figure illustrates that sensitivity and specificity are inextricably linked. If we think of the P-EMS dispatch as a diagnostic test, we aim to reach the perfect balance between specificity (undertriage) and sensitivity (overtriage). This will lead to the best dispatch accuracy and will be the most appropriate use of this limited resource. If the focus is exclusively on limiting undertriage, the result will probably be a massive overtriage and a lot of “false positives”, i.e., the P-EMS will respond to many patients who could be cared for at a lower level of care. In terms of the quality dimension of efficiency, such an overtriage is not associated with good service quality. Therefore,

one should aim to balance the P-EMS dispatch to create a practical compromise between specificity and sensitivity.

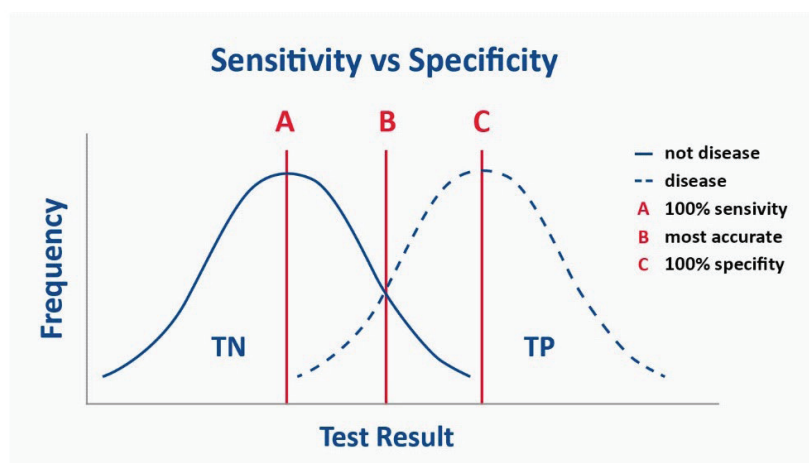


Figure 15: The relationship between sensitivity and specificity.

In paper II, we do not find undertriage and overtriage rates among the set of quality indicators developed for P-EMS. It is possible to argue that this is a limitation of the set of QIs. However, some of the quality indicators are indirect measures of the precision of triage. That is the case with the quality indicators “P-EMS necessary to provide appropriate care”, “Provision of advanced treatment”, and “Significant logistical contribution”.

### 13.2.3 Provision of advanced treatment

For the QI “Provision of advanced treatment”, the P-EMS physicians were asked if the P-EMS service provided advanced treatment in the actual response. A number of options were defined as advanced treatment: Procedures (both medical and rescue techniques) or medications offered only by P-EMS units in the actual region were defined as advanced treatment. Also, procedures or medications offered by local pre-hospital units other than the P-EMS unit were defined as advanced treatment when other local units could not be present on-scene. Avoidance of unethical or unnecessary treatment and presence in

particularly demanding situations such as a major incident were also included in the term “advanced treatment”. Thus, the term was relatively widely defined. The P-EMS bases reported that they provided advanced treatment in 49% of the responses (mean). The lowest mean value among the bases was 33%, while the highest was 77%. The large range might indicate a different use of the participating P-EMS units. The fact that almost half of the responses were completed without the provision of widely defined advanced treatment by P-EMS might indicate room for improvement to reduce a possible overtriage.

#### 13.2.4 Significant logistical contribution

Even if no advanced treatment is provided, there could still be good reasons for using P-EMS due to the logistical contribution of this resource. For time-critical incidents like ST-elevation myocardial infarction and stroke, we know that time to definitive treatment is decisive for the patient’s prognosis. If transportation by helicopter can save time for such patients, the use of P-EMS may be reasonable even though these patients rarely need the added competence of a specialist physician. The QI “Significant logistical contribution” was developed to cover this advantage of helicopter-based P-EMS: time saved due to transport by air. For this QI, the P-EMS physicians were asked if the logistical contribution by P-EMS resulted in significantly better service than the existing alternative. According to the QI definition catalogue, the answer to this question should be “yes” if the logistical contribution by P-EMS reduced the estimated time to hospital by  $\geq 15$  minutes for time-critical conditions like STEMI, stroke, and severe trauma. Furthermore, the answer should also be “yes” if the use of P-EMS was necessary to access or evacuate the patient from an area otherwise difficult to reach. Those would be mainly areas on land and sea not accessible by regular EMS. The P-EMS bases reported that the logistical contribution by P-EMS resulted in better service for patients in 43% of responses (mean). The lowest mean among the bases was 6%, while the highest mean was 80%. This might reflect a substantial difference in the use of P-EMS at the different bases, and/or a substantial difference in geography. When we look into the figures of this QI, we find that the five participating Finnish P-EMS are the five bases with the five lowest means, indicating that the Finns do not use their P-EMS resources a lot for these logistical purposes. An explanation for this finding is found in the number of patients encountered by P-EMS who are eventually transported to hospital by other resources (typically, regular EMS). We observed that the five bases whose patients were most frequently transported by other resources were all Finnish bases. The mean proportion of patients transported by others was 41% among the Finnish bases, compared to a mean of

11% among the other bases. This substantial difference reflects a conceptual variance in the way in which the P-EMS resource is used. If a difference in use was related to geography, we would expect that the most urban bases, with presumably the shortest distance to hospital, had the highest proportion of patients transported by other resources. This is not the case and reinforces the impression that this difference in use is a conceptual difference between Finland and the other nations. It remains unclear whether this is due to demography, tradition, the hospital infrastructure, or other factors.

#### 13.2.5 P-EMS necessary to provide appropriate care

For this QI, the P-EMS physicians in study III were asked whether the assistance of the P-EMS unit was necessary to give appropriate care to the patient. It was specified that decision-making by P-EMS (without any therapeutic procedures) was included in the requested competence. The participating P-EMS services determined that the assistance of the P-EMS unit was necessary or probably necessary in 39% (mean) of the responses. The lowest mean value among the P-EMS bases was 27%, while the highest was 52%. As illustrated by the mean of this QI, the P-EMS physicians assessed that 61% of the responses could be handled without P-EMS assistance, e.g., by ordinary EMS, and the patient would still be provided with appropriate care. This figure can be assessed as the overtriage rate as judged by the P-EMS physicians. Giannakopoulos et al. documented a P-EMS overtriage rate of almost 44% in a Dutch region [92]. They assessed this as high and concluded that it should be possible to reduce this rate and still keep the undertriage rate within acceptable limits. The undertriage rate in the study was 4%. If we apply the ACS-COT definition of an acceptable overtriage rate for trauma patients (<35%) to our total population, knowing that these populations are different, there might still be an indication of improvement possibilities for overtriage rates in Nordic P-EMS. As with all QIs in this study, these numbers are self-reported by the P-EMS physicians. In such self-reporting studies, cognitive bias is a threat. This is an overestimation of one's own quality, ability, and importance. In the presence of such a bias, the real necessity of P-EMS to provide appropriate care might be expected to be even lower than 39%.

### 13.2.6 Responses without patient contact

In study 3, a mean of 49.8% of the P-EMS responses involved patient contact. Reasons for responses without patient contact included cancellation due to no medical indication, poor weather, concurrency conflicts, or technical problems with helicopter or other equipment. Østerås et al. found that 62% of all dispatches to three P-EMS bases in Western Norway were completed without cancellation [90]. It is plausible to assume that the vast majority of these responses included patient contact. The remaining 38% of the responses were cancelled. Over half of these responses were cancelled due to “No indication” as evaluated by the P-EMS physician. The cancellation rate in the Norwegian study was comparable to those of a Dutch study and a Canadian study, which reported cancellation rates of 44% and 32%, respectively [92, 93].

As illustrated in figure 16, there is a substantial difference between the P-EMS bases in terms of the proportion of responses with patient contact. The proportion of responses with patient contact varied from 24.9% to 92.0%. When the same proportions were calculated only for primary responses, i.e., excluding interhospital transfers and search-and-rescue missions, the figures varied between 24.6% and 85.5%. In our study, the Finnish P-EMS bases had the lowest proportion of responses with patient contact. If a P-EMS unit systematically aborts responses when it receives information indicating that it is not needed, the proportion of responses with patient contact will obviously decrease. Whether this can explain the Finnish data remains unknown but a practice like this would, in any case, have pros and cons. Aborting responses due to the availability of more detailed information en route is beneficial to preserving the availability of the P-EMS unit – a quality in itself. On the other hand, a high rate of aborted responses may indicate that the P-EMS unit is often dispatched unnecessarily. This may be a quality problem in terms of the availability of the P-EMS unit, especially with respect to adding duty time for the crew. Moreover, a too-liberal dispatch practice can be criticised from an economic point of view if one considers these responses as a waste of resources. However, the majority of the costs of these services are basic costs, not extra costs per response. Nevertheless, if a P-EMS unit regularly has a low proportion of responses with patient contact, this might indicate that the P-EMS unit is dispatched on a too-weak information base. Spending some extra time at the EMCC or at the P-EMS base to collect more information might be beneficial. Even waiting for an en-route ambulance to arrive on the scene to provide first-hand information might be a good investment of time in selected cases – and a contribution to securing the most precise dispatch of the P-EMS unit.

In addition to having the lowest proportion of responses with patient contact in study III, the Finnish bases had the highest mortality. A possible explanation for this may be that Finnish P-EMSs manage to select the responses with the most critically ill patients. As discussed earlier in this thesis the Finnish bases also have, by far, the highest proportion of patients encountered by the P-EMS unit who are eventually transported by others. The combination of high mortality, a high proportion of patients transported by others, and a high proportion of responses aborted en route might indicate that the Finnish service has a low threshold for dispatch but a higher threshold for completing responses unless the patient is critically ill.

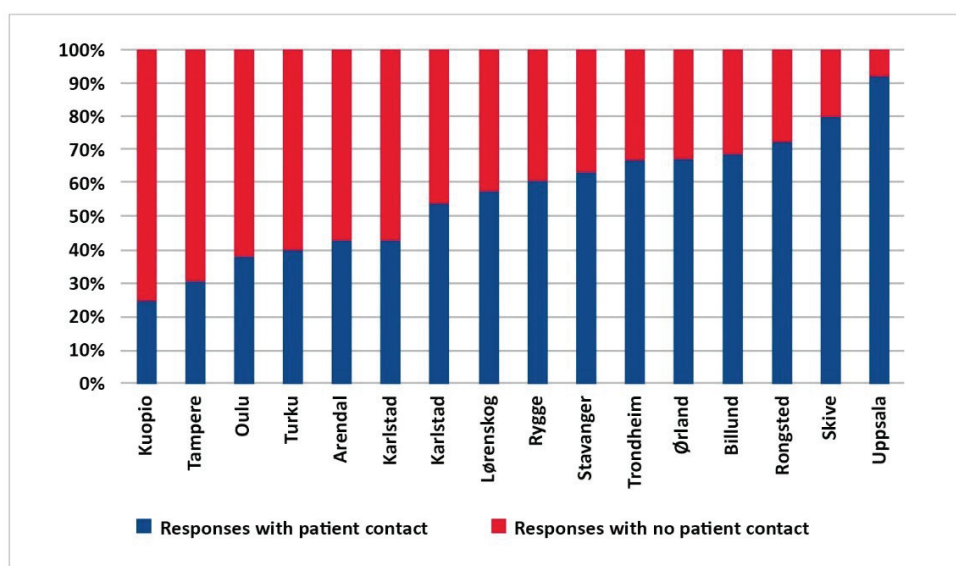


Figure 16 (Figure 1 in paper IV): Proportion of P-EMS responses with and without patient contact.

### 13.3 Response time (Time to arrival of P-EMS).

In EMS, the most commonly used structure indicator is response time [59]. For a P-EMS unit, we understand response time as the time from the alarm call to the time when the P-EMS unit arrives at the patient, as measured in the QI "Time to arrival of P-EMS". We defined the end of this interval as the time when the P-EMS unit arrives at the patient, not at the scene. In most responses, this difference is

very small. However, in selected responses, the length of time between landing and patient encounter might be longer – for instance, if one must walk the last part along steep terrain. In paper III, we documented a mean response time of 27 minutes for the participating P-EMS bases. The lowest mean response time among the P-EMS bases was 18 minutes, while the highest was 36 minutes. The figures apply to primary responses only. (Secondary responses and SAR responses were excluded.)

In Germany, the response time for virtually the whole country should be less than 15 minutes. A fully equipped medical team, including a physician if needed, will arrive within this response time [18]. Here, the density of ground-based and helicopter-based P-EMS resources is high. The coverage of the Norwegian air ambulance service is also considered comprehensive. However, the Norwegian authorities have defined that 90% of the population should be reached by helicopter-based P-EMS within 45 minutes [30]. Thus, the response time requirements indicate a different reality pertaining to the geography, demography, and tradition of pre-hospital services despite the existence of well-functioning health care in both Germany and Norway. Nevertheless, in both countries, response times are considered important political goals.

Response time is related primarily to the distance from the position of the P-EMS unit (often, the P-EMS base) to the patient [94]. Thus, it is hard to imagine what could reduce the response time significantly, except for a higher P-EMS base density. Because of the nature of this QI, it is hard to influence the performance of this QI for a P-EMS crew. However, a poor actionability for the P-EMS crew is no surprise, as this is a structure indicator. Structure indicators describe the setting in which care is provided, such as equipment, staffing, facilities, and deployment [95]. Structure indicators may reflect standards defined on an administrative level, often regionally or nationally. Thus, the actionability for the P-EMS crew is often low for structure indicators; changes in performance must be initiated by administrators or politicians. Despite poor actionability for this QI, the expert panel that developed the QIs argued that this variable is important because timely specialist competence in the pre-hospital phase of care contributes to the equity of access to health care for those living in remote areas. Thus, response time is important for establishing the population's sense of security pertaining to what service to expect when serious illness or trauma occurs. This is supported by a Nordic report on data collection and benchmarking in EMS [96].



## 13.4 Mortality in P-EMS

### 13.4.1 Strength and limitation of mortality as a QI

As noted earlier in this thesis, a widely-cited definition of quality is “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” [2]. From this definition, it seems clear that outcome is not necessarily the same as quality. Though high-quality care might increase the likelihood of a desired outcome, the outcome might still be poor. For instance, a severely injured patient might die despite receiving excellent care.

A major strength of mortality as a quality indicator is the fact that it is a hard outcome. For patients, it is hard to imagine a more important outcome when one is suffering from a serious illness or injury.

Mortality measures are easy to define and have traditionally been important in reducing preventable deaths in health care. However, as preventable deaths have decreased due to improved care, the use of mortality measures alone no longer seems adequate [97]. A falling baseline mortality complicates the evaluation of the effect of interventions due to the need for a larger sample. A decreasing mortality may indicate that death does not occur frequently enough in selected patient groups to secure the necessary event frequency to be a meaningful QI [98]. Moreover, as described by Petros et al., the use of mortality as a QI is complicated by patient heterogeneity and heterogeneous causes of death [99]. From a physiological point of view, mortality is a composite outcome. For example, patients may die from a thoracic trauma for different reasons: cardiac tamponade, tension pneumothorax, haemorrhagic shock, lethal arrhythmias, oxygenation failure due to pulmonary contusion, and more. A quality improvement project aimed at improving the identification and treatment of tension pneumothorax will therefore, at its best, be able to prevent only a fraction of the deaths. This, again, can make it difficult for a quality improvement program to have any measurable impact on all-cause mortality.

Furthermore, most patients are unlikely to see any change in mortality. This can typically be the case for patients with high co-morbidity and/or a high severity of actual illness or trauma. For these patients, mortality rates can be very high regardless of the quality of care provided. At the other end of the scale, we find otherwise-healthy patients with low severity of their actual illness. These patients will probably survive unless they are subjected to severe adverse events. In the middle of these two groups, we find the patients with an intermediate risk of death. These are the patients for whom the quality of care can most likely make a significant difference. For the other two groups, it will probably be hard to find lower

mortality due to increased quality of care. If we consider the number of undocumented interventions that we perform every day in pre-hospital care, it seems naive to believe that we should not change our practice if there is no proven benefit with respect to mortality. The need for endpoints other than mortality in the evaluation of interventions in critically ill patients is supported by studies from intensive care medicine [100, 101].

The challenge in using mortality as a quality indicator can be at least partly solved by choosing quality indicators that measure outcomes more proximal to the provided care. For instance, improvement in physiological variables from P-EMS arrival at the patient until admittance in the ED is more closely related to the provided care. QIs more proximal to the provided care are more likely to provide a useful result, first of all, because it is more common to be admitted to the ED than to die. This enables the analysis of a larger signal from the same number of patients. Moreover, a QI measuring change in physiology is a continuous variable, as opposed to mortality, which is a binary variable. This provides more details about the outcome over time.

Advanced pre-hospital care is constantly evolving, and so should our goals for the patients. Most P-EMS physicians would probably agree that our ambitions on behalf of the patient are about more than survival. For instance, we want our traumatic brain injury patients not only to survive but to survive with as few sequela as possible. In the future quality measurement of P-EMS, a stronger focus on soft quality indicators may be adequate to obtain a clear and comprehensive assessment of the quality of our care.

#### 13.4.2 Patterns of survival in Nordic P-EMS

Østerås et al. found that 88% of the patients in completed HEMS responses survived until hospital discharge [94]. In paper IV of this thesis, we found that the proportion of patients surviving until 30 days after the actual P-EMS response was 83.5% (Denmark), 76.1% (Finland), 84.1% (Norway), and 89.0% (Sweden), respectively. In both studies, all kinds of responses (primary, secondary, and others) were included. However, the endpoint of the mortality variable was different: time of discharge and 30 days after the response, respectively. In a study of a P-EMS unit in Odense, Denmark, Mikkelsen et al. found that 83.8% of the patients survived the first 30 days after the P-EMS contact [102]. This is in concordance with our results in study IV. The proportion of patients surviving until patient handover in paper IV was 93.2% (Denmark), 87.3% (Finland), 93.0% (Norway), and 95.5% (Sweden). This, again, means that the proportion of patients surviving until patient handover but not until 30 days after the P-

EMS response was 10.5% (Denmark), 12.8% (Finland), 9.6% (Norway), and 6.8% (Sweden). In the study by Mikkelsen et al., this proportion was 5.7% [102]. We consider these patients to be an especially interesting target group for quality improvement initiatives. The rationale for this idea is that most of the patients who die before hospital admittance probably suffer an injury or disease that is extremely difficult to survive. Therefore, our chances of improving survival in this group might be limited. However, the patients surviving until hospital admission proved to have at least some signs of sustained physiology. For some of the patients in this group who still die within 30 days, we believe that there is an interesting opportunity to improve their quality of care – and, thus, hopefully prolong their survival. This should not be interpreted as indicating that we believe all these patients can be saved but, rather, as an indication that these patients may be the most interesting target group for quality improvement in pre-hospital critical care.

Three Danish studies on mortality among patients cared for in EDs in Denmark allow us to compare the mortality of the Nordic P-EMS population to that of other patient populations of interest. Two of the studies demonstrated overall 30-days mortality of 3.0% and 4.7%, respectively, for the whole group of patients admitted to the Eds [103, 104]. It should be mentioned that patient contacts in these studies were included regardless of whether or not they resulted in hospital admission. Nevertheless, the mortality for these ED patients was considerably lower than that for the P-EMS patients in our paper IV. In the third study, the 30-days mortality for patients transported to hospital by ambulance after an emergency call was found to be 4.4% [105].

### 13.5 Limitations

Paper I: As always in a systematic review, a limitation of this paper is that screening, eligibility checking, and qualitative synthesis of literature is a product of the review authors' judgements, thereby allowing for subjective interpretations of the studies' content. Although screening and eligibility assessments were conducted in pairs, data extraction and quality appraisal were conducted primarily by one author. Moreover, no established tools for data extraction or quality appraisal were available. Second, some of the included studies are not explicitly presented as quality measurement studies. However, when screening the literature, we recognised that some studies were quality measurement studies despite the fact that they did not use the quality measurement terminology. A possible reason for this might be the fact that quality measurement terminology is still quite new in P-EMS and is not widely used.

Accordingly, when deciding whether a study should be included, we focused on the actual content of the screened literature and not on the presence or absence of correct quality measurement terminology. Doing this, we determined whether a paper “concerns an initiative to improve healthcare; broadly defined to include the quality, safety, effectiveness, patient-centeredness, timeliness, cost, efficiency, and equity of healthcare”, as stated by the SQUIRE guidelines [106]. We chose this approach to avoid overlooking potentially relevant aspects of quality measurement studies in P-EMS.

Paper II: In this paper, we chose a modified nominal group technique to establish QIs for P-EMS. Different methods for developing QIs exist; often, they are expensive and time-consuming [5, 107, 108]. We considered the use of an expert panel as an appropriate approach toward developing the QIs. However, other methods could have been used as long as they secured a systematic selection of QIs. For instance, the use of a systematic literature review, either alone or in combination with a consensus process, could have been possible. Developing QIs from existing clinical guidelines could have been another option – a method that normally saves time and money [109]. The use of existing governmental reports on quality measurement could have been a possibility in the choice of QIs for P-EMS. However, to the best of our knowledge, no comprehensive report on QIs in P-EMS existed at the starting point of this project. (In 2018, a Nordic project group presented a report with a selection of QIs for EMS [96]). Moreover, reports and recommendations are usually specific to health care settings. Application of QIs from other parts of the world or from other related health care services, e.g., ordinary EMS, cannot always be done without an adaption process [110]. Pertaining to the composition of the expert panel, more representatives from countries outside Scandinavia could have added valuable views and opinions with respect to the selection of quality indicators. France and Germany are both examples of countries with a long tradition of well-developed P-EMS; the recruitment of experts from these (and other) countries could have improved the scientific value of the study and increased the geographical variety. Moreover, representatives from EMCCs would have been a relevant supplement to the competence in the expert panel. EMCCs are often the gatekeepers for the P-EMS units and their views on which quality indicators are most relevant for P-EMS would likely have added value to the consensus process.

Paper III: One of the limitations of the current analysis is that the attending physicians registered all the data. They are, therefore, subject to registration bias and recall bias. Except for the feasibility of the QIs, the different QIs’ characteristics were assessed by the authors. The variability was assessed based on the data (mean and median). However, to the best of our knowledge, thresholds for defining poor, fair, and

good variability for QIs do not exist. Therefore, conclusions on this topic were the result of assessments and consensus among all authors. With that in mind, the assessment of variance is somewhat arbitrary.

Paper IV: Mortality is very much determined by the patient's status. Nonetheless, we have combined all patients in all countries and all services into one group to explore the association between QIs and mortality. This was done to secure a normal clinical setting in P-EMS, as these QIs are developed for everyday quality measurement in international P-EMS regardless of patient characteristics. Hence, this seems to be an adequate setting for our study. However, it might be that subgroup analysis of specific patient groups – for instance, high mortality diagnosis – would reveal different correlations between QIs and mortality. Regarding missing data, "Survival to handover" data were missing for only 6 out of 2,814 patients. However, 30-days mortality data were missing for 9.7% of the patients. These were either patients with foreign personal identification number or patients with unknown identities. Both patient groups are regularly taken care of by P-EMS. The problem of losing patients to follow-up because of unknown identity in the pre-hospital phase has also been reported by Christensen et al., who reported a loss to follow-up of 17.8% [104]. In all four nations, when collecting 30-days mortality data, we experienced the same difficulties pertaining to these patient groups. The data collection period was partly in the summer months, when the number of foreign tourists in Nordic countries is high. Thus, the proportion of missing data might at least partly be explained by a relatively high number of foreign citizens treated by P-EMS. We have no reason to believe that the mortality of the mentioned patient groups differs significantly from that of the rest of the patient cohort. Thus, we assess the 30-days mortality figures as representative for the patients in the study group, although the missing data for these figures ideally should be lower to secure the most valid results.

## 14 Future perspectives

Feasible and reliable quality measurement depends largely on robust documentation systems to ensure proper data quality and prevent registration fatigue among clinicians. To enable continuous quality measurement, the data necessary to feed the quality indicators should be integrated into these documentation systems. Ideally, as many variables as possible should be automatically collected for patient charts and documentation systems through electronic data capture. This seems imperative for continuous quality measurement and is an interesting challenge for the future. To ensure that quality measurement data become an integrated part of P-EMS, the presentation of figures should be intuitive. Graphical presentations, possibly in dashboard solutions, may aid in interpretation [111].

In everyday clinical work, analysis of the quality measurement data through statistical process control (SPC) may be an attractive method of displaying data with statistical thoroughness. An advantage of the dynamic SPC is the possibility of discriminating between common cause variation (inevitably, a part of daily activity) and special cause variation (indicative of something special happening) – a differentiation that may guide quality improvement initiatives [112]. Understanding variation over time – as made possible by SPC – is, in fact, the key to understanding quality improvement. Aggregated data presented in tabular formats or with summary statistics can be useful in many settings. However, they will not be the correct tools for measuring the effect of quality measures. A central principle in quality measurement is that the data should lead to improvement, not judgement. It seems more likely to achieve this by using dynamic data in SPC than by using aggregated data.

Another future challenge in P-EMS is that of standardisation. As documented in paper III, there is considerable variation in Nordic P-EMS pertaining to the dispatch practice, the provision of advanced treatment, documentation, debriefing, and more. It seems plausible that more standardisation could improve these results. Standardisation is the fundamental first step in quality improvement [113]. First, when the owners of a process agree that this process is done in a consistent, standardised manner, then the timing is correct for proceeding to the identification of measures intended to improve quality. If not, the first task is to standardise the process. Standardised processes exhibit so-called common cause variation. Common cause variation is built into a process and the factors can be known or unknown. However, the final impact that this variation has on your output is normally predictable and controllable. Special cause variation, on the other hand, occurs when something out of the ordinary happens in a

process – for instance, due to a failure of medical equipment. One should attempt to improve only processes with the signature of common cause variation. This is because such processes are stable and predictable – and, thus, good targets for quality improvement. Because of its central role in creating such stable processes, standardisation can be considered a future challenge for quality improvement in P-EMS. An important remark pertaining to the importance of stable processes in quality improvement is that stable performances without much variation do not necessarily represent a well-performing system. The whole system may be uniformly under-performing, but because of the stability of the process, it represents an opportunity for goal-directed quality improvement. The importance of standardisation and reduced variation has been emphasised by one of the leading management thinkers in the field of quality, Dr. W. Edwards Deming. He said, *“If I had to reduce my message for management to just a few words, I’d say it all had to do with reducing variation”*. Standardisation is imperative to reaching this goal.

The need for quality measurement and quality improvement in P-EMS seems clear. With or without optimal documentation systems, building a culture for quality improvement is imperative, involving leadership involvement and long-term commitment. This is a management challenge for the future. The need for an open atmosphere in every service that wants to improve its quality cannot be emphasised enough. A prerequisite for quality improvement is a workplace where improvement potential, mistakes, and omissions can be discussed openly as a matter of course. Everyone can improve but it is decisive that the workplace atmosphere supports – and does not complicate – quality improvement as described above. Creating such a workplace atmosphere is primarily a leader responsibility [114]. *“The leader should be a coach, not a judge”*, as stated by Dr. W. Edwards Deming.

To improve, we must compare ourselves to other services. This is the value of benchmarking [115]. By identifying best practices, we can learn from the best – from each other. In other words, the benchmarking is not for the judgement of services but for the improvement of patient care. Donald Berwick, the former president of the Institute of Healthcare Improvement, explained the purpose of benchmarking in a simple way: *“Benchmarking is like turning the light on! Without benchmarking and transparency we are in the dark”*. However, leaders must cautiously and wisely use benchmarking. At its worst, benchmarking can nourish rivalry and leave people bitter, discouraged, and feeling inferior due to a poor result. Sometimes it can be hard to understand the reasons why your service is performing poorer than other services. Thus, it is difficult to plan how to improve your performance. In fact, having

a target to reach, but lacking a plan for reaching that target, can have the opposite effect than desired. People may get discouraged and stop trying. Or they may start fabricating or manipulating their own results – especially if they are not able to reach the target. Moreover, performance differences may be totally attributable to the system, rather than the workers themselves. Workers in a poor-performing group might therefore believe, often rightly, that they are labelled unfair. Nevertheless, when used with knowledge and caution, benchmarking has substantial power. Thus, it seems reasonable that more benchmarking in P-EMS should be another task for the future.

As emphasised in this thesis, high dispatch precision is a prerequisite for increasing the quality of P-EMS. Thus, the competence (the P-EMS crew) and the transportation tool (the helicopter) can be used where they are needed the most. Therefore, future studies aimed at increasing dispatch precision remain pivotal. Studying the effect on dispatch precision by introducing higher operational competence in the EMCC may be one option. Another option could be to study the effect on dispatch precision of P-EMS when telemedicine is introduced in regular EMS. Is it possible to increase dispatch precision when video of the actual patient is available to the P-EMS physician? Finally, examining the effect of more defined protocols for P-EMS dispatch is a possible future research project. Even introducing more automatised dispatch by computer algorithms to study its effect on dispatch precision might be a possible study for the future.





## 15 Conclusions

### Paper I

The review demonstrated a lack of shared understanding of which QIs to use in P-EMS. Moreover, papers using only one QI dominated the literature, thereby increasing the risk of a narrow perspective in quality measurement. Future quality measurement in P-EMS should rely on a set of consensus-based QIs, ensuring a comprehensive approach to quality measurement.

### Paper II

Using a modified nominal group technique, an international expert panel successfully developed a set of QIs for international P-EMS. The QIs should be prospectively tested for feasibility, validity, and reliability in clinical datasets. The quality indicators should then allow for adjusted quality measurement across different P-EMS systems.

### Paper III

In this paper, the 15 response-specific QIs developed for P-EMS (in paper II) were tested for necessary QI characteristics. The feasibility of obtaining the necessary data for these QIs was good. The variability of the QIs was adequate for all QIs except for the QI “Adverse events”. For five QIs, actionability was assessed as poor. Three of these QIs measured the timeliness of P-EMS. Some QIs depended on characteristics of the P-EMS services that might differ, such as patient volume, distances, and patient characteristics. Thus, they should be interpreted with caution for service comparison. However, it seems more straightforward to use these QIs for internal quality measurement of a service. To aid in future quality measurements in P-EMS, benchmarks for all QIs have been proposed. Moreover, we presented a variable combining the QI performances into a single score: the Total Quality Score.

#### Paper IV

This study demonstrated that good overall process quality did not seem to correlate with good outcome quality. However, these are complementary measures, both with undeniable value in terms of identifying the total quality achieved in P-EMS. Thus, combining mortality measures and multiple process quality indicators seems adequate for future quality measurement.

## 16 References

1. Porter ME. What is value in health care? *N Engl J Med*. 2010;363:2477-81.
2. Institute of Medicine. *Emergency Medical Services at a Crossroads*. The National Academies Press, Washington DC, 2006.
3. World Health Organization. *Quality of care: a process for making strategic choices in health systems* 2006.
4. National Committee for Quality Assurance. *The essential guide to health care quality*., [https://www.ncqa.org/Portals/0/Publications/Resource%20Library/NCQA\\_Primer\\_web.pdf](https://www.ncqa.org/Portals/0/Publications/Resource%20Library/NCQA_Primer_web.pdf). Accessed 01.03. 2016.
5. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care*. 2003;15:523-30.
6. Fevang E, Lockey D, Thompson J, Lossius HM, Torpo Research C. The top five research priorities in physician-provided pre-hospital critical care: a consensus report from a European research collaboration. *Scand J Trauma Resusc Emerg Med*. 2011;19:57. doi:10.1186/1757-7241-19-57.
7. Snooks H, Evans A, Wells B, Peconi J, Thomas M, Woollard M et al. What are the highest priorities for research in emergency prehospital care? *Emerg Med J*. 2009;26:549-50. doi:10.1136/emj.2008.065862.
8. J.E. Nilsen KS, I.B.M. Tjelmeland, K.S. Halvorsen, J. Kramer-Johansen, T.M. Olasveengen. Kvalitetsindikatorer i den akuttmedisinske kjeden. [Norwegian]. [www.nakos.no](http://www.nakos.no): Nasjonal kompetansetjeneste for prehospital akuttmedisin 2015.
9. Price L. Treating the clock and not the patient: ambulance response times and risk. *Quality & safety in health care*. 2006;15:127-30. doi:10.1136/qshc.2005.015651.
10. Rehn M, Kruger AJ. Quality improvement in pre-hospital critical care: increased value through research and publication. *Scand J Trauma Resusc Emerg Med*. 2014;22:34.
11. Berwick DM. Continuous improvement as an ideal in health care. *The New England journal of medicine*. 1989;320:53-6.
12. Kruger AJ, Lossius HM, Mikkelsen S, Kurola J, Castren M, Skogvoll E. Pre-hospital critical care by anaesthesiologist-staffed pre-hospital services in Scandinavia: a prospective population-based study. *Acta Anaesthesiol Scand*. 2013;57:1175-85. doi:10.1111/aas.12181.
13. Kruger AJ, Skogvoll E, Castren M, Kurola J, Lossius HM, ScanDoc Phase 1a Study G. Scandinavian pre-hospital physician-manned Emergency Medical Services--same concept across borders? *Resuscitation*. 2010;81:427-33. doi:10.1016/j.resuscitation.2009.12.019.
14. Fischer M, Kamp J, Garcia-Castrillo Riesgo L, Robertson-Steel I, Overton J, Ziemann A et al. Comparing emergency medical service systems--a project of the European Emergency Data (EED) Project. *Resuscitation*. 2011;82:285-93.
15. Ghosh R, Pepe P. The critical care cascade: a systems approach. *Curr Opin Crit Care*. 2009;15:279-83. doi:10.1097/MCC.0b013e32832faef2.
16. Trunkey DD. Society of University Surgeons. Presidential address: On the nature of things that go bang in the night. *Surgery*. 1982;92:123-32.
17. World Health Organization DolaVP. *World report on road traffic injury prevention* 2008.
18. Roessler M, Zuzan O. EMS systems in Germany. *Resuscitation*. 2006;68:45-9. doi:10.1016/j.resuscitation.2005.08.004.
19. MacFarlane C, van Loggerenberg C, Kloock W. International EMS systems in South Africa--past, present, and future. *Resuscitation*. 2005;64:145-8. doi:10.1016/j.resuscitation.2004.11.003.
20. Langhelle A, Lossius HM, Silfvast T, Bjornsson HM, Lippert FK, Ersson A et al. International EMS Systems: the Nordic countries. *Resuscitation*. 2004;61:9-21. doi:10.1016/j.resuscitation.2003.12.008.

21. Black JJ, Davies GD. International EMS systems: United Kingdom. *Resuscitation*. 2005;64:21-9. doi:10.1016/j.resuscitation.2004.10.004.
22. Adnet F, Lapostolle F. International EMS systems: France. *Resuscitation*. 2004;63:7-9. doi:10.1016/j.resuscitation.2004.04.001.
23. Trevithick S, Flabouris A, Tall G, Webber CF. International EMS systems: New South Wales, Australia. *Resuscitation*. 2003;59:165-70.
24. Pozner CN, Zane R, Nelson SJ, Levine M. International EMS systems: The United States: past, present, and future. *Resuscitation*. 2004;60:239-44. doi:10.1016/j.resuscitation.2003.11.004.
25. Arnold JL, Dickinson G, Tsai MC, Han D. A survey of emergency medicine in 36 countries. *CJEM*. 2001;3:109-18.
26. Kurla J, Wangel M, Uusaro A, Ruokonen E. Paramedic helicopter emergency service in rural Finland - do benefits justify the cost? *Acta Anaesthesiol Scand*. 2002;46:779-84.
27. Al-Shaqsi S. Models of International Emergency Medical Service (EMS) Systems. *Oman Med J*. 2010;25:320-3. doi:10.5001/omj.2010.92.
28. Dick W. Resuscitation Great. Martin Kirschner: 1879-1942--a surgeon in prehospital care. *Resuscitation*. 2006;68:319-21. doi:10.1016/j.resuscitation.2005.07.004.
29. Luftrettungszentrum, Christoph 1, München. <https://www.christoph-1.de/index.php/geschichte>. Accessed September 17th 2019.
30. Øgar P. NOU 1998:8 Air ambulance service in Norway (Document in Norwegian). SHD,1998.
31. Roudsari BS, Nathens AB, Cameron P, Civil I, Gruen RL, Koepsell TD et al. International comparison of prehospital trauma care systems. *Injury*. 2007;38:993-1000.
32. Bøtger M, Bakke S, Christensen E. A systematic review of controlled studies: do physicians increase survival with prehospital treatment? *Scandinavian journal of trauma, resuscitation and emergency medicine*. 2009;17:12. doi:10.1186/1757-7241-17-12.
33. Lewis R, Booms B. The marketing aspects of service quality. In: Berry L, Shostack G, Upah G, editors. *Emerging perspectives on service marketing*. American Marketing Association Chicago; 1983. p. 99-104.
34. Kudzma EC. Florence Nightingale and healthcare reform. *Nurs Sci Q*. 2006;19:61-4. doi:10.1177/0894318405283556.
35. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966;44:Suppl:166-206.
36. Institute of Medicine. *To Err is Human*. <http://www.iom.edu/1999>.
37. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the Twenty-first Century*. Washington: National Academies Press; 2001.
38. World Health Organization. *The world health report 2000; health systems improving performance,2000*.
39. NHS. *High Quality Care for All: NHS Next Stage Review Final Report*.2011.
40. Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988;260:1743-8.
41. Langley GL, Moen R, Nolan KM, Nolan TW, Norman CL, Provost LP. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance* 2ed. San Fransisco: Jossey-Bass Publishers; 2009.
42. FAMILIESUSA. *Measuring Health Care Quality: An Overview of Quality Measures*. [www.familiesusa.org](http://www.familiesusa.org). Accessed 15.04 2015.
43. Sackett D, Strauss S, Richardson W. *Evidence-Based Medicin: how to practice and teach EBM*. 2. ed. Edinburgh: Churchill Livingstone; 2000.
44. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998;2:i-iv, 1-88.

45. Jones B, Vaux E, Olsson-Brown A. How to get started in quality improvement. *BMJ*. 2019;364:k5408. doi:10.1136/bmj.k5437.
46. Nelson EC, Splaine ME, Batalden PB, Plume SK. Building measurement and data collection into medical practice. *Annals of internal medicine*. 1998;128:460-6. doi:10.7326/0003-4819-128-6-199803150-00007.
47. Spaite DW, Criss EA, Valenzuela TD, Guisto J. Emergency medical service systems research: problems of the past, challenges of the future. *Annals of emergency medicine*. 1995;26:146-52.
48. Grol R. Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. *JAMA*. 2001;286:2578-85.
49. Blumenthal D. Part 1: Quality of care--what is it? *N Engl J Med*. 1996;335:891-4. doi:10.1056/NEJM199609193351213.
50. Hov MR, Zakariassen E, Lindner T, Nome T, Bache KG, Roislien J et al. Interpretation of Brain CT Scans in the Field by Critical Care Physicians in a Mobile Stroke Unit. *J Neuroimaging*. 2018;28:106-11. doi:10.1111/jon.12458.
51. Larsen K, Bache KG, Franer E, Tveit LH, Hov MR, Lund CG et al. Pre-hospital thrombolysis of ischemic stroke in the emergency service system-A case report from the Treat-NASPP trial. *Acta Anaesthesiol Scand*. 2019;63:410-3. doi:10.1111/aas.13285.
52. Samdal M, H. Haugland H, Fjeldet C, Rehn M, Sandberg M. Static Rope Evacuation by Helicopter Emergency Medical Services in Rescue Operations in Southeast Norway. 2018.
53. Porter KM. Letter to the editor. *The Journal of trauma*. 2003;55.
54. Thomas SH, Harrison T, Wedel SK, Thomas DP. Helicopter emergency medical services roles in disaster operations. *Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors*. 2000;4:338-44.
55. Bredmose PP, Forbes RA, Davies GE, Freij R, Lockey D. Medical response times to major incidents: potential benefits of a regional air ambulance mutual aid scheme. *Emergency medicine journal : EMJ*. 2008;25:851-3.
56. Johnsen AS, Fattah S, Sollid SJM, Rehn M. Utilisation of helicopter emergency medical services in the early medical response to major incidents: a systematic literature review. *BMJ Open*. 2016;6:e010307. doi:10.1136/bmjopen-2015-010307.
57. Thomas SH, Arthur AO. Helicopter EMS: Research Endpoints and Potential Benefits. *Emergency medicine international*. 2012;2012:698562.
58. Lov om pasient- og brukerrettigheter (pasient- og brukerrettighetsloven). LOV 1999-07-02-63 (1999).
59. El Sayed MJ. Measuring quality in emergency medical services: a review of clinical performance indicators. *Emergency medicine international*. 2012;161630.
60. Krafft T, Garcia Castrillo-Riesgo L, Edwards S, Fischer M, Overton J, Robertson-Steel I et al. European Emergency Data Project (EED Project): EMS data-based health surveillance system. *Eur J Public Health*. 2003;13:85-90.
61. Herlitz J, Wireklintundstrom B, Bang A, Berglund A, Svensson L, Blomstrand C. Early identification and delay to treatment in myocardial infarction and stroke: differences and similarities. *Scand J Trauma Resusc Emerg Med*. 2010;18:48. doi:10.1186/1757-7241-18-48.
62. Simonsen SA, Andresen M, Michelsen L, Viereck S, Lippert FK, Iversen HK. Evaluation of pre-hospital transport time of stroke patients to thrombolytic treatment. *Scand J Trauma Resusc Emerg Med*. 2014;22:65. doi:10.1186/s13049-014-0065-z.
63. Harmsen AMK, Giannakopoulos GF, Moerbeek PR, Jansma EP, Bonjer HJ, Bloemers FW. The influence of prehospital time on trauma patients outcome: A systematic review. *Injury*. 2015;46:602-9. doi:<https://doi.org/10.1016/j.injury.2015.01.008>.

64. Department of Health. Emergency care ten years on: reforming emergency care. Department of Health, London 2007.
65. Cooke M, Fischer J, McLeod E, et al. Reducing attendances and waits in emergency departments: a systematic review of present innovations NHS Service and Delivery Organisation 2005.
66. Meizoso JP, Valle EJ, Allen CJ, Ray JJ, Jouria JM, Teisch LF et al. Decreased mortality after prehospital interventions in severely injured trauma patients. *J Trauma Acute Care Surg.* 2015;79:227-31. doi:10.1097/TA.0000000000000748.
67. Dissmann PD, Le Clerc S. The experience of Teesside helicopter emergency services: doctors do not prolong prehospital on-scene times. *Emerg Med J.* 2007;24:59-62. doi:10.1136/emj.2006.038844.
68. Garner AA. The role of physician staffing of helicopter emergency medical services in prehospital trauma response. *Emerg Med Australas.* 2004;16:318-23. doi:10.1111/j.1742-6723.2004.00636.x.
69. Gonzalez RP, Cummings G, Mulekar M, Rodning CB. Increased mortality in rural vehicular trauma: identifying contributing factors through data linkage. *J Trauma.* 2006;61:404-9. doi:10.1097/01.ta.0000229816.16305.94.
70. Botker MT, Jacobsen L, Rudolph SS, Knudsen L. The role of point of care ultrasound in prehospital critical care: a systematic review. *Scand J Trauma Resusc Emerg Med.* 2018;26:51. doi:10.1186/s13049-018-0518-x.
71. Skaga NO, Eken T, Jones JM, Steen PA. Different definitions of patient outcome: consequences for performance analysis in trauma. *Injury.* 2008;39:612-22.
72. Spaite DW, Maio R, Garrison HG, Desmond JS, Gregor MA, Stiell IG et al. Emergency Medical Services Outcomes Project (EMSOP) II: developing the foundation and conceptual models for out-of-hospital outcomes research. *Annals of emergency medicine.* 2001;37:657-63.
73. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ.* 2008;337:a1655. doi:10.1136/bmj.a1655.
74. Donchin Y, Gopher D, Olin M, Badihi Y, Biesky M, Sprung CL et al. A look into the nature and causes of human errors in the intensive care unit. *Crit Care Med.* 1995;23:294-300.
75. Wang HE, Peitzman AB, Cassidy LD, Adelson PD, Yealy DM. Out-of-hospital endotracheal intubation and outcome after traumatic brain injury. *Annals of emergency medicine.* 2004;44:439-50.
76. Davis DP, Peay J, Serrano JA, Buono C, Vilke GM, Sise MJ et al. The impact of aeromedical response to patients with moderate to severe traumatic brain injury. *Annals of emergency medicine.* 2005;46:115-22. doi:10.1016/j.annemergmed.2005.01.024.
77. McEachin CC, McDermott JT, Swor R. Few emergency medical services patients with lower-extremity fractures receive prehospital analgesia. *Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors.* 2002;6:406-10.
78. Brown JB, Stassen NA, Bankey PE, Sangosanya AT, Cheng JD, Gestring ML. Helicopters and the civilian trauma system: national utilization patterns demonstrate improved outcomes after traumatic injury. *The Journal of trauma.* 2010;69:1030-4; discussion 4-6.
79. Hoffman JR, Cooper RJ. Overdiagnosis of disease: a modern epidemic. *Arch Intern Med.* 2012;172:1123-4. doi:10.1001/archinternmed.2012.3319.
80. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA.* 2012;307:1513-6. doi:10.1001/jama.2012.362.
81. Kruger AJ, Lockey D, Kurola J, Di Bartolomeo S, Castren M, Mikkelsen S et al. A consensus-based template for documenting and reporting in physician-staffed pre-hospital services. *Scand J Trauma Resusc Emerg Med.* 2011;19:71. doi:10.1186/1757-7241-19-71.

82. Lossius HM, Kruger AJ, Ringdal KG, Sollid SJ, Lockey DJ. Developing templates for uniform data documentation and reporting in critical care using a modified nominal group technique. *Scand J Trauma Resusc Emerg Med.* 2013;21:80. doi:10.1186/1757-7241-21-80.
83. G K. ADACOPTER-Aufzeichnungen einer Entwicklung München. Werner Wolfseffner Verlag; 2002.
84. Kelley Eea. Health Care Quality Indicators Project Conceptual Framework Paper2006 Contract No.: 23.
85. Quality Indicator Measure Development, Implementation, Maintenance, and Retirement. Agency for Healthcare Research and Quality2011.
86. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A et al. The quality of health care delivered to adults in the United States. *N Engl J Med.* 2003;348:2635-45. doi:10.1056/NEJMsa022615.
87. Turner J, Siriwardena AN, Coster J, Jacques R, Irving A, Crum A et al. Developing new ways of measuring the quality and impact of ambulance service care: the PhOEBE mixed-methods research programme. Programme Grants for Applied Research. Southampton (UK)2019.
88. Fischer M, Krep H, Wierich D, Heister U, Hoeft A, Edwards S et al. [Comparison of the emergency medical services systems of Birmingham and Bonn: process efficacy and cost effectiveness]. *Anesthesiol Intensivmed Notfallmed Schmerzther.* 2003;38:630-42. doi:10.1055/s-2003-42507.
89. Ringburg AN, Buljac M, Stolk EA, van Lieshout EM, van Beeck EF, Patka P et al. Willingness to pay for lives saved by Helicopter Emergency Medical Services. *Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors.* 2009;13:37-43. doi:10.1080/10903120802472004.
90. Osteras O, Brattebo G, Heltne JK. Helicopter-based emergency medical services for a sparsely populated region: A study of 42,500 dispatches. *Acta Anaesthesiol Scand.* 2016;60:659-67. doi:10.1111/aas.12673.
91. American College of Surgeons Committee on Trauma. Resources for optimal care of the injured patient 2014. Chicago (IL)2014.
92. Giannakopoulos GF, Lubbers WD, Christiaans HM, van Exter P, Bet P, Hugen PJ et al. Cancellations of (helicopter-transported) mobile medical team dispatches in the Netherlands. *Langenbecks Arch Surg.* 2010;395:737-45. doi:10.1007/s00423-009-0576-7.
93. Lawless J, Tallon JM, Petrie D. Aborted air medical missions: a 4-year quality review of a Canadian province-wide air medical program. *Air Med J.* 2005;24:79-82. doi:10.1016/j.amj.2004.12.006.
94. Osteras O, Heltne JK, Tonsager K, Brattebo G. Outcomes after cancelled helicopter emergency medical service missions due to concurrencies: a retrospective cohort study. *Acta Anaesthesiol Scand.* 2018;62:116-24. doi:10.1111/aas.13028.
95. Moore L. Measuring quality and effectiveness of prehospital EMS. *Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors.* 1999;3:325-31.
96. Norwegian Directorate of Health. The Nordic Emergency Medical Services. Project on data collection and benchmarking 2014-2018.2019.
97. Gruen RL, Gabbe BJ, Stelfox HT, Cameron PA. Indicators of the quality of trauma care and the performance of trauma systems. *Br J Surg.* 2012;99 Suppl 1:97-104. doi:10.1002/bjs.7754.
98. Gisvold SE, Fasting S. How do we know that we are doing a good job - can we measure the quality of our work? *Best Pract Res Clin Anaesthesiol.* 2011;25:109-22. doi:10.1016/j.bpa.2011.02.010.
99. Petros AJ, Marshall JC, van Saene HK. Should morbidity replace mortality as an endpoint for clinical trials in intensive care? *Lancet.* 1995;345:369-71. doi:10.1016/s0140-6736(95)90347-x.
100. Ospina-Tascon GA, Buchele GL, Vincent JL. Multicenter, randomized, controlled trials evaluating mortality in intensive care: doomed to fail? *Crit Care Med.* 2008;36:1311-22. doi:10.1097/CCM.0b013e318168ea3e.



101. Kress JP. Mortality is the only relevant outcome in ARDS: no. *Intensive Care Med.* 2015;41:144-6. doi:10.1007/s00134-014-3563-6.
102. Mikkelsen S, Lossius HM, Toft P, Lassen AT. Characteristics and prognoses of patients treated by an anaesthesiologist-manned prehospital emergency care unit. A retrospective cohort study. *BMJ Open.* 2017;7:e014383. doi:10.1136/bmjopen-2016-014383.
103. Sovso MB, Hermansen SB, Faerk E, Lindskou TA, Ludwig M, Moller JM et al. Diagnosis and mortality of emergency department patients in the North Denmark region. *BMC Health Serv Res.* 2018;18:548. doi:10.1186/s12913-018-3361-x.
104. Christensen EF, Larsen TM, Jensen FB, Bendtsen MD, Hansen PA, Johnsen SP et al. Diagnosis and mortality in prehospital emergency patients transported to hospital: a population-based and registry-based cohort study. *BMJ Open.* 2016;6:e011558. doi:10.1136/bmjopen-2016-011558.
105. Christensen EF, Bendtsen MD, Larsen TM, Jensen FB, Lindskou TA, Holdgaard HO et al. Trends in diagnostic patterns and mortality in emergency ambulance service patients in 2007-2014: a population-based cohort study from the North Denmark Region. *BMJ Open.* 2017;7:e014508. doi:10.1136/bmjopen-2016-014508.
106. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf.* 2016;25:986-92. doi:10.1136/bmjqs-2015-004411.
107. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *International Journal for Quality in Health Care.* 2000;12:281-95. doi:10.1093/intqhc/12.4.281.
108. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. *International Journal for Quality in Health Care.* 2003;15:i5-i11. doi:10.1093/intqhc/mzg084.
109. Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators--a systematic review. *Implementation Science.* 2012;7:21. doi:10.1186/1748-5908-7-21.
110. Marshall MN, Shekelle PG, McGlynn EA, Campbell S, Brook RH, Roland MO. Can health care quality indicators be transferred between countries? *Quality and Safety in Health Care.* 2003;12:8-12. doi:10.1136/qhc.12.1.8.
111. Weiss D, Dunn SI, Sprague AE, Fell DB, Grimshaw JM, Darling E et al. Effect of a population-level performance dashboard intervention on maternal-newborn outcomes: an interrupted time series study. *BMJ Qual Saf.* 2018;27:425-36. doi:10.1136/bmjqs-2017-007361.
112. Callahan CD, Griffen DL. Advanced statistics: applying statistical process control techniques to emergency medicine: a primer for providers. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine.* 2003;10:883-90.
113. Leotsakos A, Zheng H, Croteau R, Loeb JM, Sherman H, Hoffman C et al. Standardization in patient safety: the WHO High 5s project. *International Journal for Quality in Health Care.* 2014;26:109-16. doi:10.1093/intqhc/mzu010.
114. Alderwick H JB, Charles A, Warburton W. Making the case for quality improvement: lessons for NHS boards and leaders. [www.kingsfund.org.uk/publications/better-value-nhs](http://www.kingsfund.org.uk/publications/better-value-nhs) (Accessed November 1st 2019): The King's Fund, London 2017.
115. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: a method for continuous quality improvement in health. *Healthc Policy.* 2012;7:e101-19.

## 17 Appendix A

### Quality indicators for physician-staffed emergency medical services

	Quality indicator
1	Was the P-EMS unit <i>able</i> to respond immediately to the actual response?
2	What is the time interval after the dispatch center receives the alarm call and until the P-EMS unit arrives at the patient?
3	What is the time interval after the P-EMS unit arrives at the patient and until transportation of patient is initiated?
4	What is the time interval after the P-EMS unit received the alarm call and until the patient was delivered to the preferred destination?
5	Did the patient arrive at the hospital alive?
6	Was the P-EMS response debriefed?
7	Did you experience any adverse events during the P-EMS response?
8	Are all defined key variables measured and documented in the patient chart?
9	Was a physician and/or paramedic from P-EMS involved in deciding if the P-EMS unit should be dispatched to the particular job or not?
10	Without the assistance from the P-EMS unit: Do you consider that the level of competence on scene was sufficient to give the patient appropriate care?
11	Did the service have a guideline for the medical problem encountered in the response?
12	Did the P-EMS service provide advanced treatment during the actual response?
13	Did the logistical contribution by P-EMS give the patient significantly better service than the existing alternative?
14	Was the patient enrolled in a scientific study involving the pre-hospital care?
15	Did you ensure that the relatives' needs were addressed, either by P-EMS or by collaborating services?

**Appendix A:** Quality indicators for physician-staffed emergency medical services. For a detailed description of each quality indicator, please see the definition catalogue from paper II in which the quality indicators were developed.



18 Papers I-IV



# Paper I



## Review

# Quality measurement in physician-staffed emergency medical services: a systematic literature review

HELGE HAUGLAND<sup>1,2,3</sup>, ODDVAR ULEBERG<sup>1,2,3</sup>, PÅL KLEPSTAD<sup>3,4</sup>,  
ANDREAS KRÜGER<sup>1,2,3</sup>, and MARIUS REHN<sup>1,5,6</sup>

<sup>1</sup>Department of Research and Development, Norwegian Air Ambulance Foundation, 1441 Drøbak, Norway, <sup>2</sup>Department of Emergency Medicine and Pre-Hospital Services, St. Olavs Hospital, 7030 Trondheim, Norway, <sup>3</sup>Department of Circulation and Medical Imaging, Medical Faculty, NTNU, Norwegian University of Science and Technology, 7491 Trondheim, Norway, <sup>4</sup>Department of Anaesthesiology and Intensive Care, St. Olav University Hospital, 7030 Trondheim, Norway, <sup>5</sup>Department of Health Studies, University of Stavanger, 4021 Stavanger, Norway, and <sup>6</sup>Division of Emergencies and Critical Care, Department of Anaesthesia, Oslo University Hospital, 0424 Oslo, Norway

Address reprint requests to: Helge Haugland, Department of Research and Development, Norwegian Air Ambulance Foundation, 1441 Drøbak, Norway. Tel: 0047-41128218; Fax: 0047-73869750; Email: helge.haugland@norskluftambulans.no

Editorial Decision 10 April 2018; Accepted 25 April 2018

## Abstract

**Purpose:** Quality measurement of physician-staffed emergency medical services (P-EMS) is necessary to improve service quality. Knowledge and consensus on this topic are scarce, making quality measurement of P-EMS a high-priority research area. The aim of this review was to identify, describe and evaluate studies of quality measurement in P-EMS.

**Data sources:** The databases of MEDLINE and Embase were searched initially, followed by a search for included article citations in Scopus.

**Study selection:** The study eligibility criteria were: (1) articles describing the use of one quality indicator (QI) or more in P-EMS, (2) original manuscripts, (3) articles published from 1 January 1968 until 5 October 2016. The literature search identified 4699 records. 4543 were excluded after reviewing title and abstract. An additional 129 were excluded based on a full-text review. The remaining 27 papers were included in the analysis. Methodological quality was assessed using an adapted critical appraisal tool.

**Data extraction:** The description of used QIs and methods of quality measurement was extracted. Variables describing the involved P-EMSs were extracted as well.

**Results of data synthesis:** In the included papers, a common understanding of which QIs to use in P-EMS did not exist. Fifteen papers used only a single QI. The most widely used QIs were 'Adherence to medical protocols', 'Provision of advanced interventions', 'Response time' and 'Adverse events'.

**Conclusion:** The review demonstrated a lack of shared understanding of which QIs to use in P-EMS. Moreover, papers using only one QI dominated the literature, thus increasing the risk of a narrow perspective in quality measurement. Future quality measurement in P-EMS should rely on a set of consensus-based QIs, ensuring a comprehensive approach to quality measurement.

**Key words:** quality measurement < quality management, quality improvement < quality management, emergency care < setting of care



## Background

Emergency medical services (EMS) provide pre-hospital treatment and transportation to definitive care for patients in need of urgent medical care. EMSs are well integrated in health services in most countries and normally consist of ground ambulances staffed by paramedics, emergency medical technicians or nurses [1–6]. As a supplement to regular EMS, physician-staffed rapid response vehicles or helicopters exist in some areas [7]. Although the operational concept of this physician-staffed EMS (P-EMS) may differ, a common feature is the involvement of a specially trained physician in pre-hospital care of critically ill or injured patients. Depending on the country, these physicians are often anaesthesiologists, surgeons, internists or emergency physicians [1, 3].

The focus on quality measurement in healthcare is increasing [8–12]. As an example, quality dimensions such as efficiency, patient-centeredness and safety have been assessed in several emergency departments [13–15]. In P-EMS, a valid model for quality assessment is needed to achieve appropriate governance, quality assurance and quality improvement [16]. Snooks *et al.* [17] define the development of meaningful quality indicators (QIs) for EMS as the most important issue for future research in emergency pre-hospital care. For P-EMS, measuring quality of care is considered a priority area of research [18].

Quality measurement can be defined as measuring the extent to which set targets are achieved [19]. A QI is used to measure performance against a recognized standard of care. Donabedian defines three categories of QIs: structure, process and outcome of healthcare [20, 21]. Structure indicators describe the infrastructure of a healthcare system, such as competence of the staff, equipment and deployment and response times. Process indicators evaluate the care provided to the patient, and outcome indicators address the change in patient health status. None of these categories of indicators provide a complete description of the quality of care but address single components. Thus, different types of QIs should be combined to assess the quality of a service [19].

QIs inform clinicians and organizations how the health system performs and aid in the improvement in care. Ideally, all QIs are based upon evidence of their relevance and importance. The process of developing QIs generally includes stakeholders who evaluate the evidence and define the QI parameters [22]. These QIs for P-EMS should be evaluated against patient-oriented outcomes, e.g. pain intensity, morbidity or mortality. However, P-EMS quality can also relate to system factors such as training of traditional EMS, major incident management and the concept of providing equity of access to healthcare. Different stakeholders have different perspectives on what represents quality in healthcare [23, 24], and various QIs for P-EMS are possible.

A widely cited definition of quality that also might be applicable for P-EMS systems is 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' [9]. Often, a few high impact clinical conditions are identified (in terms of morbidity, mortality, costs and incidence). These conditions are termed 'tracer conditions'. Examples of tracer conditions for emergency medicine are cardiac arrest or trauma patients with severely reduced consciousness (Glasgow Coma Score < 8). Measuring the outcomes of tracer conditions can predict a system's response to other clinical states and the overall quality of a service [25].

This systematic review aims to identify, describe and appraise the methodological quality of the literature pertaining to the quality assessment of P-EMS.

## Methods

For the purpose of this review, physicians who staff P-EMS should be trained in critical care, exceeding the competency of a general practitioner on call [26]. Moreover, we define the term 'pre-hospital' as relating to procedures administered or care provided prior to patient arrival at the hospital [8]. The studies identified in the review do not address the potential benefit from P-EMS compared to other EMS.

### Literature search strategy

A systematic literature search of MEDLINE and Embase to identify relevant literature was conducted (see Additional file 1 for search strategy). Four different sets of entry terms were applied and combined. These entry terms describe pre-hospital setting, emergency care, physician staffing and finally the concept of quality measurement. All records were collated in an Endnote bibliographic database (©2007 Thompson Reuters).

The study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines, including the PICOS (Population, Intervention, Comparator, Outcome, Study design) methodology [27]. Here, the participants were all the identified articles describing quality estimation in P-EMS. Our evaluation of intervention, comparison and outcomes was carried out using the data extraction and quality appraisal variables in Tables 1 and 3. The study was registered at PROSPERO (<http://www.crd.york.ac.uk/prospero/>, registration number CRD42015024421).

### Inclusion/exclusion criteria

Articles were included in the systematic literature review if they fulfilled all the following criteria: (1) literature describing methods of quality estimation in P-EMS, i.e. the use of one or more QIs based on quantitative methods, qualitative methods or both; (2) original manuscripts and (3) literature published after 1 January 1968 and until the date of the literature search, 5 October 2016. The rationale for including literature from 1 January 1968, is the establishment of the world's first civilian physician-staffed helicopter EMS in Munich this year [28].

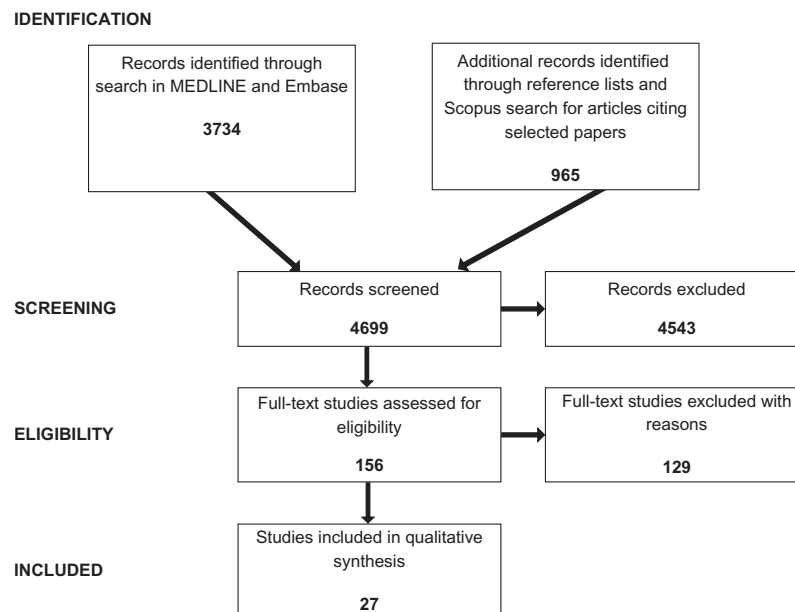
Articles without abstract, book chapters, editorials, comments and letters to the editor were excluded. Articles in English, French, German and Scandinavian languages were identified. The translation competency of these languages was present in the author group.

### Literature identification

The records from the literature search were exported to [www.covidence.org](http://www.covidence.org). Here, all titles and corresponding abstracts were

**Table 1** Reasons for excluding 129 out of 156 full-text studies in the eligibility-check of the systematic review

Reasons for exclusion	No.
Wrong study design	77
Not about quality measurement in P-EMS	29
Not enough information for quality appraisal and data extraction	9
Only abstracts	6
Not about P-EMS	3
Comparative studies pertaining to new procedures	2
Commentary, letter to editor or editorial	1
Not original article	1
Duplicate	1



**Figure 1** Information flow through the different phases of the systematic review.

screened independently by two of the authors for eligibility (O.U. and H.H.) (Figure 1). In the case of uncertainty, a third reviewer (M.R.) reviewed the title/abstract. Articles clearly not meeting the inclusion criteria were excluded. Articles accepted for full-text screening were assessed in pairs of authors (O.U. and M.R., A.J.K. and H.H.) using the inclusion/exclusion criteria listed above. Excluded articles were listed with the reason for exclusion. If there was any uncertainty about whether an article should be included, there was a discussion until consensus was reached among the authors. One author (H.H.) performed data extraction and quality appraisal and consulted another author (M.R.) in uncertain cases. Further, one author hand-searched references in included articles to identify additional relevant articles (H.H.). Finally, a search in Scopus was conducted to identify articles citing the included articles.

Prior to the literature search, the authors made templates for data extraction and quality appraisal. The data extraction and quality appraisal variables were based on the authors' assumptions on what is important to report in quality measurement studies in P-EMS. However, these variables do not represent a reference standard, since such a standard does not exist, to our knowledge. As a part of data extraction, fixed-system variables were included. Fixed-system variables relate to system characteristics concerning the organization, staffing and operational capacities of the service and are necessary for interpreting the results [29].

## Results

A total of 4699 articles were identified by the search strategy, 156 of which were accepted for full-text screening. Of these, 129 articles were excluded. The main reason for excluding articles were 'Wrong study design', pertaining to articles exclusively comparing different treatment modalities and without any quality measurement objectives

(Table 1). A total of 26 articles from the main database search were included for data extraction and quality appraisal [30–55]. One additional article was included from the Scopus search for citing articles [56]. A review of the literature lists of included articles did not result in additional findings.

None of the papers gave a complete report of fixed-system variables, thus complicating the comparison of involved P-EMS concepts (Table 2). Twenty-four of the 27 papers use QIs that can be identified as process indicators. Structure indicators and outcome indicators are used less frequently, in two and seven papers, respectively. Twenty-five different QIs were identified, all of which were considered suitable for international use and transferable to other P-EMSs (Table 3). The most widely used QI was 'Adherence to medical protocols'. This QI measures if medical guidelines are followed, as done by Viergutz *et al.* [53] who investigated whether guidelines for preclinical care of patients with traumatic head injury were followed. The second most used QI was 'Provision of advanced interventions', investigating if the P-EMS unit provided treatment that exceeds the competences of the attending EMS, as done by Mikkelsen *et al.* [44]. The two following QIs are 'Response time' and 'Adverse events', the latter exemplified by Nakstad *et al.* [45], who studied the incidence of desaturation during pre-hospital rapid sequence intubation. Fifteen papers used one single QI, and twelve papers applied a set of QIs. Moreover, three papers used tracer conditions as their approach to quality assessment (Table 2). Pertaining to the internal validity of the papers, ten of the 27 papers did not clearly explain the methodology for developing the QIs (Table 4).

## Discussion

This systematic review identified 27 papers that reported the use of QIs in P-EMS. Fifteen papers used one single QI, and twelve papers

**Table 2** Data extraction of included articles

	Are the following fixed-system variables reported?												Multiple, mixed quality indicators	Tracer conditions used
	Quality indicators used													
	Transportation mode	SAR-capability	Rural and/or urban	Response types	Speciality of physician	Physician experience	Funding	Description of dispatch system	Other fixed-system variables	Structure	Process	Outcome		
Akin Paker [30]	✓	×	×	✓	×	×	×	×	×	✓	×	×	✓	×
Anadolic [31]	✓	×	✓	Partly	×	×	✓	✓	✓	✓	×	×	✓	×
Arntz [32]	✓	×	✓	✓	Partly	×	✓	✓	✓	✓	×	×	✓	×
Breckwoldt [33]	✓	×	✓	×	✓	×	×	✓	×	✓	✓	×	×	×
Duchateau [34]	✓	×	✓	×	×	×	×	✓	×	✓	×	×	×	×
Fjeldstad [35]	✓	×	✓	✓	×	×	×	✓	×	✓	×	×	×	×
Flabouris [36]	✓	✓	×	✓	✓	×	✓	✓	×	✓	×	×	✓	×
Helm [37]	✓	×	×	×	×	×	×	×	×	✓	×	×	×	×
Helm [38]	✓	×	×	×	✓	×	×	×	×	✓	×	×	×	×
Hennes [39]	✓	×	✓	✓	×	×	×	×	×	×	✓	×	×	×
Klemenc-Ketis [40]	✓	✓	×	✓	×	×	✓	×	×	✓	×	×	✓	×
Leicht [41]	✓	×	×	✓	✓	×	Partly	×	×	✓	×	×	×	×
Lossius [42]	✓	×	✓	✓	✓	×	×	×	×	✓	×	×	×	×
Messelken [43]	✓	×	✓	✓	✓	×	×	×	×	✓	×	×	✓	×
Mikkelsen [44]	✓	×	✓	✓	✓	×	×	×	×	✓	×	×	×	×
Nakstad [45]	✓	×	✓	✓	✓	×	×	×	×	✓	×	×	×	×
Neukamm [46]	✓	×	✓	✓	×	×	Partly	✓	✓	✓	×	×	×	×
Pedersen [47]	✓	×	✓	✓	×	×	×	✓	✓	✓	×	×	✓	×
Regel [48]	✓	×	✓	✓	×	×	×	✓	×	✓	×	×	✓	×
Rognås [49]	✓	×	✓	✓	✓	✓	✓	✓	×	✓	×	×	×	×
Rognås [50]	✓	×	✓	✓	✓	×	Partly	✓	✓	✓	×	×	×	×
Sollid [51]	✓	×	✓	✓	✓	×	×	✓	×	✓	×	×	×	×
Van der Velden [52]	✓	×	✓	✓	×	×	✓	✓	×	✓	×	×	×	×
Viergut [53]	✓	×	✓	✓	×	×	×	✓	×	✓	×	×	✓	×
Von Knobelsdorff [54]	✓	×	✓	✓	×	×	×	✓	×	✓	✓	×	✓	×
Weltermann [55]	✓	×	✓	✓	×	×	×	×	×	✓	×	×	✓	×
Schlechterien [56]	✓	×	×	×	×	×	×	×	×	✓	×	×	✓	✓

✓ = yes, × = no.

**Table 3** Quality indicators used in the included literature

Quality indicator	Category	No. of papers it is used in
Adherence to medical protocols	Process	11
Provision of advanced interventions	Process	8
Response time	Process	7
Adverse events	Process	7
Medication administration	Process	5
Transport to appropriate facility	Process	4
Time on scene	Process	3
Improved care due to clinical decision making	Process	3
Reliability of the primary diagnosis made by the P-EMS physician	Process	3
Survival	Outcome	3
ROSC in cardiac arrest	Outcome	3
Time from alarm to patient handover	Process	2
Time from arrival at patient until hospital admission	Process	2
Altered physiology	Outcome	2
Pain management	Outcome	2
Proportion of intubated patients adequately oxygenated and ventilated	Process	2
Time gain by air transportation	Process	1
The number of patients with a NACA-score $\geq 4$ with an intravenous line	Process	1
The proportion of patients successfully intubated	Process	1
Life years gained	Outcome	1
Morbidity/disability	Outcome	1
Amount of yearly CPR training	Structure	1
Precision of dispatch	Process	1
Rate of CPR started within 8 min of the call to the dispatch center	Process	1
Patient satisfaction	Outcome	1

ROSC, return of spontaneous circulation; NACA, National Advisory Committee for Aeronautics (see Additional file 2); CPR, cardiopulmonary resuscitation.

applied a set of QIs. Twenty-four of the 27 papers used QIs pertaining to process. Generally, the systematic review demonstrated a lack of a shared understanding of which QI to use in P-EMS.

The lack of agreement on QIs has also been described for regular EMS [57, 58]. This lack of consensus on QI, the heterogeneity of diagnosis, and the challenge of isolating the effect of pre-hospital care from the effect of in-hospital care complicates quality measurement in pre-hospital emergency medicine [59, 60].

The characteristics of the P-EMSs in the identified literature vary and are described sufficiently in only ten papers. This complicates the comparison of studies because the concept of involved P-EMSs remains unclear. Five services use helicopters, eleven use rapid response cars, and another eleven use both transportation modes. Further, nine services are urban, one is rural and eleven are both urban and rural. This mixed representation of transportation modes and urban versus rural profile seems to reflect the heterogeneity of P-EMS. Another relevant aspect of P-EMS is the specialty training and competency of the staffing physicians. Fourteen papers do not report the physicians' medical specialty. Emergency medicine and anesthesiology are reported as the physicians' specialties in one and seven papers, respectively. Five studies report a mix of medical specialties in the actual services. All mixes are different, with anesthesiology as the only medical specialty represented in all five papers. Regarding country, eleven of the 27 papers are German. Germany has a long history of P-EMS as an integrated and natural part of the emergency medical system. However, the considerable contribution of German papers is not only because of the service's long existence. Quality measurement in German hospitals has developed substantially during the last two decades [61], and gradually, German P-EMS has adopted this quality measurement initiative. In addition to a formalized and common understanding of the need for quality measurement in P-EMS [62], the establishment of common

documentation systems seems to have been the first necessary step towards quality measurement in German P-EMS and might set an example for QI work in P-EMS in other countries [63, 64].

In 23 papers, the QIs are well defined. However, for many of the QIs identified in this systematic review, the development process does not seem optimal. Ten papers do not clearly describe the methodology for developing the QIs, and only in two papers were the QIs systematically developed by a group of experts. Finally, only seven papers report the professional background and funding of those involved in the development of QIs. A key point in the development of QIs is a systematical and objective approach; this allows for the assessment of the evidence base for the QIs and secures legitimacy. Inadequate QI development may influence the validity, reliability and feasibility of the QIs.

Structure indicators and outcome indicators are used by only two and seven papers, respectively. This may be because process indicators are easier to collect. Routine use of outcome indicators will require automated data exchange between pre- and in-hospital databases, which might not be feasible. This calls for a more integrated electronic patient chart system covering patient data capture through all phases of care. Nevertheless, the major role of process indicators in quality measurement literature of P-EMS is as expected. Process indicators are considered useful for short time frames and when it is difficult to adjust for patient factors [65], and process indicators are therefore particularly relevant for P-EMS. Moreover, process indicators provide a direct assessment of quality of care, as opposed to structure and outcome indicators, which measure care quality by an indirect approach [66]. Process indicators are easy to interpret and well suited for the evaluation of adherence to medical protocols and other quality improvement programmes.

As seen in Table 3, many of the process indicators are time variables. When setting targets and measuring EMS quality, time

Table 4 Quality appraisal of included articles

	Internal validity				External validity										
	Was the methodology for developing the QIs clearly explained?	Are the QIs clearly defined?	Is the rationale for the QIs sufficiently described?	Is handling of missing data sufficiently described?	Do the authors address conflicts of interest?	Has an ethics committee approved the study?	Is professional background and funding of those involved in QI development stated?	Were the QIs developed by the systematic work of a group of experts?	Is it reported in which country/organization the QIs were developed?	Is (are) the P-EMS (s) involved in quality measurement sufficiently described?	Are the QIs transferable to other countries or P-EMSs?	Are limitations of the study discussed?	Are possible sources of bias discussed?	Have the QIs been used in other publications?	Was the feasibility of the QIs evaluated?
Akin Paker [30]	✓	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	?	✓
Anadolic [31]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Arntz [32]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Breckwoldt [33]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Duchateau [34]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	✓	✓
Fjaeldstad [35]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Flabouris [36]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Helm [37]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Helm [38]	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?	✓
Hennes [39]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	✓	✓
Klemenc-Ketiš [40]	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?	✓
Leicht [41]	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?	✓
Lossius [42]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	✓	✓
Messelken [43]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	✓	✓
Mikkelsen [44]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Nakstad [45]	✓	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓
Neukamm [46]	✓	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	?	✓
Pedersen [47]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	✓
Regel [48]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	✓	✓
Rognås [49]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rognås [50]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sollid [51]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	✓
Van der Velden [52]	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?	✓
Viergutz [53]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	✓
Von Knobelsdorff [54]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	?	✓
Wettermann [55]	✓	✓	✓	✓	✓	?	✓	?	✓	✓	✓	✓	✓	✓	✓
Schlechtriemen [56]	✓	✓	✓	–	✓	–	✓	?	✓	–	✓	✓	✓	?	✓

✓ = yes; ✗ = no; ? = uncertain; – = not relevant.

variables have been widely applied [17, 66]. EMS evolved as a response to the need for rapid access to healthcare in time critical conditions (war injuries, cardiac arrest, major trauma) [66, 67]. However, the current patient population for P-EMS is increasingly heterogeneous, including conditions that are not highly time critical. For a high proportion of patients, shorter pre-hospital time intervals do not improve outcome, and in some cases, longer on-scene time for initial treatment and stabilization is desirable [68, 69]. Accordingly, if P-EMS quality measurement focuses too much on time variables, the results will have poor relevance for a high proportion of the services' patient population. Moreover, it can lead to an undesired attention shift, resulting in decreased quality for quality dimensions not measured. However, this undesired attention shift is not specific for time variables but rather a universal challenge in quality measurement when the used QIs are few. Finally, our findings indicate that literature pertaining to pre-hospital time variables often lacks information about the competency of the responding unit and the quality of provided care, resulting in even less information from the time variables measured.

Adherence to medical protocols is the most frequently used QI in the included papers. Guidelines and protocols are developed to improve quality of care and to reduce unwanted variation in care. Quality measurement can be defined as measuring the extent to which set targets are achieved [19], and to explore the gap between guidelines and clinical practice is an adequate approach in quality measurement. Ebben *et al.* [70] demonstrated a wide variation in different EMS professionals' adherence to guidelines and protocols, indicating that a substantial number of patients do not receive appropriate pre-hospital care. However, this conclusion presumes that there is an evident relationship between adherence to guidelines and patient outcome. Few studies have explored this relationship, and in the review by Ebben *et al.*, only three studies showed that adherence to guidelines improved patient outcomes [71–73]. Finally, it is recommended that guidelines should define QIs to aid monitoring and assessment of guideline adherence [74, 75]. Thus, QIs should ideally be a part of the guideline development process.

For the fifteen papers relying on only one single QI, there is a risk for narrowing the perspective on quality, and important aspects of quality in the actual healthcare service may be ignored. Using a set of mixed QIs that cover different aspects of the service is preferable [76]. Three papers evaluated the care quality of selected 'tracer conditions', i.e. high-priority clinical conditions [77]. The evaluation of a service's response to tracer conditions with condition-specific QIs is used to predict the overall performance of P-EMS. The chosen tracer conditions are not identical, illustrating the following challenge: when different clinical conditions are used for quality measurement, it may complicate the evaluation and comparison of the quality of different P-EMS. To overcome this, the Institute for Healthcare Improvement (IHI) has recommended the use of 'whole system measures', defined as a set of QIs aligned with the Institute of Medicine's (IOM's) six dimensions of quality, which are not disease- or condition-specific [78, 79]. The six quality dimensions that define high-quality care are timeliness, safety, efficiency, equity, effectiveness and patient-centeredness. Each of these is distinct, and all are equally important. To obtain an adequate and comprehensive quality measurement of P-EMS, future quality measurement should therefore cover these six dimensions as far as possible. This will require the use of multiple QIs, developed specifically for P-EMS. A set of QIs for this purpose has been developed recently, with IOM's six quality dimensions as the most important part of the conceptual framework [76].

## Strengths and limitations

We recognize some strengths and limitations of this study. A strength is that in the literature search, several languages were eligible, allowing inclusion of non-English literature. This is important as eleven of the 27 identified papers were non-English. First, a limitation is that as always in a systematic review, screening, eligibility-check and qualitative synthesis of literature is a product of the review authors' judgements, allowing subjective interpretations of the content of the studies. Although screening and eligibility assessment were conducted in pairs, data extraction and quality appraisal were primarily conducted by one author. Moreover, no established tools for data extraction or quality appraisal were available. Second, some of the included studies are not explicitly presented as quality measurement studies. However, when screening the literature, we recognized that some studies are quality measurement studies despite not using the quality measurement terminology. A possible reason for this might be the fact that quality measurement terminology is still quite new in P-EMS and not widely used. Accordingly, we have focused on the actual content of the screened literature when deciding if a study should be included—not on the presence or absence of correct quality measurement terminology. Doing this, we have appraised if a paper 'concerns an initiative to improve healthcare; broadly defined to include the quality, safety, effectiveness, patient-centeredness, timeliness, cost, efficiency and equity of healthcare', as stated by the SQUIRE guidelines [80]. This approach was chosen to avoid overlooking potentially relevant aspects of quality measurement studies in P-EMS.

## Conclusion

This systematic literature review served the purpose of identifying, describing and evaluating studies of quality measurement in P-EMS. The review demonstrated a lack of a shared understanding of which QIs to use in P-EMS. Process indicators were dominant in the included papers, and the most emphasized QIs were 'Adherence to medical protocols', 'Provision of advanced interventions', 'Response time' and 'Adverse events'. Moreover, fifteen of the 27 papers used only a single QI to measure quality, thus increasing the risk of a narrow perspective on quality. The remaining papers used multiple QIs in their quality measurement, which is considered preferable.

Future quality measurement in P-EMS should rely on a set of consensus-based QIs, securing a comprehensive approach to quality measurement and offering the possibility of comparing results from different P-EMS systems.

## Supplementary material

Supplementary material is available at *International Journal for Quality in Health Care* online.

## Acknowledgements

The authors would like to thank Mrs Ingrid Ingeborg Riphagen, Unit for Applied Clinical Research, Norwegian University of Science and Technology, for her competent and invaluable assistance in designing and conducting the literature search. We would also like to thank members and donors of the Norwegian Air Ambulance Foundation for their financial contribution to this study.

## Additional file

Additional file 1: 'Search strategy'.

Additional file 2: 'NACA-score; elaboration document'.

## Authors' contribution

H.H. and M.R. conceived the idea. All authors were part of the study design. H.H. and O.U. screened the identified literature. H.H., O.U., A.J.K. and M.R. considered the eligibility of uncertain literature. H.H. performed data extraction and quality appraisal of the included literature. All authors have approved the final version of the manuscript.

## Ethics approval and consent to participate

Not applicable.

## References

- Roessler M, Zuzan O. EMS systems in Germany. *Resuscitation* 2006;**68**: 45–9. doi:10.1016/j.resuscitation.2005.08.004.
- Langhelle A, Lossius HM, Silfvast T *et al.* International EMS Systems: the Nordic countries. *Resuscitation* 2004;**61**:9–21. doi:10.1016/j.resuscitation.2003.12.008.
- Black JJ, Davies GD. International EMS systems: United Kingdom. *Resuscitation* 2005;**64**:21–9. doi:10.1016/j.resuscitation.2004.10.004.
- Adnet F, Lapostolle F. International EMS systems: France. *Resuscitation* 2004;**63**:7–9. doi:10.1016/j.resuscitation.2004.04.001.
- Trevithick S, Flabouris A, Tall G *et al.* International EMS systems: New South Wales, Australia. *Resuscitation* 2003;**59**:165–70. doi:10.1016/j.resuscitation.2003.11.004.
- Pozner CN, Zane R, Nelson SJ *et al.* International EMS systems: the United States: past, present, and future. *Resuscitation* 2004;**60**:239–44. doi:10.1016/j.resuscitation.2003.11.004.
- Garner AA. The role of physician staffing of helicopter emergency medical services in prehospital trauma response. *Emerg Med Australas* 2004;**16**: 318–23. doi:10.1111/j.1742-6723.2004.00636.x.
- Porter ME. What is value in health care? *N Engl J Med* 2010;**363**: 2477–81.
- Institute of Medicine. *Emergency Medical Services at a Crossroads*. Washington DC: The National Academies Press, 2006.
- World Health Organization. Quality of care: a process for making strategic choices in health systems 2006.
- National Committee for Quality Assurance. The essential guide to health care quality, [https://www.ncqa.org/Portals/0/Publications/Resource%20Library/NCQA\\_Primer\\_web.pdf](https://www.ncqa.org/Portals/0/Publications/Resource%20Library/NCQA_Primer_web.pdf). (1 March 2016, date last accessed).
- Scott KW, Jha AK. Putting quality on the global health agenda. *N Engl J Med* 2014;**371**:3–5. doi:10.1056/NEJMp1402157.
- Hsia RY, Niedzwiecki M. Avoidable emergency department visits: a starting point. *Int J Qual Health Care* 2017;**29**:642–5. doi:10.1093/intqhc/mzx081.
- Male L, Noble A, Atkinson J *et al.* Measuring patient experience: a systematic review to evaluate psychometric properties of patient reported experience measures (PREMs) for emergency care service provision. *Int J Qual Health Care* 2017;**29**:314–26. doi:10.1093/intqhc/mzx027.
- Stenlund M, Sjö Dahl R, Pia Yngman-Uhlin RN. Incidence and potential risk factors for hospital-acquired pneumonia in an emergency department of surgery. *Int J Qual Health Care* 2017;**29**:290–4. doi:10.1093/intqhc/mzx018.
- Rehn M, Kruger AJ. Quality improvement in pre-hospital critical care: increased value through research and publication. *Scand J Trauma Resusc Emerg Med* 2014;**22**:34.
- Snooks H, Evans A, Wells B *et al.* What are the highest priorities for research in emergency prehospital care? *Emerg Med J* 2009;**26**:549–50. doi:10.1136/emj.2008.065862.
- Fevang E, Lockey D, Thompson J *et al.* The top five research priorities in physician-provided pre-hospital critical care: a consensus report from a European research collaboration. *Scand J Trauma Resusc Emerg Med* 2011;**19**:57. doi:10.1186/1757-7241-19-57.
- Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003;**15**:523–30.
- Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988; **260**:1743–8.
- Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;**44**:166–206.
- FAMILIESUSA. Measuring Health Care Quality: An Overview of Quality Measures. [www.familiesusa.org](http://www.familiesusa.org). (15 April 2015, date last accessed).
- Grol R. Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. *JAMA* 2001; **286**:2578–85.
- Blumenthal D. Part 1: Quality of care—what is it? *N Engl J Med* 1996; **335**:891–4. doi:10.1056/NEJM199609193351213.
- El Sayed MJ. Measuring quality in emergency medical services: a review of clinical performance indicators. *Emerg Med Int* 2012;**2012**:161630. doi:10.1155/2012/161630.
- Kruger AJ, Skogvoll E, Castrén M *et al.* ScanDoc Phase 1a Study Group. Scandinavian pre-hospital physician-manned Emergency Medical Services—same concept across borders? *Resuscitation* 2010;**81**:427–33. doi:10.1016/j.resuscitation.2009.12.019.
- Moher D, Liberati A, Tetzlaff J *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6**:e1000097. doi:10.1371/journal.pmed.1000097.
- G K. ADACOPTER-Aufzeichnungen einer Entwicklung München. Werner Wolfsehlner Verlag; 2002.
- Kruger AJ, Lockey D, Kurola J *et al.* A consensus-based template for documenting and reporting in physician-staffed pre-hospital services. *Scand J Trauma Resusc Emerg Med* 2011;**19**:71. doi:10.1186/1757-7241-19-71.
- Akin Paker S, Dagar S, Gunay E *et al.* Assessment of prehospital medical care for the patients transported to emergency department by ambulance. *Turk J Emerg Med* 2015;**15**:122–5. doi:10.1016/j.tjem.2015.11.005.
- Anadelic S, Ivancevic N, Bogunovic S. Reaction times as indicators of the quality of expert work of belgrade municipal institutions for emergency medical services. *Cent Eur J Med* 2013;**8**:90–5. doi:10.2478/s11536-012-0102-0.
- Arntz HR, Klatt S, Stern R *et al.* Are emergency physicians' prehospital diagnoses reliable? [German]. *Anaesthesist* 1996;**45**:163–70. doi:10.1007/s001010050251.
- Breckwoldt J, Müller D, Overbeck M *et al.* Prehospital care of acute coronary syndrome by anaesthetists. Prospective comparison with the care standards of cardiologists. [German]. *Anaesthesist* 2008;**57**:131–8. doi:10.1007/s00101-007-1290-x.
- Duchateau FX, Garnier-Connois D, Ricard-Hibon A *et al.* Response time evaluation for emergency medical service as a part of its performance. *Emerg Med J* 2013;**30**:763–5. doi:10.1136/emered-2012-201714.
- Fjældstad A, Kirk MH, Knudsen L *et al.* Physician-staffed emergency helicopter reduces transportation time from alarm call to highly specialized centre. *Dan Med J* 2013;**60**:A4666.
- Flabouris A. A description of events associated with scene response by helicopter based medical retrieval teams. *Injury* 2003;**34**:847–52. doi:10.1016/S0020-1383(2803)2900033-0.
- Helm M, Faul M, Unger T *et al.* Reliability of emergency medical field triage. Exemplified by traffic accident victims. [German]. *Anaesthesist* 2013; **62**:973–80. doi:10.1007/s00101-013-2255-x.
- Helm M, Hauke J, Sauermüller G *et al.* About the quality of prehospital emergency ventilation—A prospective study in trauma patients. [German]. *Unfallchirurg* 1999;**102**:347–53. doi:10.1007/s001130050416.
- Hennes HJ, Reinhardt T, Otto S *et al.* Prehospital care: the efficacy of pre-hospital intervention by emergency medicine physicians. [German]. *Anaesthesist* 1993;**42**:455–61.
- Klemenc-Ketiš Z, Tomazin I, Kersnik J. HEMS in Slovenia: one country, four models, different quality outcomes. *Air Med J* 2012;**31**:298–304. doi:10.1016/j.amj.2012.09.002.
- Leicht MJ, Melanson SW, Albright D. Air medical tracheal intubation: establishing a threshold for this QA indicator. *J Air Med Transp* 1991;**10**: 15–20.
- Lossius HM, Soreide E, Hotvedt R *et al.* Prehospital advanced life support provided by specially trained physicians: is there a benefit in terms of life years gained? *Acta Anaesthesiol Scand* 2002;**46**:771–8.
- Messelken M, Kehrberger E, Dirks B *et al.* The quality of emergency medical care in Baden-Württemberg (Germany): four years in focus. [German],



- English]. *Dtsch Arztebl* 2010;107:523–30. doi:10.3238/arztebl.2010.0523.
44. Mikkelsen S, Kruger AJ, Zwisler ST *et al*. Outcome following physician supervised prehospital resuscitation: a retrospective study. *BMJ Open* 2015;5:e006167. doi:10.1136/bmjopen-2014-006167.
  45. Nakstad AR, Heimdal HJ, Strand T *et al*. Incidence of desaturation during prehospital rapid sequence intubation in a physician-based helicopter emergency service. *Am J Emerg Med* 2011;29:639–44. doi:10.1016/j.ajem.2010.01.013.
  46. Neukamm J, Grasner JT, Schewe JC *et al*. The impact of response time reliability on CPR incidence and resuscitation success: a benchmark study from the German Resuscitation Registry. *Crit Care* 2011;15:R282. doi:10.1186/cc10566.
  47. Pedersen UG, Jorgensen S, Nielsen SL. A survey of patients' satisfaction with the Mobile Emergency Care Unit in Copenhagen. [Danish]. *Ugeskr Laeger* 2006;168:3322–5.
  48. Regel G, Seekamp A, Pohlemann T *et al*. Does the accident patient need to be protected from the emergency doctor?. [German]. *Unfallchirurg* 1998;101:160–75. doi:10.1007/s001130050250.
  49. Rognas L, Hansen TM, Kirkegaard H *et al*. Pre-hospital advanced airway management by experienced anaesthesiologists: a prospective descriptive study. *Scand J Trauma Resusc Emerg Med* 2013;21:58. doi:10.1186/1757-7241-21-58.
  50. Rognas L, Hansen TM, Kirkegaard H *et al*. Anaesthesiologist-provided prehospital airway management in patients with traumatic brain injury: an observational study. *Eur J Emerg Med* 2014;21:418–23. doi:10.1097/MEJ.000000000000103.
  51. Sollid SJ, Lossius HM, Soreide E. Pre-hospital intubation by anaesthesiologists in patients with severe trauma: an audit of a Norwegian helicopter emergency medical service. *Scand J Trauma Resusc Emerg Med* 2010;18:30. doi:10.1186/1757-7241-18-30.
  52. van der Velden MW, Ringburg AN, Bergs EA *et al*. Prehospital interventions: time wasted or time saved? An observational cohort study of management in initial trauma care. *Emerg Med J* 2008;25:444–9. doi:10.1136/emj.2007.052662.
  53. Viergutz T, Schmittner MD, Bess S *et al*. Quality management of patients with severe traumatic brain injury in emergency medicine. [German]. *Notfall und Rettungsmedizin* 2015;18:381–90. doi:10.1007/s10049-015-0033-9.
  54. Von Knobelsdorff G, Reiferscheid F, Straakholder TM *et al*. Do paramedics and emergency physicians adhere to the algorithm of the European Resuscitation Council (ERC) for advanced cardiovascular life support (ACLS)? [German]. *Intensivmedizin und Notfallmedizin* 2004;41:22–8. doi:10.1007/s00390-004-0447-y.
  55. Weltermann B, Vom Eyser D, Kleine-Zander R *et al*. Deployment of emergency physicians for patients with stroke in the Munster area: a cross-sectional study of the quality of medical care in one region. [German]. *Dtsch Med Wochenschr* 1999;124:1192–6.
  56. Schlechtriemen T, Lackner CK, Moecke H *et al*. Medizinisches Qualitätsmanagement mit Hilfe ausgewählter Zieldiagnosen. [German]. *Notfall Rettungsmedizin* 2003;6:175–88. doi:10.1007/s10049-003-0561-6.
  57. Sobo EJ, Andriese S, Stroup C *et al*. Developing indicators for emergency medical services (EMS) system evaluation and quality improvement: a statewide demonstration and planning project. *Jt Comm J Qual Improv* 2001;27:138–54.
  58. Spaite DW, Criss EA, Valenzuela TD *et al*. Emergency medical service systems research: problems of the past, challenges of the future. *Ann Emerg Med* 1995;26:146–52.
  59. Spaite DW, Maio R, Garrison HG *et al*. Emergency Medical Services Outcomes Project (EMSOP) II: developing the foundation and conceptual models for out-of-hospital outcomes research. *Ann Emerg Med* 2001;37:657–63.
  60. Maio RF, Garrison HG, Spaite DW *et al*. Emergency medical services outcomes project I (EMSOP I): prioritizing conditions for outcomes research. *Ann Emerg Med* 1999;33:423–32.
  61. Busse R, Nimptsch U, Mansky T. Measuring, monitoring, and managing quality in Germany's hospitals. *Health Aff* 2009;28:294–304. doi:10.1377/hlthaff.28.2.w294.
  62. DIVI. Gemeinsame Empfehlung der Deutschen Interdisziplinären Vereinigung für Intensiv- und Notfallmedizin (DIVI) und des Vorstandes des Bundesärztekammer zum Qualitätsmanagement in der Notfallmedizin. [German]. *Intensivmed* 1995;32:387–8.
  63. Moecke H, Schafer J, Herden H *et al*. Das Bundes einheitliche Rettungsdienstprotokoll-Empfehlung der DIVI. [German]. *Intensivmed* 1994;31:96–9.
  64. Herden H, Moecke H. Bundes einheitliches Notarzteinsetzprotokoll. [German]. *Anasth Intensivmed* 1991;33:166–9.
  65. Palmer RH. Using health outcomes data to compare plans, networks and providers. *Int J Qual Health Care* 1998;10:477–83.
  66. El Sayed MJ. Measuring quality in emergency medical services: a review of clinical performance indicators. *Emerg Med Int* 2012;2012:161630.
  67. Nilsen JE, Tjelmeland IBM, Halvorsen KS *et al*. Kvalitetsindikatorer i den akuttmedisinske kjeden. [Norwegian]. www.nakos.no: Nasjonal kompetansetjeneste for prehospital akuttmedisin 2015.
  68. Department of Health. *Emergency care ten years on: reforming emergency care*. London: Department of Health, 2007.
  69. Cooke M, Fischer J, McLeod E *et al*. Reducing attendances and waits in emergency departments: a systematic review of present innovations NHS Service and Delivery Organisation 2005.
  70. Ebben RH, Vloet LC, Verhofstad MH *et al*. Mintjes-de Groot JA, van Achterberg T. Adherence to guidelines and protocols in the prehospital and emergency care setting: a systematic review. *Scand J Trauma Resusc Emerg Med* 2013;21:9. doi:10.1186/1757-7241-21-9.
  71. Charpentier S, Sagnes-Raffy C, Cournot M *et al*. Determinants and prognostic impact of compliance with guidelines in reperfusion therapy for ST-segment elevation myocardial infarction: results from the ESTIM Midi-Pyrenees Area. *Arch Cardiovasc Dis* 2009;102:387–96. doi:10.1016/j.acvd.2009.02.011.
  72. Aylwin CJ, König TC, Brennan NW *et al*. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. *Lancet* 2006;368:2219–25. doi:10.1016/S0140-6736(06)69896-6.
  73. Kirves H, Skrifvars MB, Vahakuopus M *et al*. Adherence to resuscitation guidelines during prehospital care of cardiac arrest patients. *Eur J Emerg Med* 2007;14:75–81. doi:10.1097/MEJ.0b013e328013f88c.
  74. Wollersheim H, Burgers J, Grol R. Clinical guidelines to improve patient care. *Neth J Med* 2005;63:188–92.
  75. Forsner T, Wistedt AA, Brommels M *et al*. An approach to measure compliance to clinical guidelines in psychiatric care. *BMC Psychiatry* 2008;8:64. doi:10.1186/1471-244x-8-64.
  76. Haugland H, Rehn M, Klepstad P *et al*, EQUIPE collaboration group. Developing quality indicators for physician-staffed emergency medical services: a consensus process. *Scand J Trauma Resusc Emerg Med* 2017;25:14. doi:10.1186/s13049-017-0362-4.
  77. Kessner DM, Kalk CE, Singer J. Assessing health quality—the case for tracers. *N Engl J Med* 1973;288:189–94. doi:10.1056/NEJM197301252880406.
  78. Martin LA, Nelson EC, Lloyd RC, Nolan TW. *Whole system measures*. Cambridge, Massachusetts: Institute for Healthcare Improvement, 2007.
  79. Institute of Medicine. *Crossing the quality chasm: a new health system for the twenty-first century*. Washington: National Academies Press, 2001.
  80. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QQuality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* 2016;25:986–92. doi:10.1136/bmjqs-2015-004411.





## Paper II



ORIGINAL RESEARCH

Open Access



# Developing quality indicators for physician-staffed emergency medical services: a consensus process

Helge Haugland<sup>1,2,3\*</sup> , Marius Rehn<sup>1,4,5</sup>, Pål Klepstad<sup>3,6</sup>, Andreas Krüger<sup>1,2,3</sup> and The EQUIPE-collaboration group

## Abstract

**Background:** There is increasing interest for quality measurement in health care services; pre-hospital emergency medical services (EMS) included. However, attempts of measuring the quality of physician-staffed EMS (P-EMS) are scarce. The aim of this study was to develop a set of quality indicators for international P-EMS to allow quality improvement initiatives.

**Methods:** A four-step modified nominal group technique process (expert panel method) was used.

**Results:** The expert panel reached consensus on 26 quality indicators for P-EMS. Fifteen quality indicators measure quality of P-EMS responses (response-specific quality indicators), whereas eleven quality indicators measure quality of P-EMS system structures (system-specific quality indicators).

**Discussion:** When measuring quality, the six quality dimensions defined by The Institute of Medicine should be appraised. We argue that this multidimensional approach to quality measurement seems particularly reasonable for services with a highly heterogenic patient population and complex operational contexts, like P-EMS. The quality indicators in this study were developed to represent a broad and comprehensive approach to quality measurement of P-EMS.

**Conclusions:** The expert panel successfully developed a set of quality indicators for international P-EMS. The quality indicators should be prospectively tested for feasibility, validity and reliability in clinical datasets. The quality indicators should then allow for adjusted quality measurement across different P-EMS systems.

**Keywords:** Quality indicators, Physician-staffed emergency medical services, Modified nominal group technique

## Background

The European Resuscitation Council has identified five critical conditions that require immediate pre-hospital management; cardiac arrest, severe respiratory failure, severe trauma, chest pain and stroke. Four of these conditions are among the leading causes of death in the European Union [1]. An observational study on Scandinavian physician-staffed emergency medical services (P-EMS) observed a pre-hospital incidence of severe illness or injury of 25–30 per 10 000 person-years [2]. Many of these conditions benefit from interventions that

rapidly correct deranged physiology and improve tissue oxygen delivery [3]. Services delivering pre-hospital critical care remain a critical link in the chain of survival for several frequent and life-threatening conditions.

Pre-hospital emergency care is primarily delivered by paramedics or nurses in ambulance EMS. In addition many health care systems employ P-EMS to respond to selected patients [4–6]. These P-EMS normally use rapid response cars or helicopters depending on distance to the scene and receiving hospital, weather, and the characteristics of each assignment [7]. However, although P-EMS is widely established in many countries little is known about the quality delivered by P-EMS.

\* Correspondence: helge.haugland@norskluftambulans.no

<sup>1</sup>Department of Research and Development, Norwegian Air Ambulance Foundation, Drøbak, Norway

<sup>2</sup>Department of Emergency Medicine and Pre-Hospital Services, St. Olavs Hospital, Trondheim, Norway

Full list of author information is available at the end of the article



© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

The importance of quality measurement in health care is widely recognized [8–11]. Moreover, defining quality indicators (QI) for P-EMS and EMS is identified as a high priority research area [12, 13]. QIs are instrumental to aid clinicians, organizations, health care managers and societies to achieve improvements in health care quality [14]. Further, QIs should integrate the best research evidence with clinical expertise and patient values [15] and allow measurement of health care quality by creating a quantitative basis that indicates performance.

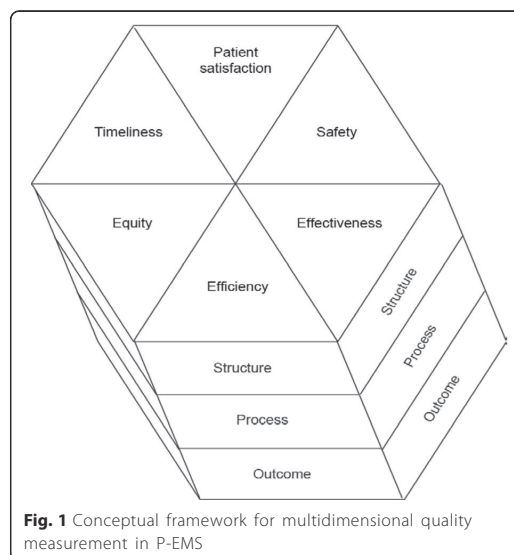
The literature on QIs in pre-hospital critical care is scarce [13, 16] and there is no international agreement on conceptual framework or choice of QIs for P-EMS. Appropriate QIs are needed to identify both high-quality care as well as areas where there is room for improvement in care. The current study describes the development of a comprehensive set of QIs for P-EMS and is a necessary initial step towards quality measurement in this field of health care.

## Methods

### Conceptual framework

For the purpose of this study, we used the framework described by Donabedian, which groups QIs in three broad categories; structure, process or outcome of health care [17, 18]. Structure indicators describe the infrastructure of a health care system, such as competence of the staff, available equipment, deployment and response times. Process indicators evaluate the care provided to the patient, whereas outcome indicators address the change in the patient's health status as a result of the provided care. Each type of QI will not give a complete description of the quality of care, but rather addresses a component of the care. Thus, different types of QIs should be combined when estimating the quality of a service [14].

To identify potential QIs, a widely used method is a combination of a systematic review of current literature and a formal process to obtain expert opinions. In this study, we tasked an expert panel to develop QIs for P-EMS using the modified nominal group technique [19, 20]. We defined P-EMS as a dedicated unit staffed with physicians trained in emergency care exceeding the competency of a general practitioner on call [21]. The QIs should be feasible to collect during the pre-hospital time interval or in the emergency department at hand-over. Further, the QIs should as far as possible cover the six quality dimensions that define high-quality care, stated by the Institute of Medicine [22], and appreciated by the World Health Organization [10]. The six quality dimensions are timeliness, safety, efficiency, equity, effectiveness, and patient-centeredness. An overview of the conceptual framework for this study; using structure-, process- and outcome-indicators to address six established quality dimensions, is depicted in Fig. 1.



**Fig. 1** Conceptual framework for multidimensional quality measurement in P-EMS

### The experts

When developing QIs the expert panel should consist of people considered experts in the appropriate area and who have credibility in the target audience [19]. Clinical expertise is represented by physicians, scientific expertise by researchers and user-expertise by patients. Accordingly, this study's expert panel consisted of clinicians and researchers from different P-EMSs, but also of stakeholders representing other perspectives in P-EMS. More specifically the 18 members of the expert panel consisted of, three general practitioners, two P-EMS medical directors, a director of a public health institute, a specialist in community medicine, a patient-organization leader and ten physicians working in P-EMS. All panel members were in different ways considered experts in P-EMS or in collaborating services of P-EMS, and practiced in Australia, Austria, Denmark, Finland, Norway, Scotland, United Kingdom and the United States of America. The experts were recruited through PubMed and Google Scholar searches, and via the professional network of the study group. 26 experts were invited by e-mail or telephone. 18 accepted the invitation, two declined and two did not respond. Non-responders were reminded three times by e-mail and three times by telephone.

### The modified nominal group process

In our study, the expert panel developed the QIs through a four-step modified nominal group technique. Stage 1, 2 and 4 were e-mail correspondences. In stage 3, the expert panel gathered for a 2-day consensus meeting in Oslo, Norway.

**Stage 1.** The members of the expert panel were asked to propose QIs for P-EMS according to the following predefined instructions: A total of 3–10 QIs should be proposed for each of the three categories of QIs; structure, process and outcome. A fourth category (Other indicators) was available for proposing QIs difficult to fit into the Structure-Process-Outcome - system. All proposed QIs should be possible to obtain during the pre-hospital time interval. The experts were asked to consider both evidence base and feasibility of data-collection when proposing QIs. However, it was not required that the proposed QIs could be extracted from existing databases in P-EMS.

The panel members returned their proposals to the project group pr. e-mail in a predesigned Excel spreadsheet (Microsoft Office 2013, Microsoft Inc., USA). QIs with identical meaning were merged. No proposed QIs were deleted. Further, the QIs within each category (structure, process and outcome) were ranked according to the number of experts who had included each QI in their proposal.

**Stage 2.** The experts were asked to use the revised spreadsheet to rank the ten most important QIs in each of the three categories (structure, process and outcome). In each category, the quality indicator ranked in first place was given a point value of 10, second of 9, third of eight and so forth, until the tenth place that was given one point. The point values from all panel members were added, and quality indicators with no ranking were removed from the list. The list with the remaining quality indicators, prioritized according to achieved point value, formed the basis for the consensus meeting.

**Stage 3.** The expert panel gathered for a 2-day consensus meeting in Oslo, Norway. A moderator (MR) led the experts through discussions on the QIs in the spreadsheet developed in stage 2. The experts decided which QIs should be included in the final set. Further, preliminary definitions and limitations were defined. All debates and discussions were plenary.

**Stage 4.** Based on the results from the consensus meeting, the project group prepared a document with the selected QIs, including definitions. This document was submitted to the panel members for comments. At this stage, no additional QIs were accepted. However, minor changes pertaining to definitions were allowed.

Consensus was defined as agreement on the proposed QIs during the meeting among the attending experts.

## Results

The 18 experts proposed and ranked QIs in stage 1 and 2 (one expert did not submit rankings in stage 2). In stage 1, 358 QIs were proposed by the expert panel. After merging, 179 QIs entered stage 2. At stage 2, 45 QIs obtained 0 points and were excluded, leaving 134 indicators to be discussed at the consensus meeting.

Thirteen experts attended the consensus meeting. During the consensus meeting the expert panel recommended the QIs from stage 2 to be classified into two different categories for clarity; response-specific QIs and system-specific QIs. The former is data from the pre-hospital time interval, measuring quality on the response level, and should be feasible to collect from any P-EMS response by the P-EMS physician. The latter should be administrative data describing fixed system characteristics, and should be registered once a year for services using the set of quality indicators continuously or for study purposes. The expert panel argued that the combined information from response- and system-specific QIs allows for a more thorough quality measurement than exclusively relying on response-specific QIs.

Consensus was reached on 15 response- and 11 system-specific QIs (Tables 1 and 2). More specific definitions for each QI are given in the explanation and elaboration document (Additional file 1). The expert panel allowed the project group to finalize the definitions of the indicators and propose them to all 18 experts in stage 4, where the final result was agreed upon. The QIs were allocated into one of the six quality dimensions as defined by the Institute of Medicine. All six quality dimensions were covered by the QIs, and both structure-, process- and outcome-indicators were represented. An overview of the distribution of the QIs is presented in Table 3.

## Discussion

This paper presents a set of potential QIs for quality measurement of P-EMS. Using a modified nominal group technique an international expert panel achieved consensus on these QIs that describe six quality dimensions and include structure-, process- and outcome-indicators.

Quality measurement of pre-hospital services has been identified as a high-priority research area and pivotal to achieve improvement in care [12, 13, 23, 24]. However, identifying valid quality indicators that are feasible to collect in the operational context of pre-hospital services has been a challenge [25]. We deliberately asked the experts to propose quality indicators themselves, not simply selecting from a pre-defined list. The rationale behind this was to make this process as open as possible in order to achieve a broad selection of proposals. The multidisciplinary composition of the expert panel was partly to facilitate this broad approach.

A premise for this study is that the principles for quality measurement in health care also applies to P-EMS. P-EMS is the practice of medicine outside hospitals, and we find it reasonable to accept this premise. The six core characteristics of quality depicted in Fig. 1 were defined by Institute of Medicine, naming them dimensions of quality [22]. Each of these is distinct and no

**Table 1** Response-specific quality indicators for physician-staffed emergency medical services

#	Quality indicator	Type of quality indicator	Quality dimension
1	Was the P-EMS unit <i>able</i> to respond immediately to the actual response?	Structure	Timeliness
2	What is the time interval from the dispatch center receives the alarm call until P-EMS unit arrives at the patient?	Structure	Timeliness
3	What is the time interval from P-EMS unit arrives at the patient until transportation of patient is initiated?	Process	Timeliness
4	What is the time interval from the P-EMS unit received the alarm call until the patient was delivered at the preferred destination?	Process	Timeliness
5	Did the patient arrive hospital alive?	Outcome	Timeliness
6	Was the P-EMS response debriefed?	Process	Safety
7	Did you experience any adverse events during the P-EMS response?	Process	Safety
8	Are all defined key variables measured and documented in the patient chart?	Process	Efficiency
9	Did the service have a guideline for the medical problem encountered in the response?	Structure	Equity
10	Was a physician and/or a paramedic from P-EMS involved in deciding if the P-EMS unit should be dispatched to the particular job or not?	Process	Equity
11	Without the assistance of the P-EMS unit: Do you consider that the level of competence on scene was sufficient to give the patient appropriate care?	Process	Equity
12	Did P-EMS provide advanced treatment in the actual response?	Process	Effectiveness
13	Did the logistical contribution by P-EMS give the patient a significant better service than the existing alternative?	Process	Effectiveness
14	Was the patient enrolled in a scientific study involving the pre-hospital care?	Structure	Effectiveness
15	Did you ensure that the relatives' needs were addressed; either by P-EMS or by collaborating services?	Process	Patient-centeredness

**Table 2** System-specific quality indicators for physician-staffed emergency medical services

#	Quality indicator	Type of quality indicator	Quality dimension
16	Is the dispatch center staffed 24/7 by specially trained pre-hospital physician?	Structure	Effectiveness
17	What is the number of P-EMS units per 100 000 inhabitants in the service area?	Structure	Equity
18	What is the number of P-EMS units per km <sup>2</sup> in the area covered by the service?	Structure	Equity
19	Does the service regularly perform interfacility transports coordinated by a dispatch centre?	Structure	Effectiveness
20	What level of regular in-hospital service do the P-EMS doctors practice in addition to their pre-hospital work?	Structure	Effectiveness
21	Proportion of P-EMS doctors with achieved speciality in: 1; anesthesiology 2; emergency medicine 3; other specialities.	Structure	Effectiveness
22	Proportion of P-EMS doctors who have attended and passed formalized training in major incident management.	Structure	Efficiency
23	Proportion of P-EMS doctors' assistants with the following qualification: Paramedic or nurse with supplemental regular training in assisting during induction of general anesthesia and/or formal education in anesthesia or intensive care.	Structure	Safety
24	Does the P-EMS service collect data pertaining to patient satisfaction?	Structure	Patient- centeredness
25	What is the number of documented complaints from patients, relatives or receiving hospitals per total number of P-EMS events (ratio)?	Outcome	Patient- centeredness
26	Does it exist a system for registration and reviewing of adverse events, critical incidents and educational events in the service?	Structure	Safety

**Table 3** Classification of quality indicators from the consensus process

	Timeliness	Safety	Efficiency	Equity	Effectiveness	Patient-centeredness	Number	Percent
Structure	0	2	1	2	6	1	12	46,2
Process	4	1	1	2	2	1	11	42,3
Outcome	1	1	0	0	0	1	3	11,5
n	5	4	2	4	8	3	26	
%	19,2	15,4	7,7	15,4	30,8	11,5		

one is defined more important than the others. When measuring quality, all six quality dimensions should be appraised. We argue that this multidimensional approach to quality measurement seems particularly reasonable for services with a highly heterogeneous patient population, like P-EMS. Patients cared for by P-EMS differ a lot: Neonates vs. elderly patients, medical vs. surgical diagnoses, patients rescued from open water vs. Intensive Care Unit transfers [2, 5]. What is considered high quality care for each patient will be context-specific. With this complexity of cases, treatments and operational contexts, we argue that adequate quality measurement of P-EMS should be multidimensional.

#### Quality dimensions

Timely care is about reducing needless and potentially harmful delays before the patient receives specialized care from the P-EMS. Traditionally, attempts on quality measurement of pre-hospital services, have been limited to data on time-variables corresponding the quality dimension “*timeliness*” [13, 26]. Studies of EMS have shown that response time affects outcome only for a small group of patients [27, 28]. Moreover, time variables describe the logistics, but not the provided care. Response time of P-EMS is measured in QI 2 “Time to arrival of P-EMS” and is indeed important for some time critical conditions such as cardiac arrest and major trauma [29]. However, the importance of short response times cannot be generalized to all emergency responses [30]. In selected situations, too much emphasis on timeliness is misleading in respect of what really represents quality for the patient. In the United Kingdom paramedics criticized the use of a time target structure measure (eight-minute response time for 75% of category A or emergency calls) as the main performance indicator in EMS. They argued this QI was “too simplistic and narrow” and that it could also increase risk for patients and ambulance crews [31]. An example may illustrate the limitation of time-variables as the sole QIs in P-EMS: Performing an ultra-sound scan of the traumatized patient may prolong the time on scene slightly. However, the examination can result in changes in treatment or triage [32], hence making the extra time spent on scene well worth.

The quality dimension “*safety*” focus on safety issues related to P-EMS responses for patient, EMS-staff or

others. The safety issues can be medical, technical or operational. P-EMS operates rapid response cars and helicopters, all activities associated with operative risks for patients, bystanders and crew [33]. Additionally, P-EMS care for severely injured or ill patients without access to safety initiatives as seen in hospitals e.g. senior assistance or access to patient history. Moreover, the pre-hospital environment can be associated with hazards like extreme temperatures, traffic and difficult access requiring application of rescue techniques [34].

The quality dimension “*efficiency*” is about avoiding medical waste; including waste of use of P-EMS personal, equipment and energy. Advanced major incident management reduce over-triage and is an example of how to prevent waste of resources [35]. This issue is covered in QI 22, which measures the proportion of P-EMS doctors who have completed a major incident management program.

“*Equity*” is about ensuring that quality of care is provided equally regardless of the patient's gender, ethnicity, geographic location and socioeconomic status. P-EMS contributes to equitable care by reducing transportation times (when using a helicopter) and by bringing the hospital competencies to the pre-hospital environment. This role of P-EMS can also be defined a governmental objective [36] as an initiative to give people living in scattered spread populations specialized care within due time. Thus, a more equitable access to centralized medical treatments like neurosurgery and invasive cardiology can be provided. The expert panel argued that the involvement of a physician or a paramedic from P-EMS in the dispatch decision would secure the most correct use of P-EMS, thus contributing to equitable care. This is addressed in QI 9 “P-EMS involvement in dispatch decision”.

“*Effectiveness*” is about ensuring that provided treatment is evidence-based. Care proven effective should be provided, thereby preventing undertreatment. Similarly, care proven ineffective should not be provided, thereby preventing overtreatment. There is some evidence that the use of physicians in EMS for selected patient groups, improve outcome or proxy outcomes such as physiological variables [1, 37]. However, the current documentation on the impact of P-EMS initiatives is controversial and, therefore, effectiveness QIs are difficult to derive from the literature. The expert



panel combined existing evidence with the experience and considerations of all panel members. One of the resulting QIs, QI 12 “Advanced Treatment”, addresses care considered indicated, but not feasible without the competence of P-EMS. Please note that withholding unethical or unnecessary treatment by the P-EMS physician also was defined as “advanced treatment” by the expert panel. Thus, critical decision making as illustrated for pre-hospital advanced airway management by Rognås et al.[38], is recognized as a part of quality care.

“Patient-centeredness” is about ensuring that care is responsive to individual needs. Although most stakeholders and clinicians in P-EMS presumably put the patient in the center of the care, the study group wanted to secure that the patients were represented in the expert panel. Therefore, a leading representative from a major patient organization was invited to join the expert panel. Developing quality indicators for this quality dimension in P-EMS is challenging, primarily because many of the patients cared for by the service are unconscious or at least not capable of expressing their own needs in their usual manner. This can be due to the clinical condition itself, stressful situation or pharmacological interventions. The needs of the patient’s family, however, can be expressed more easily. Moreover, the term “patient-centeredness” has been argued expanded to “patient- and family-centeredness” [39]. Patient- and family-centered care is based on the beneficial partnership between patient, family and health care workers, and it can be applied to patients in all ages and in any health care setting [39, 40]. As a surrogate for measuring the patient’s needs, the needs of the patient’s relatives could be addressed, as defined in QI 15 “Care for relatives”. This QI addresses the relatives’ needs, including the need for practical and emotional assistance.

#### Types of quality indicators

J. Mainz has reviewed the strengths of structure-, process- and outcome-indicators [14]. Structure indicators are found most useful when they predict variations in processes or outcomes of care. Process indicators are particularly useful when coping with short time frames, low volume providers and when tools to adjust or stratify for patient related factors are difficult to apply. Comparison of process indicators are generally easier to interpret and more sensitive to small differences than comparison of outcomes data. Based on these characteristics, we consider process indicators particularly suitable for continuous quality measurement of P-EMS. Although necessary to get information about a patient’s final outcome, long-term outcome indicators appear less feasible for measuring the isolated quality of P-EMS. From a patient is admitted to hospital by P-EMS until a long-

term outcome is measured, the patient has received care from numerous units, each potentially influencing outcome [41]. Unless performing risk adjustment and outcome measurement for each of these care intervals, it will be problematic to use long term outcome measures as indicators of the isolated quality of P-EMS. Instead, quality indicators from the pre-hospital care interval should be developed for this purpose [23]. The Institute of Medicine has stated that “quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” [42]. This definition of quality is a reminder that good quality is not identical to good outcomes. Despite excellent health care is provided, outcome for patients can be poor. Opposite, patients receiving poor quality health care can have good outcome.

#### Strengths and limitations

Using the professional network of the study group for recruitment of panel members, may have limitations: Colleagues that share our own professional interests may have been easier to identify and invite, than those with other views and mindsets. This practice can possibly lead to an imbalance in the composition of the expert panel. Although the expert panel reflected the inter-disciplinary nature of EMS, we recognize that we did not include a representative from an Emergency Medical Communication Central (EMCC). There was a trade-off between a manageable number of experts and the need for an inter-disciplinary composition of the expert panel. Consensus methodology literature describe an optimal group size of eight to twelve members [43]. Our efforts in making the panel sufficiently inter-disciplinary resulted in a group size of 18 experts. However, due to a rigorous time schedule throughout the process, the slightly larger expert panel did not lead to any unnecessary delay.

Eight nations were represented in the expert panel; all from developed countries and the majority from Scandinavia. Therefore, we recognize that other areas may have other QIs which should be implemented locally. However, P-EMS as a service is usually only delivered in developed countries. Hence, for these services the nationalities included should be representative.

In the consensus process we used a system of ranking and scoring to identify the QIs supported by the most experts in the panel. There are different methods to prioritize proposals and obtain consensus, and no method is considered clearly superior to the others [44]. At the consensus meeting, any proposal was omitted if vigorously opposed by one or more of the participants.

The use of a Likert scale is another recognized method for defining the level of consensus. Likert scores are used for QI selection in several studies, including a recent Danish study selecting QIs for hospital-based emergency care [45, 46]. Whether the use of a Likert scale would have improved our consensus process remains unclear. Moreover, it is methodological important to prevent that verbally skilled panel members dominate the consensus process. This issue also relates to “strong” personalities or experts enjoying a higher reputation than the other panel members [19]. Therefore, proposals and rankings in stage 1 and 2 were anonymous.

The value of this study is the development of multidimensional quality indicators for P-EMS. This represents a starting point for future studies on measuring and improving quality of P-EMS. The necessary next step should be to test the feasibility and validity of the QIs in a sample of P-EMSs. Thus, a more final set of QIs for P-EMS can be developed.

## Conclusion

Using a modified nominal group technique, an international expert panel reached consensus on 15 response specific and 11 system specific quality indicators for P-EMS. All six quality dimensions stated by Institute of Medicine are covered, and the quality indicators represent structure, process and outcome indicators. This 26 quality indicators large set is developed to represent a broad and comprehensive approach to quality measurement in international P-EMS, allowing future quality measurement comparable across different P-EMS systems.

## Additional file

**Additional file 1:** “Definition catalogue from the EQUIPE-project”. This explanation and elaboration document contains the definitions of all quality indicators, as well as explanation of the response alternatives where necessary. (ZIP 134 kb)

## Abbreviations

EMCC: Emergency Medical Communication Central; EMS: Emergency medical service; P-EMS: Physician-staffed emergency medical service; QI: Quality indicator

## Acknowledgements

The project group would like to thank the following members of the “EQUIPE-collaboration group” for their valuable participation: Gry Elise Albrektzen (Norway), Peter Anthony Berlac (Denmark), Geir Sverre Braut (Norway), Per Bredmose (Norway), Robert Burman (Norway), Brian Burns (Australia), Alasdair Corfield (Scotland), Marta Ebbing (Norway), Magnus Hjortdahl (Norway), Freddy Lippert (Denmark), Pål Madsen (Norway), Per Oretorp (Norway), Leif Rognås (Denmark), Julian Thompson (UK), Oddvar Uleberg (Norway), Janne Virta (Finland), Wolfgang G Voelckel (Austria) and Ryan Wubben (USA).

The authors would like to thank all members and involved staff from NAAF for making this study possible.

## Funding

This project was financed by the Norwegian Air Ambulance Foundation (NAAF), and the first author (HH) holds a position as a PhD-student financed by NAAF.

## Availability of data and materials

The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

## Authors' contributions

HH was the main author, communicated with the members of the expert panel and collected the data. AK conceptualized and initiated the project. MR chaired the consensus meeting, and all authors were involved in the planning of the meeting and the writing process. All authors read and approved the final version of the manuscript.

## Competing interests

MR is Deputy Editor of *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*. He had no involvement in the editorial management of this manuscript.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Author details

<sup>1</sup>Department of Research and Development, Norwegian Air Ambulance Foundation, Drøbak, Norway. <sup>2</sup>Department of Emergency Medicine and Pre-Hospital Services, St. Olavs Hospital, Trondheim, Norway. <sup>3</sup>Department of Circulation and Medical Imaging, Medical Faculty, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. <sup>4</sup>Department of Health Studies, University of Stavanger, Stavanger, Norway. <sup>5</sup>Division of Emergencies and Critical Care, Department of Anaesthesia, Oslo University Hospital, Oslo, Norway. <sup>6</sup>Department of Anaesthesiology and Intensive Care, St. Olav University Hospital, Trondheim, Norway.

Received: 16 October 2016 Accepted: 10 February 2017

Published online: 15 February 2017

## References

- Fischer M, Kamp J, Garcia-Castrillo Riesgo L, Robertson-Steel I, Overton J, Ziemann A, et al. Comparing emergency medical service systems—a project of the European Emergency Data (EED) Project. *Resuscitation*. 2011;82(3):285–93.
- Kruger AJ, Lossius HM, Mikkelsen S, Kurola J, Castren M, Skogvoll E. Pre-hospital critical care by anaesthesiologist-staffed pre-hospital services in Scandinavia: a prospective population-based study. *Acta Anaesthesiol Scand*. 2013;57(9):1175–85. doi:10.1111/aas.12181.
- Ghosh R, Pepe P. The critical care cascade: a systems approach. *Curr Opin Crit Care*. 2009;15(4):279–83. doi:10.1097/MCC.0b013e32832faef2.
- Lockey D. International EMS, systems: Geographical lottery and diversity but many common challenges. *Resuscitation*. 2009;80(7):722. doi:10.1016/j.resuscitation.2009.04.006.
- Kruger AJ, Skogvoll E, Castren M, Kurola J, Lossius HM. ScanDoc Phase 1a Study G. Scandinavian pre-hospital physician-manned Emergency Medical Services—same concept across borders? *Resuscitation*. 2010;81(4):427–33. doi:10.1016/j.resuscitation.2009.12.019.
- Garner AA. The role of physician staffing of helicopter emergency medical services in prehospital trauma response. *Emerg Med Australas*. 2004;16(4):318–23. doi:10.1111/j.1742-6723.2004.00636.x.
- Rehn M, Davies G, Smith P, Lockey DJ. Structure of Rapid Response Car Operations in an Urban Trauma Service. *Air Med J*. 2016;35(3):143–7. doi:10.1016/j.amj.2015.12.015.
- Porter ME. What is value in health care? *N Engl J Med*. 2010;363(26):2477–81.
- Institute of Medicine. *Emergency Medical Services at a Crossroads*. Washington DC, USA: The National Academies Press; 2006.
- World Health Organization. *Quality of care: a process for making strategic choices in health systems*2006.

11. National Committee for Quality Assurance. The essential guide to health care quality. <http://www.ncqa.org/Newsroom/ResourceLibrary/EssentialGuidetoHealthCareQuality.aspx>. Accessed 15 Feb 2016.
12. Fevang E, Lockey D, Thompson J, Lossius HM, Torpo Research C. The top five research priorities in physician-provided pre-hospital critical care: a consensus report from a European research collaboration. *Scand J Trauma Resusc Emerg Med*. 2011;19:57. doi:10.1186/1757-7241-19-57.
13. Snooks H, Evans A, Wells B, Peconi J, Thomas M, Woollard M, et al. What are the highest priorities for research in emergency prehospital care? *Emerg Med J*. 2009;26(8):549–50. doi:10.1136/emj.2008.065862.
14. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care*. 2003;15(6):523–30.
15. Sackett D, Strauss S, Richardson W. Evidence-Based Medicine: how to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000.
16. Spaite DW, Criss EA, Valenzuela TD, Guisto J. Emergency medical service systems research: problems of the past, challenges of the future. *Ann Emerg Med*. 1995;26(2):146–52.
17. Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988;260(12):1743–8.
18. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966;44 Suppl 1:166–206.
19. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998;2(3):i–iv. 1–88.
20. Lossius HM, Kruger AJ, Ringdal KG, Sollid SJ, Lockey DJ. Developing templates for uniform data documentation and reporting in critical care using a modified nominal group technique. *Scand J Trauma Resusc Emerg Med*. 2013;21:80. doi:10.1186/1757-7241-21-80.
21. Kruger AJ, Lockey D, Kuroki J, Di Bartolomeo S, Castren M, Mikkelsen S, et al. A consensus-based template for documenting and reporting in physician-staffed pre-hospital services. *Scand J Trauma Resusc Emerg Med*. 2011;19:71. doi:10.1186/1757-7241-19-71.
22. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the Twenty-first Century. Washington: National Academies Press; 2001.
23. Rehn M, Kruger AJ. Quality improvement in pre-hospital critical care: increased value through research and publication. *Scand J Trauma Resusc Emerg Med*. 2014;22:34.
24. Jones SE, Brenneis AT. Study design in prehospital trauma advanced life support-basic life support research: a critical review. *Ann Emerg Med*. 1991;20(8):857–60.
25. Moore L. Measuring quality and effectiveness of prehospital EMS. *Prehosp Emerg Care*. 1999;3(4):325–31.
26. El Sayed MJ. Measuring quality in emergency medical services: a review of clinical performance indicators. *Emerg Med Int*. 2012;2012:161630.
27. Department of Health. Emergency care 10 years on: reforming emergency care. London: Department of Health; 2007.
28. Cooke M, Fischer J, McLeod E, et al. Reducing attendances and waits in emergency departments: a systematic review of present innovations. London: National Coordinating Centre for NHS Service Delivery and Organisation (NCCSDO); 2005. [http://www.netscc.ac.uk/hsdr/files/project/SDO\\_ES\\_08-1204-029\\_V01.pdf](http://www.netscc.ac.uk/hsdr/files/project/SDO_ES_08-1204-029_V01.pdf).
29. Eisenberg MS, Bergner L, Hallstrom A. Cardiac resuscitation in the community. Importance of rapid provision and implications for program planning. *JAMA*. 1979;241(18):1905–7.
30. Pons PT, Haukoos JS, Bludworth W, Cribley T, Pons KA, Markovchick VJ. Paramedic response time: does it affect patient survival? *Acad Emerg Med*. 2005;12(7):594–600. doi:10.1197/jaem.2005.02.013.
31. Price L. Treating the clock and not the patient: ambulance response times and risk. *Qual Saf Health Care*. 2006;15(2):127–30. doi:10.1136/qshc.2005.015651.
32. Walcher F, Weinlich M, Conrad G, Schweigkofler U, Breikreutz R, Kirschning T, et al. Prehospital ultrasound imaging improves management of abdominal trauma. *Br J Surg*. 2006;93(2):238–42. doi:10.1002/bjs.5213.
33. Chesters A, Grieve PH, Hodgetts TJ. A 26-year comparative review of United Kingdom helicopter emergency medical services crashes and serious incidents. *J Trauma Acute Care Surg*. 2014;76(4):1055–60. doi:10.1097/TA.0000000000000170.
34. Kruger AJ, Lippert F, Brattebo G. Pre-hospital care and hazardous environments. *Acta Anaesthesiol Scand*. 2012;56(2):135–7. doi:10.1111/j.1399-6576.2011.02619.x.
35. Aylwin CJ, König TC, Brennan NW, Shirley PJ, Davies G, Walsh MS, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. *Lancet*. 2006;368(9554):2219–25. doi:10.1016/S0140-6736(06)69896-6.
36. Stortingsmelding nr. 50. Samarbeid og styring. 1993:94.
37. Roudsari BS, Nathens AB, Cameron P, Civil I, Gruen RL, Koepsell TD, et al. International comparison of prehospital trauma care systems. *Injury*. 2007;38(9):993–1000.
38. Rognas L, Hansen TM, Kirkegaard H, Tonnesen E. Refraining from pre-hospital advanced airway management: a prospective observational study of critical decision making in an anaesthesiologist-staffed pre-hospital critical care service. *Scand J Trauma Resusc Emerg Med*. 2013;21:75. doi:10.1186/1757-7241-21-75.
39. Institute for Patient- and Family-centered Care. [www.ipfcc.org](http://www.ipfcc.org). Accessed 3 July 2016.
40. Committee On Hospital C, Institute For P, Family-Centered C. Patient- and family-centered care and the pediatrician's role. *Pediatrics*. 2012;129(2):394–404. doi:10.1542/peds.2011-3084.
41. Spaite DW, Maio R, Garrison HG, Desmond JS, Gregor MA, Stiell IG, et al. Emergency Medical Services Outcomes Project (EMSOP) II: developing the foundation and conceptual models for out-of-hospital outcomes research. *Ann Emerg Med*. 2001;37(6):657–63.
42. Institute of Medicine LK. Medicare: a strategy for quality assurance. Washington D.C.: National Academy Press; 1990.
43. Waggoner J, Carline JD, Durning SJ. Is There a Consensus on Consensus Methodology? Descriptions and Recommendations for Future Consensus Research. *Acad Med*. 2016;91(5):663–8. doi:10.1097/ACM.0000000000001092.
44. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health*. 1984;74(9):979–83.
45. Madsen MM, Eiseit AH, Mackenhauer J, Odby A, Christiansen CF, Kurland L, et al. Selection of quality indicators for hospital-based emergency care in Denmark, informed by a modified-Delphi process. *Scand J Trauma Resusc Emerg Med*. 2016;24(1):11. doi:10.1186/s13049-016-0203-x.
46. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS ONE*. 2011;6(6):e20476. doi:10.1371/journal.pone.0020476.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research



Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



## Paper III



# BMJ Open Testing quality indicators and proposing benchmarks for physician-staffed emergency medical services: a prospective Nordic multicentre study

Helge Haugland <sup>1,2</sup>, Anna Olkinuora,<sup>3</sup> Leif Rognås <sup>4,5</sup>, David Ohlen,<sup>6</sup> Andreas Krüger<sup>1,2</sup>

**To cite:** Haugland H, Olkinuora A, Rognås L, et al. Testing quality indicators and proposing benchmarks for physician-staffed emergency medical services: a prospective Nordic multicentre study. *BMJ Open* 2019;9:e030626. doi:10.1136/bmjopen-2019-030626

► 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-2019-030626>).

Received 25 March 2019  
Revised 18 September 2019  
Accepted 11 October 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Helge Haugland;  
[helge.haugland@norskluftambulans.no](mailto:helge.haugland@norskluftambulans.no)

## ABSTRACT

**Objectives** A consensus study from 2017 developed 15 response-specific quality indicators (QIs) for physician-staffed emergency medical services (P-EMS). The aim of this study was to test these QIs for important characteristics in a real clinical setting. These characteristics were feasibility, rankability, variability, actionability and documentation. We further aimed to propose benchmarks for future quality measurements in P-EMS.

**Design** In this prospective observational study, physician-staffed helicopter emergency services registered data for the 15 QIs. The feasibility of the QIs was assessed based on the comments of the recording physicians. The other four QI characteristics were assessed by the authors. Benchmarks were proposed based on the quartiles in the dataset.

**Setting** Nordic physician-staffed helicopter emergency medical services.

**Participants** 16 physician-staffed helicopter emergency services in Finland, Sweden, Denmark and Norway.

**Results** The dataset consists of 5638 requests to the participating P-EMSs. There were 2814 requests resulting in completed responses with patient contact. All QIs were feasible to obtain. The variability of 14 out of 15 QIs was adequate. Rankability was adequate for all QIs. Actionability was assessed as being adequate for 10 QIs. Documentation was adequate for 14 QIs. Benchmarks for all QIs were proposed.

**Conclusions** All 15 QIs seem possible to use in everyday quality measurement and improvement. However, it seems reasonable to not analyse the QI 'Adverse Events' with a strictly quantitative approach because of a low rate of adverse events. Rather, this QI should be used to identify adverse events so that they can be analysed as sentinel events. The actionability of the QIs 'Able to respond immediately when alarmed', 'Time to arrival of P-EMS', 'Time to preferred destination', 'Provision of advanced treatment' and 'Significant logistical contribution' was assessed as being poor. Benchmarks for the QIs and a total quality score are proposed for future quality measurements.

## INTRODUCTION

### Background/rationale

The importance of quality improvement in healthcare has been recognised by leading

## Strengths and limitations of this study

- This is the first study putting the EQUIPE (Establishing Quality Indicators in Physician-staffed Emergency Medical Services) quality indicators (QIs), developed specifically for physician-staffed emergency medical services, into a clinical setting.
- A prospective multicentre study involving 16 Nordic physician-staffed helicopter emergency medical services.
- The QIs are assessed for important QI characteristics.
- Benchmarks for future quality measurement are proposed.
- Except from the feasibility of the QIs, the assessment of the different QI characteristics was done by the author group.

health organisations and in landmark publications.<sup>1–4</sup> However, publications on quality measurement in physician-staffed emergency medical services (P-EMS) are rare.<sup>5</sup> For prehospital services in general, and P-EMS specifically, more research on quality measurement has been warranted.<sup>6,7</sup> Moreover, it has been argued that quality assurance and even quality improvement in P-EMS requires a model for quality estimation to achieve appropriate governance.<sup>8</sup> Quality measurements are an obvious prerequisite for quality improvement. A first initial step is the development of appropriate tools for quality measurement, that is, quality indicators (QIs). A QI can be defined as a measurable element of performance for which there is evidence or consensus that it can be used to assess the quality and hence change the quality of care provided.<sup>9</sup>

No comprehensive set of systematically developed QIs are registered in P-EMS in Sweden, Denmark, Finland and Norway. Attempts on extracting information concerning the quality of the service have

primarily been limited to time variables.<sup>10</sup> Response time has been widely used for quality assessment but may have been overemphasised and is not applicable for all prehospital emergency medical activity.<sup>11</sup> Time variables primarily describe the transport component of P-EMS. This information is necessary but not sufficient for quality assessment. The care component of P-EMS also has to be addressed. In fact, The Institute of Medicine, a US independent non-governmental research organisation, has defined six quality dimensions that should be addressed when measuring the overall quality of a health service<sup>12</sup>: patient centredness, safety, effectiveness, efficiency, equity and timeliness. If only one or a few of these quality dimensions are addressed, the result can be a simplistic and narrow quality measurement.

In 2018, we published a systematic literature review describing quality measurement studies in P-EMS.<sup>5</sup> There was no common understanding in the studies as to which QIs to use. Moreover, 15 out of the 27 identified studies used only one QI. This increases the risk of a one-sided approach in quality measurement. The review concludes that future quality measurement in P-EMS should be done based on a consensus-based set of QIs rather than a single QI to ensure a comprehensive quality measurement. In another recent study, we developed a set of multidimensional QIs for P-EMS through a consensus process. These QIs were called the EQUIPE (Establishing Quality Indicators in Physician-staffed Emergency Medical Services) QIs (online supplementary file 1). Panellists from different stakeholder groups agreed on 15 response-specific QIs for P-EMS.<sup>13</sup> These are QIs that should be feasible to collect from any P-EMS response during the prehospital time interval or in the emergency department at handover. Despite methodically correct development, QIs are not necessarily suitable in real datasets. The actual QIs have not yet been tested in clinical datasets. Based on modern framework for QI efforts, the next stage in the development of QIs for P-EMS should be testing for critical QI characteristics (feasibility, rankability, variability, actionability and documentation).

## Objectives

The aim of this study was to test the multidimensional QIs for the above-mentioned characteristics in a real clinical setting. We further aimed to propose benchmarks for future quality measurement in P-EMS based on the data in this study.

## METHODS

### Study design and setting

In this prospective observational study, 16 physician-staffed helicopter emergency services in Finland, Sweden, Denmark and Norway registered data for the EQUIPE quality indicators. There has previously been documented significant system similarities in the P-EMS of the four participating countries, making them a suitable arena for multicentre studies.<sup>14</sup> The Nordic countries have a mix

of urban of rural areas with a rather low overall population density (19.6 inhabitants/km<sup>2</sup>). The prehospital incidence of critical illness and injury in these countries has been documented to be 25–30/10 000 person-years.<sup>15</sup> The physicians staffing Nordic P-EMS are usually experienced anaesthesiologists, most of them working both in P-EMS and in hospitals.<sup>14 16</sup> All Nordic services do primary responses, and the Swedish, Danish and Norwegian services also do secondary responses; the former is defined as responses where the patient is located outside a hospital, and the latter is interhospital transfers. Moreover, the Norwegian services also do search and rescue responses (SAR responses). In addition, one Swedish (Karlstad) and all Finnish and Norwegian bases dispose a rapid response car for responses close to the base and for responses in poor weather conditions that prevent flight operations. The study applied Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.<sup>17</sup>

### Inclusion criteria and data variables

We included every request to the P-EMS to dispatch the P-EMS unit. Thus, we could include both completed and cancelled responses, as well as stand-downs (responses cancelled by dispatch or crews on-scene) and rejected responses. Examples of reasons for rejecting a response might be weather conditions or the lack of medical need as judged by the P-EMS physician. The latter is possible in Sweden, Finland and Norway where the acceptance or rejection of a response is at the P-EMS physicians' discretion. Inquiries with counselling as the only purpose were excluded. Primary and secondary responses as well as SAR responses were included. For bases with both a helicopter and a rapid response car, responses were included regardless of the mode of transportation. All 15 EQUIPE QIs were registered in responses involving patient contact.<sup>13</sup> Only 4 of the 15 QIs were registered in responses not involving patient contact (QIs 1, 6, 7, 10). Data were collected for 3 months (from 10 June to 12 September 2016).

### Data sources/measurement

Finland collected the necessary data by including the QIs as part of their existing documentation database (Finn-HEMS database, FHDB). FHDB is a national database, including both response and patient data where all HEMS units register all responses. Some QIs could be gathered from the existing data (eg, time stamps) and those that could not were implemented either as permanent variables or on a separate study sheet. It was mandatory to fill in all the QIs in the system. The other nations registered the same data by using a web-based questionnaire (Form-site; Vroman Systems, Chicago, Illinois, USA). In all nations, the data were collected after completed response by the P-EMS physician. The four national investigators monitored the documentation of participating P-EMS bases to secure accurate data collection.



The first 2 weeks of the data collection period (from 10 June to 24 June 2016) was a feasibility test; we wanted to study if the QIs from the consensus process were feasible to collect in the everyday of P-EMS. The feasibility test was done as a pilot study involving the same Finnish, Swedish and Danish bases that participated in the main study. However, only two Norwegian bases participated in this pilot study (Trondheim and Ørland). We considered this sample sufficient because feasibility tests can be run in a small scale.<sup>18</sup> Here, all the recording physicians could comment on the feasibility of obtaining the necessary data. An assessment of the feasibility of the QIs was done after these 2 weeks. This was done based on comments from the recording physicians. After these 2 weeks of feasibility testing, we adapted and clarified the wording of some QIs and then continued the data collection for a total of 3 months.

We assessed four other important characteristics of QIs in addition to feasibility: rankability, variability, actionability and documentation.<sup>19 20</sup> This was done according to the criteria for good QIs defined by the Organisation for Economic Cooperation and Development and the Agency for Healthcare Research and Quality.

Rankability is assessed by judging if a QI has a clear direction of good and bad, that is, the QI has a good rankability if high values for a QI are always better than low values. Conversely, rankability is poor if high values are better than low values but *very* high values are worse than low values.

According to criteria for QIs, a good QI must have enough variability to allow for improvement. To assess variability, we calculated the mean and median as well as the corresponding variance for each of the QIs based on the data collected after the feasibility test. This illustrates both the average performance and the variation in the participating Nordic P-EMSs. To the best of our knowledge, there is no definition of how much variability a QI should have to be useful. This implies that the assessment of variance is somewhat arbitrary.

Actionability is the possibility of influencing the QI performance. For instance, a P-EMS has limited opportunity to reduce the time to definitive care because this mainly depends on the distances that the P-EMS unit has to work with. In that case, actionability is rather low.

Furthermore, for a QI to be valid, the process or structure of defining the QI must have been documented to give better outcome. The degree of such documentation was assessed for each QI.

We do not report which results belong to the specific P-EMS bases simply because the aim of this study was to assess the characteristics of the QIs and not to compare the performance of the participating services.

### Missing data

Due to technical solutions, the QIs 'P-EMS involvement in dispatch' and 'Debriefed responses' were registered only in responses with patient contact in Finland; however, these QIs were registered for all responses in the other

three nations. The proportion of missing data for the QIs varied between 0.2% and 0.9%. Missing observations were acknowledged and omitted from the analysis. All analyses were done on variables present, thus minimising information loss.

### Statistical methods

Descriptive statistics are reported. The QI proportions were recorded for QIs that are categorical variables; time was recorded in minutes for QIs that were continuous time variables. All QIs are reported by the mean and the corresponding 95% CI as well as the median with corresponding IQR.

We also used figures from the 16 P-EMS bases to propose benchmarks for all QIs. We set the benchmark at the lower end of the fourth quartile for QIs where higher values reflect better performance. For QIs where lower values reflect better performance, we have set the benchmark at the highest end of the first quartile. We depicted the benchmarking graphically so that performances within the IQR are shown in yellow. Performances better than the IQR level are in green, and those worse than the IQR level are red.

### Ethics approval and consent to participate

According to the approvals from all four countries, the data were obtained without informed consent from patients or their next-of-kin. As stated in the study protocol, there was no deviation from regular clinical practice during the study period.

### Patient and public involvement

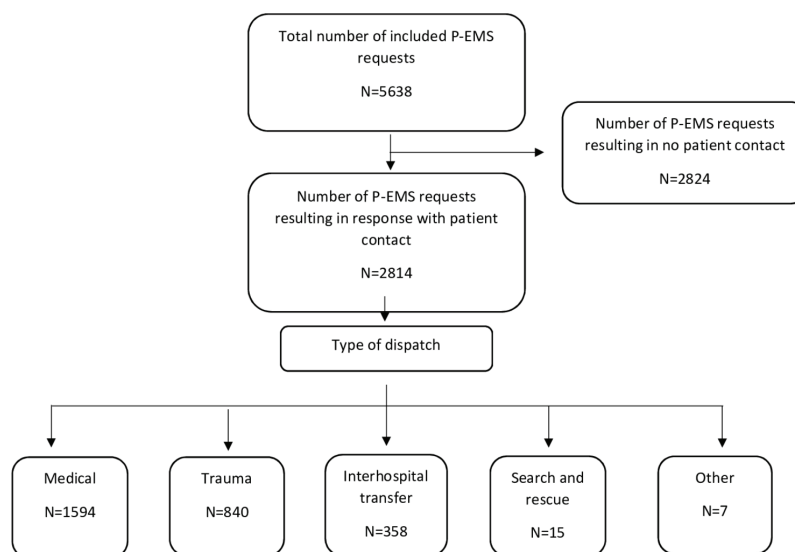
The QIs used in this study were developed by an expert panel through a consensus process.<sup>13</sup> One of the 18 members of the expert panel was a leader from a leading Norwegian patient organisation. This was done to secure user-expertise in the development of QIs.

For this particular study, no patients were involved in setting the research question, nor were they involved in the design or conduct of the study. No patients were asked to advise on the interpretation or writing up of results. The results will be disseminated via our local authorities and conference presentations. There are no plans to disseminate the results of the research to study participants.

### RESULTS

Despite the thorough and explicit definitions of QIs, a feasibility test was done first because this generally identifies variables that require modification. Omitting the feasibility test is not recommended.<sup>18</sup> Based on the experiences and comments from both recording physicians and the national coordinators during the 2 weeks feasibility test, we concluded that the necessary input data for the QIs were available in the participating services. There was no feedback indicating that the data were difficult to obtain. However, the definition of four QIs required





**Figure 1** Study population with type of dispatch for physician-staffed emergency medical services (P-EMS) responses with patient contact.

clarification. The changes done by the study group are documented in online supplementary file 2.

#### Participants and descriptive data

The dataset consists of 5638 requests for P-EMS. There were 2814 requests that resulted in completed responses with patient contact. Reasons for requests without patient contact may be cancelled responses, rejected responses due to weather or no need for P-EMS as judged by the P-EMS physician. The different dispatch types for the responses with patient contact are depicted in figure 1.

#### Outcome data and main results

The assessment of the QI feasibility, variability, rankability, actionability and documentation is depicted in table 1. The feasibility assessment was done based on comments from the recording physicians. The other four QI characteristics were assessed by the authors. The variability assessment of the QIs was based on the figures in table 2; the base-specific mean and median values with corresponding variances are shown for each QI. Documentation was assessed based on the existing literature.

**Table 1** Essential characteristics of the applied quality indicators

Quality indicator	Feasibility	Rankability	Variability	Actionability	Documentation
Able to respond immediately when alarmed	Good	Good	Good	Poor	Fair <sup>28 29</sup>
Time to arrival of P-EMS	Good	Good	Good	Poor	Fair <sup>28 29</sup>
On scene time	Good	Fair	Fair	Good	Fair <sup>11 30–32</sup>
Time to preferred destination	Good	Good	Good	Poor	Good <sup>33 34</sup>
Patients arriving hospital alive	Good	Good	Fair	Fair	Good <sup>35 36</sup>
Debriefed responses	Good	Good	Good	Good	Fair <sup>37 38</sup>
Adverse events	Good	Good	Poor	Good	Good <sup>39 40</sup>
Complete documentation	Good	Good	Good	Good	Good <sup>41 42</sup>
Guidelines for actual medical problem	Good	Good	Good	Good	Fair <sup>43–46</sup>
P-EMS involvement in dispatch	Good	Good	Good	Fair	Poor <sup>47</sup>
P-EMS necessary to provide appropriate care	Good	Good	Good	Fair	Fair <sup>48 49</sup>
Provision of advanced treatment	Good	Good	Good	Poor	Fair <sup>50 51</sup>
Significant logistical contribution	Good	Good	Good	Poor	Good <sup>33 34 52</sup>
Patients enrolled in research projects	Good	Good	Fair	Good	Fair <sup>7</sup>
Care for relatives	Good	Good	Fair	Good	Fair <sup>53–55</sup>

P-EMS, physician-staffed emergency medical services.

**Table 2** Variability of QIs (note: the columns ‘minimum mean value’ and ‘maximum mean value’ refer to the lowest and highest mean values from the participating P-EMS bases)

QI	No. of responses included	Missing (N)	Unit of QI	Mean (95% CI)	Median (IQR)	Minimum mean value	Maximum mean value
Able to respond immediately when alarmed	5599	39	%	89 (86 to 92)	90 (84–94)	78	97
Time to arrival of P-EMS	2428	6	minutes	27 (24 to 30)	26 (23–31)	18	36
On scene time	2427	7	minutes	20 (19 to 22)	21 (19–22)	14	26
Time to preferred destination	2226	19	minutes	63 (59 to 67)	63 (58–69)	46	74
Patients arriving hospital alive	2809	5	%	91 (89 to 93)	92 (88–94)	85	98
Debriefed responses	2809	5	%	74 (64 to 83)	78 (64–88)	29	97
Adverse events	5572	27	%	2 (1 to 3)	1 (1–3)	1	7
Complete documentation	2798	16	%	64 (51 to 76)	76 (34–80)	25	91
Guidelines for actual medical problem	2802	12	%	60 (48 to 72)	64 (45–77)	15	87
P-EMS involvement in dispatch	3669	29	%	47 (27 to 66)	34 (12–94)	7	98
P-EMS necessary to provide appropriate care	2808	6	%	39 (35 to 43)	39 (34–43)	27	52
Provision of advanced treatment	2804	10	%	49 (43 to 55)	48 (39–58)	33	71
Significant logistical contribution	2795	19	%	43 (32 to 55)	51 (24–58)	6	80
Patients enrolled in research projects	2788	26	%	6 (–1 to 13)	0 (1–3)	0	40
Care for relatives	2803	11	%	94 (92 to 96)	94 (93–97)	87	100

P-EMS, physician-staffed emergency medical services; QI, quality indicator.

Actionability was assessed as adequate for 10 QIs. The actionability of the QI ‘Able to respond immediately when alarmed’ was assessed as being poor because this is primarily determined by weather and concurrency conflicts. Further, the actionability was assessed as being poor for the QIs ‘Time to arrival of P-EMS’ and ‘Time to preferred destination’ because these time variables largely depend on where the patient is located geographically, and the P-EMS service cannot influence this. Moreover, the actionability was assessed as being poor for the QIs ‘Provision of advanced treatment’ and ‘Significant logistical contribution’. In our opinion, this is primarily the case for P-EMS services who are not involved in the dispatch decision. The actionability of these two QIs is fair in P-EMS services where the acceptance of a request is at the P-EMS physician’s discretion.

We used the data from the participating bases as a description of the current performance status pertaining to the QIs. Based on these figures, we proposed a benchmark level and a graphical presentation of three performance levels for the different QIs. Yellow area represents average performance, red represents low performance and green is high performance. Our objective was that these benchmarks serve as a tool for quality improvement in comparable P-EMSs in the future. The benchmarking is presented in figure 2.

Table 3 shows how the benchmarking system can compare the performance of different bases. In the actual

example, we used two of the participating bases as examples and call them Base 1 and Base 2. In the table, the actual value for each QI and its corresponding benchmark colour is depicted for all 15 QIs. For every high performance, the bases are given one point. For every low performance, the bases are given –1 point. The average performances are given 0 point. Thus, we end up with a sum or a total quality score that is between –15 and 15 for each base.

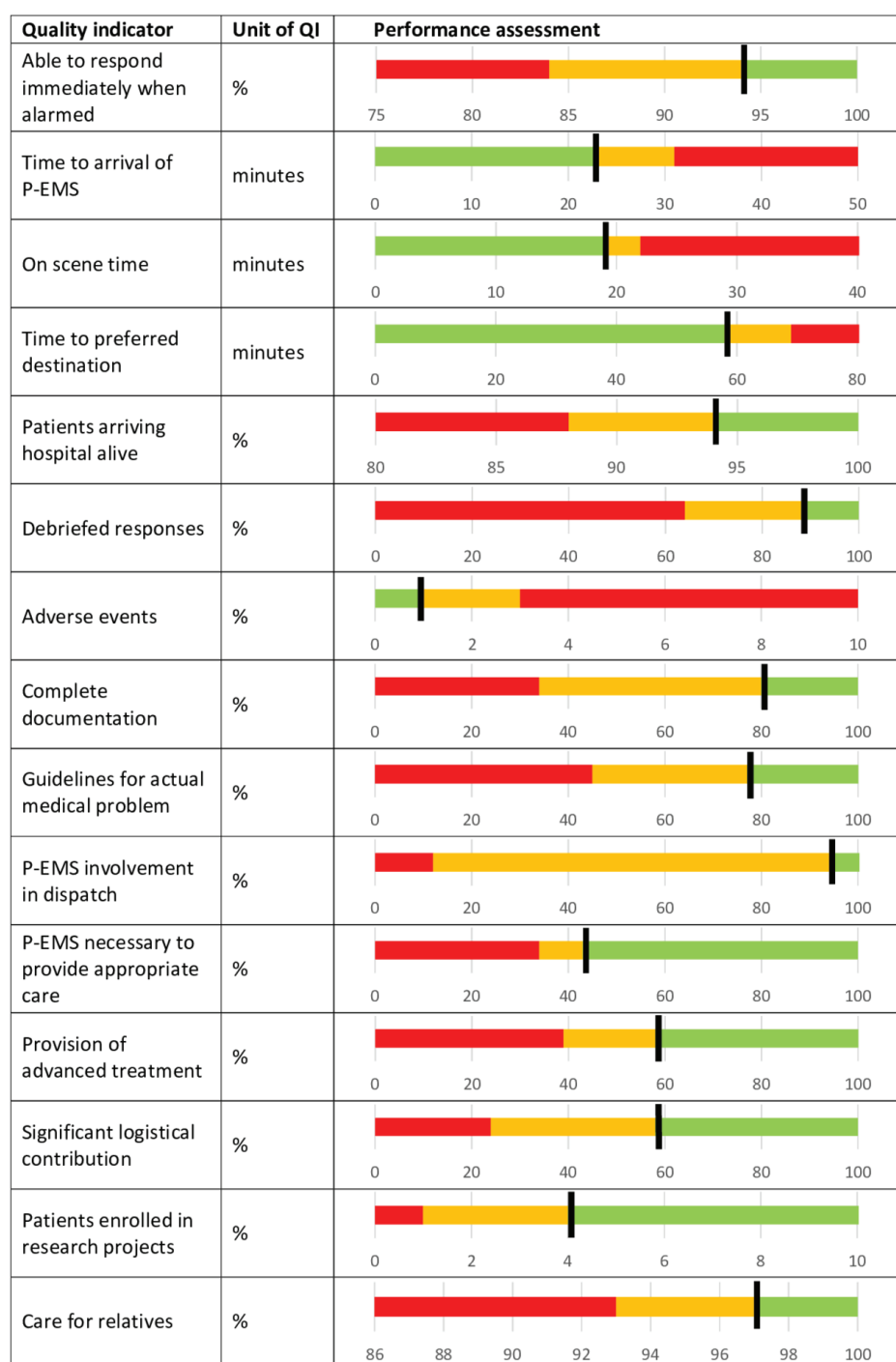
## DISCUSSION

### Key results

A set of 15 QIs were developed by an expert panel for P-EMS and were tested by applying the QIs in 5638 responses from 16 Nordic P-EMS bases. The feasibility of obtaining the necessary data for these QIs was good. The variability of the QIs was evaluated and is acceptable for all QIs except from the QI ‘Adverse events’. We used the dataset to propose benchmarks for all QIs as well as a total quality score: both of these can be used as tools for future quality measurement in P-EMS. Nonetheless, we assessed the actionability of some QIs to be low. That is especially true for QIs that measure the timeliness of P-EMS.

### Interpretation and generalisability

The patients treated by Nordic P-EMS services are heterogeneous: primary trauma and medical responses for every age group, secondary transports including neonatal



■ Low ■ Average ■ High

**Figure 2** Benchmarking of quality indicators (QIs). Green zone, high performance; yellow zone, average performance; red zone, low performance. The benchmark is set at the transition between green and yellow zones and marked with a black and fat vertical line.

**Table 3** Illustration of comparison between services using the proposed benchmarks

Quality indicator	Unit of QI	P-EMS base	
		Base 1	Base 2
Able to respond immediately when alarmed	%	96	83
Time to arrival of P-EMS	minutes	27	24
On scene time	minutes	12	14
Time to preferred destination	minutes	62	62
Patients arriving hospital alive	%	92	95
Debriefed responses	%	88	66
Adverse events	%	3	1
Complete documentation	%	27	34
Guidelines for actual medical problem	%	15	41
P-EMS involvement in dispatch	%	48	95
P-EMS necessary to provide appropriate care	%	35	43
Provision of advanced treatment	%	33	37
Significant logistical contribution	%	50	52
Patients enrolled in research projects	%	0	10
Care for relatives	%	100	96
Total quality score	Points (Scale: -15,15)	-1	1

Time variables are presented as medians as they are not normally distributed. The remaining QIs are presented as means of proportions.  
P-EMS, physician-staffed emergency medical services.

transports and SAR responses, among others. The reason for including all kinds of P-EMS responses was to get as accurate of a picture as possible to the actual patient panorama. The reason for also including P-EMS requests without patient contact was to get an impression of safety issues, availability and P-EMS involvement in dispatch for these responses.

When interpreting quality measurements, it is important to be aware that some QI performances may intercorrelate. Imagine a mountaineer traumatised with spinal injury and neurogenic shock after suffering a fall. Packing the patient well to prevent further hypothermia and placement of an arterial line followed by vasopressors for adequate blood pressure might prevent further neurological injury—even if it takes time. In

this example, too much focus on reducing on scene time could lead to a higher threshold for providing advanced treatment to correct deranged physiology. For some patients, this can be detrimental. For other patient groups, however, for example, patients with severe intra-abdominal bleeding and short transportation time to the nearest hospital, refraining from advanced treatment is likely to be beneficial. This illustrates that QIs must be interpreted with caution and that too much focus on one QI may lead to an undesired attention shift in clinical practice.

### Variability

According to Davies *et al*, there must be a certain degree of variability in the corresponding data for a QI to be meaningful.<sup>21</sup> If all P-EMS services report that they have 100% complete documentation every month—for example, because the electronic journal system does not allow the physicians to document incompletely—then it is not an interesting QI for quality improvement initiatives. However, a stable performance without much variation does not necessarily represent good system performance. The entire system may be uniformly underperforming, and thus goal-directed quality improvement may be indicated.

Even though the variation for a QI may be low within a single P-EMS service, there may be a high variation when assessing data from all services as a whole. When it is considered appropriate to compare single services with one another, a QI can still have enough variability to be useful. Due to the documented similarities between Nordic P-EMSs, including a comparable patient population, it is not reasonable to think that a high variability is merely a result of different case-mix.<sup>14</sup> It plausibly reflects real differences in performance.

### Low rate QIs

As supported by Gisvold *et al*, we conclude that events used as QIs must occur with a certain frequency.<sup>22</sup> In our dataset, we would describe the QI 'Adverse events' as a 'Low rate QI'. Low rate of an event limits statistical appraisal, as variation may be the result of chance. Moreover, it is difficult to use low rate of events as a continuous QI because changed rates of the event due to improvement efforts are difficult to separate from natural variation. A strictly quantitative approach to such data might therefore be less useful. However, analysing these data as 'sentinel events', where problems are studied individually to identify causal relationships and preventative measures, might be an adequate approach. Using the QI 'Adverse events' for this purpose in the future seems reasonable. When rates are too low to do statistically meaningful comparisons, qualitative data can be effective—even from small samples. Qualitative data in quality measurement can uncover issues that quantitative data may never reveal.<sup>23</sup>



### Documentation/validity

The validity of a QI depends on a demonstrated link between a process or a structure and a higher probability of a favourable outcome. These relationships are preferably based on scientific literature. However, where little evidence exists, these linkages can be judged important to patient outcomes by clinical experts in a consensus process.<sup>18 24</sup> The selection process of the QIs tested in this study is thus widely accepted.<sup>13</sup>

If a QI does not satisfy the criteria above (especially feasibility, rankability and variability, indicating that the variable is 'statistically' inappropriate), but the QI is still regarded clinically important, the QI may be revised to be used for the intended purpose in the future.

### Benchmarking

The data in this study are assumed representative for the P-EMS patient population and therefore transferable to other P-EMS bases in the Nordic countries. The number of responses is also relatively high. Thus, it seems reasonable to use the performances in this study as a basis for proposing benchmarks for each QI. When doing so, there are principally two approaches. The first option is to let the average score for the whole group (peer group level) serve as the average performance, and then refer to low-performance and high-performance groups related to average score. The average score will then serve as a threshold—and the aim is to perform above this level. The second option is defining a higher score, an 'excellent level' based on the performances of the best P-EMS bases. Performances above this higher level will now be the goal; in other words, this is a more ambitious form of benchmarking. How to choose the peer group is also debatable: the more homogeneous the group, the better for reliability. However, a larger group with more diversity increases the chance to learn from 'excellent performers'.<sup>25</sup>

According to Moore, 'benchmarking is an improvement process used to discover and incorporate best practices into an operation'.<sup>26</sup> When excellent performers are known, and benchmarks set, different services can measure their performance in relation to these benchmarks, which can be considered as standards. When services reach these standards, new benchmarks can be set, thus taking the quality improvement work to an even higher level. Moreover, although QIs exist for many areas in healthcare, methods to combine them into a single total score are underdeveloped.<sup>27</sup> We consider that the total quality score for P-EMS, as described in this paper, can be an additional tool in future quality measurement.

### Future needs

Feasible and reliable quality measurement largely depends on robust documentation systems to ensure proper data quality and to avoid added documentation workload for the clinicians. Ideally, as many variables as possible should be collected automatically through electronic data capture.

The relationship between different QI performance and a hard endpoint, such as 30-day mortality, remains unknown. Therefore, a study exploring this relationship is warranted.

### Limitations

One of the limitations of the current analysis is that the attending physicians registered all the data. They are therefore subject to registration bias and recall bias.

Except from the feasibility of the QIs, the different QI characteristics were assessed by the authors. The variability was assessed based on the data (mean and median). However, thresholds for defining poor, fair and good variability for QIs do not exist, to the best of our knowledge. Therefore, conclusions on this topic were a result of assessments and consensus among all authors. Conclusions on rankability, actionability and documentation were also resulting from assessment and consensus among the authors.

### CONCLUSIONS

In this study, a set of 15 QIs developed for P-EMS have been tested for necessary QI characteristics. The feasibility of obtaining the necessary data for these QIs was good. The variability of the QIs was adequate for all QIs except from the QI 'Adverse events', which was a 'Low rate QI'. Therefore, it seems reasonable to use this QI simply for identifying adverse events and then analyse them as 'sentinel events', rather than using these data in a quantitative analysis. The actionability was assessed poor for five QIs. Three of these QIs are measuring the timeliness of P-EMS. Some QIs depend on characteristics of the P-EMS services that might differ, such as patient volume, distances and patient characteristics; thus, they should be interpreted with caution for service comparison. However, it seems more straightforward to use these QIs for internal quality measurement of a service. To aid future quality measurements in P-EMS, benchmarks for all QIs have been proposed. In addition, we have presented a variable combining the QI performances into one single score, the total quality score.

### Author affiliations

<sup>1</sup>Norwegian Air Ambulance Foundation, Oslo, Norway

<sup>2</sup>Department of Emergency Medicine and Pre-Hospital Services, St. Olav University Hospital, Trondheim, Norway

<sup>3</sup>Research and Development Unit, FinnHEMS Ltd, Vantaa, Finland

<sup>4</sup>Department of Anaesthesiology, Aarhus University Hospital, Aarhus N, Denmark

<sup>5</sup>Danish Air Ambulance, Aarhus, Denmark

<sup>6</sup>Airborne Intensive Care Unit, Department of Anaesthesia, Perioperative Management and Intensive Care Medicine, Uppsala University Hospital, Uppsala, Sweden

**Contribution** We thank the following physician-staffed emergency medical services for participating in the data collection: Vantaa HEMS, Turku HEMS, Tampere HEMS, Oulu HEMS, Kuopio HEMS, all Finland. Skive HEMS, Billund HEMS, Ringsted HEMS, all Denmark. Uppsala HEMS, Karlstad HEMS, both Sweden. Lørenskog HEMS, Rygge SAR, Arendal HEMS, Stavanger HEMS, Trondheim HEMS, Ørland SAR, all Norway. We thank Bjørn Henrik Moshuus, IT Manager at The Norwegian Air Ambulance Foundation, for developing the web-based database. We thank



Päivi Laukkanen-Nevala for statistical support and Jukka Tennilä for IT support at FinnHEMS Research and Development Unit, and Sasu Liuhanen at Absolute Imaginary for the adaption of the FinnHEMS database. We thank all the donors of The Norwegian Air Ambulance Foundation for the financial support that made this study possible.

**Funding** This study was funded by The Norwegian Air Ambulance Foundation.

**Competing interests** HH and AK holds research positions in The Norwegian Air Ambulance Foundation, a non-commercial charity owning The Norwegian Air Ambulance, which is the contractor of the national air ambulance service in Norway.

**Patient consent for publication** Not required.

**Ethics approval** The study was approved by the Committees for Medical and Health Research Ethics in Sweden (reference number: 2016/109) and Finland (reference number: R16031), respectively. In Denmark, application was waved by The Committee for Medical and Health Research Ethics due to the strictly descriptive nature of the study. The Norwegian Committee for Medical and Health Research Ethics defined the study to fall outside their legislation (reference number: 2016/371). This necessitated applications to The Norwegian Data Protection Authority (reference number: 16/01113-2/SB0), The Norwegian Directorate of Health (reference number: 16/14024-3) and the Data Protection Officers at the participating Norwegian health services who all approved the study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Helge Haugland <http://orcid.org/0000-0002-5869-6675>

Leif Rognås <http://orcid.org/0000-0002-2542-565X>

#### REFERENCES

- Porter ME. What is value in health care? *N Engl J Med Overseas Ed* 2010;363:2477–81.
- Institute of Medicine. *Emergency medical services at a crossroads*. Washington DC: The National Academies Press, 2006.
- World Health Organization. Quality of care: a process for making strategic choices in health systems 2006.
- National Committee for Quality Assurance. The essential guide to health care quality. Available: [https://www.ncqa.org/Portals/0/Publications/Resource%20Library/NCQA\\_Primer\\_web.pdf](https://www.ncqa.org/Portals/0/Publications/Resource%20Library/NCQA_Primer_web.pdf) [Accessed 01 Mar 2016].
- Haugland H, Uleberg O, Klepstad P, et al. Quality measurement in physician-staffed emergency medical services: a systematic literature review. *Int J Qual Health Care* 2019;31:2–10.
- Snooks H, Evans A, Wells B, et al. What are the highest priorities for research in emergency prehospital care? *Emerg Med J* 2009;26:549–50.
- Fevang E, Lockey D, Thompson J, et al. The top five research priorities in physician-provided pre-hospital critical care: a consensus report from a European research collaboration. *Scand J Trauma Resusc Emerg Med* 2011;19:57.
- Rehn M, Krüger AJ. Quality improvement in pre-hospital critical care: increased value through research and publication. *Scand J Trauma Resusc Emerg Med* 2014;22:34.
- Lawrence M, Olesen F. Indicators of quality in health care. *Eur J Gen Pract* 1997;3:103–8.
- Nilsen KS, Tjelmeland KS, Halvorsen J. Olasveengen. Kvalitetsindikatorer i den akuttmedisinske kjeden. [Norwegian], 2015. Available: [www.nakos.no](http://www.nakos.no)
- Reid BO, Rehn M, Uleberg O, et al. Physician-provided prehospital critical care, effect on patient physiology dynamics and on-scene time. *Eur J Emerg Med* 2018;25:114–9.
- Institute of Medicine. *Crossing the quality chasm: a new health system for the twenty-first century*. Washington: National Academies Press, 2001.
- Haugland H, Rehn M, Klepstad P, et al. Developing quality indicators for physician-staffed emergency medical services: a consensus process. *Scand J Trauma Resusc Emerg Med* 2017;25:14.
- Krüger AJ, Skogvoll E, Castrén M, et al. Scandinavian pre-hospital physician-manned Emergency Medical Services--same concept across borders? *Resuscitation* 2010;81:427–33.
- Krüger AJ, Lossius HM, Mikkelsen S, et al. Pre-Hospital critical care by anaesthesiologist-staffed pre-hospital services in Scandinavia: a prospective population-based study. *Acta Anaesthesiol Scand* 2013;57:1175–85.
- Langhelle A, Lossius HM, Silfvast T, et al. International EMS systems: the Nordic countries. *Resuscitation* 2004;61:9–21.
- von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495–9.
- Rubin HR, Pronovost P, Diette GB. Methodology matters. from a process of care to a measure: the development and testing of a quality indicator. *Int J Qual Health Care* 2001;13:489–96.
- Eea K. Health care quality indicators project conceptual framework paper contract No: 23 2006.
- Agency for Healthcare Research and Quality. Quality indicator measure development, implementation, maintenance, and retirement 2011.
- Davies SM, Geppert J, McClellan M, et al. Refinement of the HCUP quality indicators. agency for healthcare research and quality, technical reviews No.4 2001.
- Gisvold SE, Fasting S. How do we know that we are doing a good job - can we measure the quality of our work? *Best Pract Res Clin Anaesthesiol* 2011;25:109–22.
- NHS Institute for Innovation and Improvement. *The good indicators guide: understanding how to use and choose indicators*, 2008.
- Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003;15:523–30.
- Romano PS. Peer group benchmarks are not appropriate for health care quality report cards. *Am Heart J* 2004;148:921–3.
- Moore L. Measuring quality and effectiveness of prehospital EMS. *Prehosp Emerg Care* 1999;3:325–31.
- Ken Lee KH, Matthew Austin J, Pronovost PJ. Developing a measure of value in health care. *Value in Health* 2016;19:323–5.
- Wilde ET. Do emergency medical system response times matter for health outcomes? *Health Econ* 2013;22:790–806.
- Al-Shaqsi SZK. Response time as a sole performance indicator in EMS: pitfalls and solutions. *Open Access Emerg Med* 2010;2:1–6.
- Walcher F, Weinlich M, Conrad G, et al. Prehospital ultrasound imaging improves management of abdominal trauma. *Br J Surg* 2006;93:238–42.
- Brown JB, Rosengart MR, Forsythe RM, et al. Not all prehospital time is equal: influence of scene time on mortality. *J Trauma Acute Care Surg* 2016;81:93–100.
- Harmen AMK, Giannakopoulos GF, Moerbeek PR, et al. The influence of prehospital time on trauma patients outcome: a systematic review. *Injury* 2015;46:602–9.
- Jauch EC, Saver JL, Adams HP, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American heart Association/ American stroke association. *Stroke* 2013;44:870–947.
- Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *The Lancet* 2003;361:13–20.
- Wong MKY, Morrison LJ, Qiu F, et al. Trends in short- and long-term survival among out-of-hospital cardiac arrest patients alive at hospital arrival. *Circulation* 2014;130:1883–90.
- Clark DE, Doolittle PC, Winchell RJ, et al. The effect of hospital care on early survival after penetrating trauma. *Inj Epidemiol* 2014;1.
- Greenberg N. A critical review of psychological Debriefing: the management of psychological health after traumatic experiences. *J R Nav Med Serv* 2001;87:158–61.
- Kessler DO, Cheng A, Mullan PC. Debriefing in the emergency department after clinical events: a practical guide. *Ann Emerg Med* 2015;65:690–8.
- O'Connor RE, Slovis CM, Hunt RC, et al. Eliminating errors in emergency medical services: realities and recommendations. *Prehosp Emerg Care* 2002;6:107–13.
- Institute of Medicine. To err is human. Available: <http://www.iom.edu/1999>
- Krüger AJ, Lockey D, Kurola J, et al. A consensus-based template for documenting and reporting in physician-staffed pre-hospital services. *Scand J Trauma Resusc Emerg Med* 2011;19:71.
- Helm M, Hauke J, Schlechtriemen T, et al. Paper-assisted digital Mission documentation in air rescue services. Quality management in preclinical emergency medicine]. *Anaesthesist* 2007;56:877–85.



- 43 Gundersen L. The effect of clinical practice guidelines on variations in care. *Ann Intern Med* 2000;133:317–8.
- 44 NHS Centre for Reviews and Dissemination. Getting evidence into practice. *Effective Health Care* 1999;5:1–16.
- 45 Carnett WG. Clinical practice guidelines: a tool to improve care. *Qual Manag Health Care* 1999;8:13–21.
- 46 Lugtenberg M, Burgers JS, Westert GP. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. *Quality and Safety in Health Care* 2009;18:385–92.
- 47 Munro S, Joy M, de Coverly R, et al. A novel method of non-clinical dispatch is associated with a higher rate of critical helicopter emergency medical service intervention. *Scand J Trauma Resusc Emerg Med* 2018;26:84.
- 48 Haner A, Örringe P, Khorram-Manesh A. The role of physician-staffed ambulances: the outcome of a pilot study. *J Acute Dis* 2015;4:63–7.
- 49 van Schuppen H, Bierens J. Understanding the prehospital physician controversy. step 2: analysis of on-scene treatment by ambulance nurses and helicopter emergency medical service physicians. *Eur J Emerg Med* 2015;22:384–90.
- 50 Pakkanen T, Kämäräinen A, Huhtala H, et al. Physician-Staffed helicopter emergency medical service has a beneficial impact on the incidence of prehospital hypoxia and secured airways on patients with severe traumatic brain injury. *Scand J Trauma Resusc Emerg Med* 2017;25:94.
- 51 Bøtker MT, Bakke SA, Christensen EF. A systematic review of controlled studies: do physicians increase survival with prehospital treatment? *Scand J Trauma Resusc Emerg Med* 2009;17:12.
- 52 Samdal M, Haugland HH, Fjeldet C, et al. Static rope evacuation by helicopter emergency medical services in rescue operations in Southeast Norway. *Wilderness Environ Med* 2018;29:315–24.
- 53 Committee on Hospital Care and Institute for Patient- and Family-Centered Care. Patient- and family-centered care and the pediatrician's role. *Pediatrics* 2012;129:394–404.
- 54 Institute for Patient- and Family-Centered care. Available: [www.ipfcc.org](http://www.ipfcc.org) [Accessed 03 Jul 2016].
- 55 Dowling J, Vender J, Guilianelli S, et al. A model of family-centered care and satisfaction predictors: the critical care family assistance program. *Chest* 2005;128:81s–92.

## Paper IV





# Exploring the relationship between mortality and quality scores in Nordic physician-staffed emergency medical services

**Authors:** Helge Haugland<sup>1,2</sup>, Anna Olkinuora<sup>3</sup>, Leif Rognås<sup>4,5</sup>, David Ohlén<sup>6</sup>, Andreas Krüger<sup>1,2</sup>.

**Author details:** <sup>1</sup> Department for Research and Development, The Norwegian Air Ambulance Foundation, Oslo, Norway. <sup>2</sup> Department of Emergency Medicine and Pre-Hospital Services, St. Olav University Hospital, Trondheim, Norway. <sup>3</sup> Research and Development Unit, FinnHEMS Ltd, Vantaa, Finland. <sup>4</sup> Department of Anaesthesia, Aarhus University Hospital, Aarhus, Denmark. <sup>5</sup> Danish Air Ambulance, Aarhus, Denmark. <sup>6</sup> Airborne Intensive Care Unit, Department of Anaesthesia, Perioperative Management and Intensive Care Medicine, Uppsala University Hospital, Uppsala, Sweden.

**Corresponding author:** Helge Haugland MD. The Norwegian Air Ambulance Foundation. Postbox 414 Sentrum, 0103 OSLO. E-mail: [helge.haugland@norskluftambulanse.no](mailto:helge.haugland@norskluftambulanse.no) Phone: 0047-64904444. Fax-nr: 0047-72512380

**Name and location of institutions where the work was carried out:** Vantaa HEMS, Turku HEMS, Tampere HEMS, Oulu HEMS, Kuopio HEMS, all Finland. Skive HEMS, Billund HEMS, Ringsted HEMS, all Denmark. Uppsala HEMS, Karlstad HEMS, both Sweden. Lørenskog HEMS, Rygge SAR, Arendal HEMS, Stavanger HEMS, Trondheim HEMS, Ørland SAR, all Norway.

(HEMS=Helicopter Emergency Medical Service. SAR=Search and Rescue)

**Conflict of interest:**

The authors declare the following conflict of interests: HH holds a position as a PhD-student in The Norwegian Air Ambulance Foundation, a non-commercial charity owning The Norwegian Air Ambulance Ltd, which is the contractor of the national air ambulance service in Norway.

## **ABSTRACT:**

**Background:** In a recent study 15 response-specific quality indicators for physician-staffed emergency medical services were developed. These quality indicators primarily measure the process quality of care. The aim of this study was to assess if there was a correlation between these quality indicators and 30-days mortality.

**Methods:** In this prospective observational study, 16 physician-staffed emergency medical services in Finland, Sweden, Denmark, and Norway registered data to score the 15 quality indicators. All requests to the physician-staffed emergency medical services were included in the study regardless of patient characteristics. Quality indicator means for the groups "Alive after 30 days" and "Dead after 30 days" were compared using Chi-Square test for categorical variables and Mann-Whitney U test for continuous variables. Moreover, a linear correlation analysis was done to assess the correlation between 30-days mortality and a composite variable adding up the performance of the quality indicators; The Total Quality Score.

**Results:** We recorded 2,808 responses in the study period. For nine out of fifteen quality indicators, there was significant difference in mean scores between the 30 days survivors and non-survivors. In the correlation analysis between The Total Quality Score and 30-days mortality the Pearson r coefficient was 0,125, indicating no significant correlation.

**Conclusion:** Good overall process quality did not seem to correlate with good outcome quality. However, these are complementary measures, both with undeniable value to identify the total quality achieved in physician-staffed emergency medical services. Thus, combining mortality measures and multiple process quality indicators seems adequate for future quality measurement.

## **INTRODUCTION:**

### **Background/rationale:**

The literature on quality indicators in pre-hospital care is scarce and research initiatives on this topic have been warranted[1, 2]. In a study from 2017 we therefore developed a set of multi-dimensional quality indicators for physician-staffed emergency medical services (P-EMS) through a consensus process. The expert panel agreed on 15 response-specific quality indicators (QIs) for P-EMS; the so called EQUIPE quality indicators[3]. These quality indicators are primarily process indicators; i.e. they describe the process of care provided by P-EMS, rather than the outcome of this care. There are good reasons for using process indicators for this purpose. Process indicators are considered useful for short time frames and when it is difficult to adjust for patient factors[4], and they are therefore particularly relevant for P-EMS. Further, process indicators often provide a more direct measurement of quality of care, whereas structure and outcome indicators often measure this quality more indirectly[5].

The fact that process indicators seem particularly suitable for prehospital services does not make outcome indicators like mortality less important. Mortality within a defined period after hospital admission (commonly 30 days) is considered an appropriate measure of the hospital care outcome[6]. Some have even argued that outcome indicators are the “ultimate measure of quality in care”[7]. However, for in-hospital services the use of mortality as a quality measure has been questioned because the proportion of preventable deaths has been found lower than earlier estimates, thus making mortality less suitable as a quality measure[8]. Hospital mortality has also been used when assessing the effects of pre-hospital care. A paradox is that outstanding pre-hospital care in fact may increase hospital mortality because patients survive until hospital admission rather than die on scene or en route[9, 10].

A widely cited definition of quality that also might be applicable for P-EMS systems is “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge”[11]. This definition supports the idea that outcome alone is not sufficient to describe the total quality of care. However, it seems reasonable that good processes ultimately lead to better outcome. This principle is used in other high-risk businesses as well; in aviation, petroleum industry and nuclear power plants for instance, the process quality is measured – assuming that good process quality will prevent a major incident[12]. In medical research, however, the use of hard end points, especially mortality, has been the gold standard. Nevertheless, a study on the relationship between quality and mortality for acute hospitals in England concluded that high mortality was not an adequate marker of overall poor quality[13]. In P-EMS, no study has explored the relationship between quality indicators and mortality, to the best of our knowledge. In this study we wanted to explore if high performance measured by QIs is correlated to low mortality and vice versa.

As such, the aim of this study was to assess if there is a correlation between the EQUIPE quality indicators for P-EMS and 30-days mortality.

## **METHODS:**

### **Study design and setting**

In this prospective observational study, 16 physician-staffed helicopter emergency services in Finland, Sweden, Denmark, and Norway registered data for the EQUIPE quality indicators. Additionally, 30-days mortality data was collected for all included patients. There has previously been documented significant system similarities in the P-EMS of the four participating countries making them a suitable arena for multi-centre studies[14]. All services do primary responses, and the Swedish, Danish, and Norwegian services also do secondary responses; i.e. responses where the patient is located outside a hospital, and the latter is inter-hospital transfers. The Finnish HEMS units do inter-hospital transfers only in rare exceptions. Moreover, the Norwegian services also do search and rescue responses (SAR-responses). In addition, one Swedish (Karlstad) and all Finnish and Norwegian bases dispose a rapid response car for responses close to the base and for responses in poor weather conditions that prevent flight operations. The study applied Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Guidelines[15].

### **Inclusion criteria and data variables**

All P-EMS requests were included in the study. Thus, we could include both completed and cancelled responses as well as stand-downs (responses cancelled by dispatch or crews on-scene) and rejected responses. Examples of reasons for rejecting a response might be weather conditions or the lack of medical need as judged by the P-EMS physician. The latter is possible in Sweden, Finland, and Norway where the acceptance or rejection of a response is at the P-EMS physicians' discretion. Inquiries with the provision of advice via telephone or radio as the only purpose were excluded. Primary and secondary responses as well as search and rescue (SAR) responses were included. For bases with both a helicopter and a rapid response car, responses were included regardless of the mode of transportation.

For the analysis of time variables we have omitted inter-hospital transfers and SAR-responses, as the nature of these responses is not comparable to primary responses pertaining to time consumption.

### **Data sources/measurement**

The Swedish, Danish and Norwegian services registered the data by using a web-based questionnaire (Formsite; Vroman systems, Inc., Chicago, Illinois). The Finnish HEMS collected the necessary data by including the quality indicators as part of their existing documentation database (FinnHEMS database, FHDB). FHDB is a national database, including both response and patient data, where all HEMS units register their responses. Some QIs could be gathered from the existing data, other QIs were either implemented as permanent variables or on a separate study sheet. Filling in all QIs was mandatory.

In all nations the data were collected after completed response by the P-EMS physician. Four national investigators performed data quality assurance and collected 30-days mortality in their respective nations. The data collection period was July 2016 – April 2017.

### **Statistical methods**

Results are presented using descriptive statistics. The QI proportions were recorded for QIs that are categorical variables; time was recorded in minutes for QIs that were continuous time variables. All quality indicators are reported by their mean and the corresponding 95% confidence interval. To explore a possible correlation between the quality indicators and 30-days mortality, the QI means for the groups “Alive after 30 days” and “Dead after 30 days” are compared using Chi-Square test for the categorical variables and Mann-Whitney U test for the continuous variables due to non-normality. Defined significance level is  $p < 0,05$ . A linear correlation analysis was done to assess the correlation between 30 days mortality and Total Quality Score.

### **Missing data**

Responses with missing data pertaining to 30-days mortality are omitted from the analysis.

### **Ethics approval and consent to participate**

The study was approved by the Committees for Medical and Health Research Ethics in Sweden (reference number: 2016/109) and Finland (reference number: R16031), respectively. In Denmark, application was waved by The Committee for Medical and Health Research Ethics due to the strictly descriptive nature of the study. The Norwegian Committee for Medical and Health Research Ethics defined the study to fall outside their legislation (reference number: 2016/371). According to the approvals from all four countries, the data was obtained without informed consent from patients or their next-of-kin. There was no deviation from regular clinical practice during the study period.

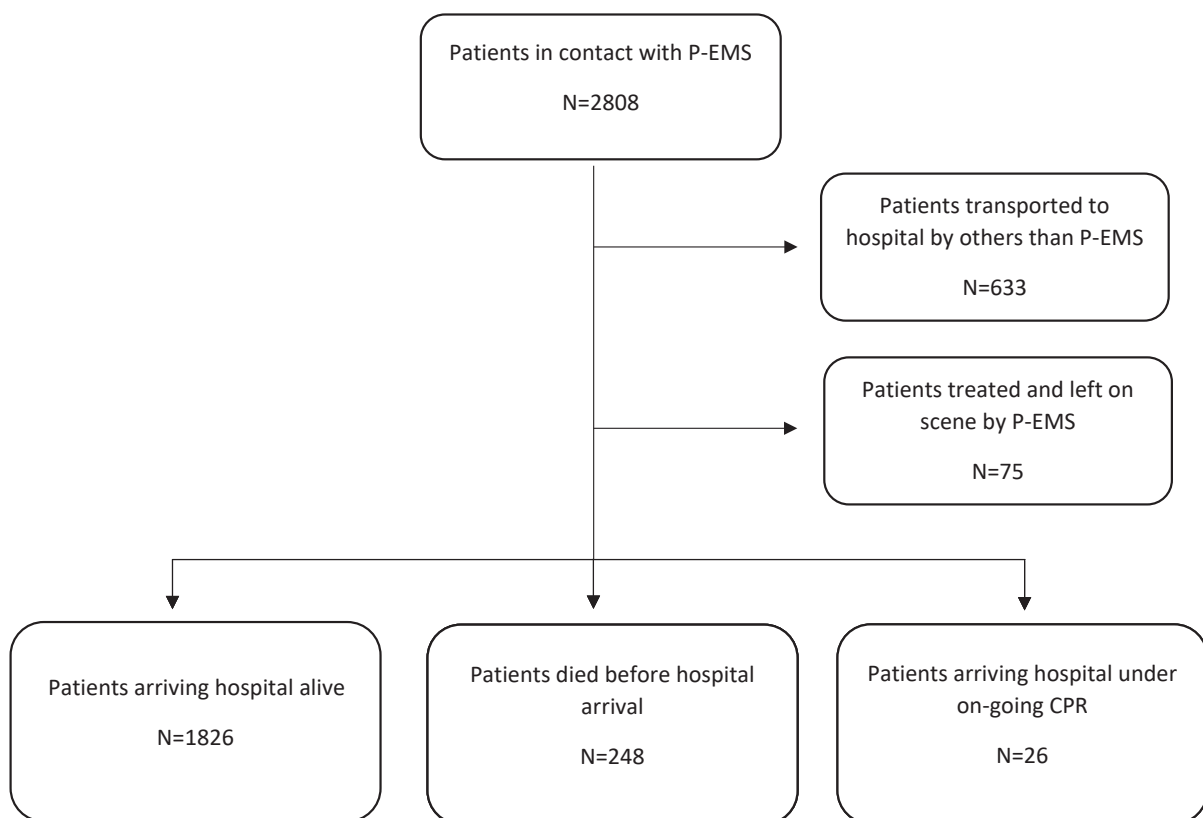
## RESULTS:

### Participants and descriptive data:

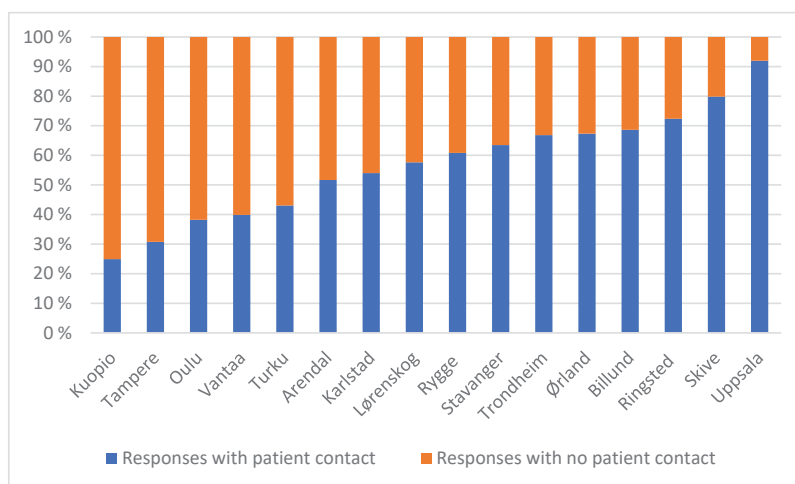
The dataset consisted of 5,638 requests for P-EMS. There were 2,808 requests that resulted in patient contact with P-EMS. The patient flow and survival from these requests is depicted in figure 1. The remaining 2,830 responses were either rejected or aborted. The proportion of responses with patient contact varied from 24,9 % and 92,0 % as depicted in figure 2. When the same proportions were calculated only for primary responses, i.e. excluding interhospital- transfers and Search and Rescue missions, the figures varied between 24,6 % 85,5 %.

### Outcome data and main results:

Of the 2808 patients cared for by P-EMS a total of 633 (22,5 %) patients were eventually transported by other services than P-EMS. 525 (82,9%) of these patients were still alive 30 days after the P-EMS response, 45 (7,1%) were dead and data were missing for 63 patients (10,0%).



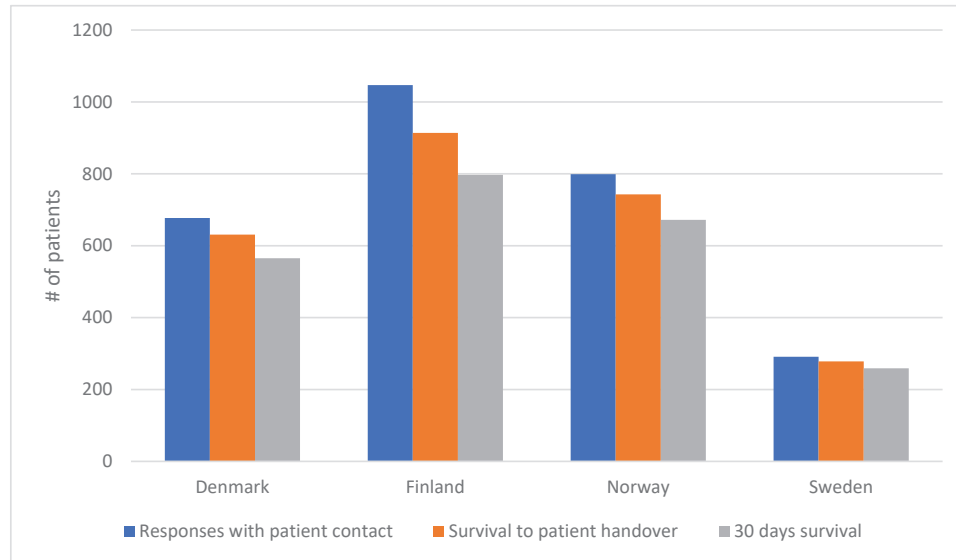
**Figure 1: Flowchart of the study.** “Treat and leave” = patients left on scene and not going to hospital. Missing data: 6 out of 2814 patients.



**Figure 2: Proportion of P-EMS responses with and without patient contact.**

In figure 3 “Survival to patient handover” and “30-days survival” is depicted for all four participating nations. Survival to patient handover is defined as survival until the patient is handed over in the hospital or as survival until handover to EMS when transported by others than P-EMS. Survival to patient handover was 93,2 % (Denmark), 87,3 % (Finland), 93,0 % (Norway) and 95,5 % (Sweden). The proportion of patients surviving until 30 days after the actual P-EMS response was 83,5 % (Denmark), 76,1 % (Finland), 84,1 % (Norway) and 89,0 % (Sweden), respectively.





**Figure 3: Survival for patients cared for by Nordic P-EMS services**

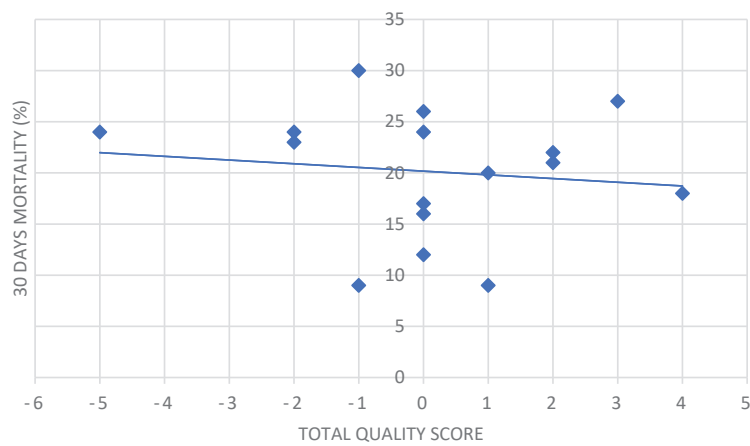
For all 15 QIs developed specifically for P-EMS, comparisons of QI score and survival are presented in table 1. For nine out of fifteen QIs, there was significant difference in mean QI scores between the survivors and non-survivors after 30 days.

Quality indicator	Unit	Alive after 30 days	Dead after 30 days	p-value
		mean (95% CI)	mean (95% CI)	
Ability to respond immediately when alarmed	%	95 (94-96)	95 (95%CI: 93-97)	0,226
Time to arrival of P-EMS	minutes	33 (31-34)	30 (28-32)	0,106
On scene time	minutes	19 (18-20)	29 (27-30)	0,000
Time to preferred destination	minutes	83 (76-89)	79 (68-89)	0,542
Survival to hospital	%	100 (100-100)	54 (50-58)	0,000
Debriefed responses	%	71 (69-73)	73 (70-77)	0,293
Adverse events	%	2 (1-2)	3 (1-4)	0,054
Complete documentation	%	64 (62-66)	74 (70-78)	0,000
Guidelines for actual medical problem	%	58 (55-60)	78 (74-82)	0,000
P-EMS involvement in dispatch	%	44 (42-46)	37 (33-41)	0,000
P-EMS necessary to provide appropriate care	%	35 (33-37)	55 (51-60)	0,000
Provision of advanced treatment	%	42 (40-44)	76 (73-80)	0,000
Significant logistical contribution	%	44 (42-46)	27 (23-31)	0,000
Patient enrolment in research projects	%	7 (6-9)	11 (8-14)	0,013
Care for relatives	%	93 (92-95)	95 (92-97)	0,413

**Table 1: Comparison of performance in different quality indicators for patients alive after 30 days and patients not alive after 30 days, respectively.**

In a recent study we have presented a composite variable adding up the performance of all 15 Quality Indicators, named The Total Quality Score (TQS)[16]. The performance for each QI will result in either

0 point (average performance; i.e. performances within the interquartile range (IQR)), 1 point (high performance; i.e. performances above the IQR) or -1 point (low performance; i.e. performances below the IQR). This gives the TQS a range from -15 to 15 points for each service. In this study, we did a correlation analysis between 30 days mortality and TQS to study if a high TQS results in a lower 30 days mortality and vice versa (figure 4). However, we omitted the QI “Survival to hospital” from the TQS due to the bias this would lead to in the analysis of 30-days survival. Thus, the range for the TQS in this analysis is between -14 and 14. The Pearson  $r$  coefficient is 0,125, indicating no significant correlation. This is supported by the  $p$ -value of 0,644.



**Figure 4: The correlation between Total Quality Score and 30 days mortality.**

## DISCUSSION:

For nine out of fifteen QIs we found a correlation between QI score and 30-days mortality. These QIs were “On scene time”, “Survival to hospital”, “Complete documentation”, “Guidelines for actual medical problem”, “P-EMS involvement in dispatch”, “P-EMS necessary to provide appropriate care”, “Provision of advanced treatment”, “Significant logistical contribution” and “Patient enrolment in research projects”. However, for these nine QIs there is no consistent pattern showing that presumptive good QI performances lead to lower mortality than poor performances and vice versa. Further, we did not find a significant correlation between Total Quality Score and 30-days mortality.

When exploring the relationship between QIs and 30-days mortality, we found that the on-scene time was significantly higher for the group of patients who were dead within 30 days after the P-EMS response. A possible explanation for this is that these patients presumably are in a more critical condition when P-EMS arrives, and that this necessitates time-consuming interventions. This finding is supported by the significantly higher proportion of advanced interventions in the group of non-survivors. On the other hand, some might argue that this correlation implies that long on-scene time is harmful for severely injured or ill patients. However, there is no adequate basis for drawing conclusions regarding causality in this study, only conclusions regarding correlation. Complete documentation was found more frequently in the group of non-survivors. This might be because the P-EMS physician feels a greater need for both obtaining and documenting clinical parameters when the patient is severely ill. Nevertheless, the defined key parameters should be relevant documentation for all patients[17]. Also the presence of guidelines was significantly higher in the group of non-survivors. A possible explanation for this might be that the services primarily establish guidelines for the most severe conditions, because a timely and efficient approach is of paramount importance in these situations. Involvement of the P-EMS physician when deciding if dispatch is appropriate occurred more often in the group of survivors. This might indicate that the alarm calls to EMCC for the most severely ill or injured patients leaves less doubt regarding the dispatch of P-EMS, while patients of lower severity are discussed more frequently with the P-EMS physician prior to dispatch. The proportion of responses in need of P-EMS to secure appropriate care was significantly higher in the group of non-survivors. This might be due to the higher need for advanced interventions and critical decision making in this group.

The Total Quality Score combines the different QIs into one single total score. Such combination scores have been described as underdeveloped in health care[18]. A combination score secures that multiple QIs form the basis for quality measurement; thus securing an adequate and comprehensive quality measure. However, we did not find a significant correlation between TQS and 30-days mortality.

Mortality measures are easy to define and have traditionally been important in reducing preventable deaths in health care. However, as preventable deaths have decreased due to improved care, the use of mortality measures alone no longer seems adequate[19]. Moreover, it is difficult to isolate the pre-hospital care so that mortality numbers refer to this phase only. Rather, mortality numbers refer to the complete chain of care until mortality is measured, i.e. pre-hospital care, emergency department, operation theatre, intensive care unit etc. Thus, mortality should not stand alone as QI, but be part of a comprehensive quality measurement approach, as one of several quality indicators. We argue that our findings underline these statements, as no correlation between many QIs and mortality was found. Moreover, for the QIs where a correlation with mortality was found, the results seem to be explained first and foremost by the patient population.

A major strength of mortality as quality indicator is the fact that it is a hard outcome. Mortality has an undisputable importance for the patients and their relatives. Nevertheless, mortality is not necessarily

an optimal QI. This is partly due to patient heterogeneity, heterogeneous causes of death and the rarity of many conditions. But it is also due to the falling baseline mortality rates and the fact that death simply does not occur often enough in some patient groups to secure the necessary frequency an event must occur with to be a meaningful QI[20]. In a Danish study for instance, 30-days mortality for pre-hospital patients varied between 2,3% (Trauma) and 49,3% (Unconsciousness/Cardiac arrest)[21]. For patient groups or systems with very low mortality, it might seem more reasonable to measure the quality of the process of care rather than outcome. In fact, the quality of care for all pre-hospital patients should also be measured by process measures, because outcome and process quality are two different concepts. A patient might receive state-of-the-art care, but still die due to the severity of the disease or trauma. Using only mortality as QI in such a case, will not reflect the high quality in the process of care. Opposite, a patient might receive poor care but still survive. In such a case, using mortality as the only QI will not reflect the low quality in the process of care. The fact that we did not find any significant correlation between 30-days mortality and Total Quality Score, can be an indication of just this; that the measurements of mortality and process quality are complementary, and both are central to identify the total quality achieved in a system.

As shown in figure 3 survival to patient handover varies significantly between 95,5 % and 87,3 % ( $p=0,00$ ). Survival until 30 days after the P-EMS response also varies significantly; between 89,0 % and 76,1 % ( $p=0,00$ ). For both variables, Sweden has the lowest mortality and Finland has the highest mortality in this study. This may be a reflection of different use of the P-EMS units. In Finland, the proportion of inter-hospital transports is lower than in the other nations; 3,1% (F) vs 20,4%(DK), 34,9%(S) and 41,6%(N). These differences in use may be contributing to the different mortality numbers because transferred patients normally are in a more stable phase than the patients cared for in primary responses. Moreover, the combined information from figure 2 and figure 3 illustrates that the Finnish P-EMS bases have the lowest proportion of responses with patient contact and at the same time the highest mortality. A possible explanation for this may be that Finnish P-EMSs manage to select the responses with the most critical ill patients. For both mortality variables in figure 3, the numbers for Denmark and Norway are quite similar to each other, which may reflect a more comparable dispatch practice and patient population between these nations.

## Limitations

Mortality is very much determined by the patient's status. Nonetheless, we have combined all patients in all countries and all services into one group to explore the association between QIs and mortality. This was done to secure a normal clinical setting in P-EMS, as these QIs are developed for everyday quality measurement in international P-EMS regardless of patient characteristics. Hence, this seems to be the adequate setting for our study. However, it might be that subgroup analysis on specific patient groups, for instance high mortality diagnosis, would reveal different correlations between QIs and mortality.

Regarding missing data, "Survival to handover"-data were missing for only 6 out of 2814 patients. However, 30 days mortality-data were missing for 9,7% of the patients. These are either patients with foreign personal identification number or patients with unknown identity. Both patient groups are taken care of regularly by P-EMS. In all four nations we experienced the same difficulties pertaining to these patient groups when collecting 30 days mortality data. The data collection period was partly in the summer months, when the number of foreign tourists in the Nordic countries is high. Thus, the proportion of missing data might at least partly be explained by a relatively high number of foreign citizens treated by P-EMS. We have no reason to believe that the mortality of the mentioned patient

groups differs significantly from the rest of the patient cohort. Thus, we assess the 30 days mortality figures as representative for the patients in the study group, although the missing data for these figures ideally should be lower to secure the most valid results.

## **CONCLUSION:**

In this study we have explored the statistical relationship between quality scores and mortality in P-EMS. Good overall process quality did not seem to correlate with good outcome quality. However, these are complementary measures, both with undeniable value and central to identify the total quality achieved in P-EMS. Thus, an adequate approach to quality measurement in P-EMS seems to be the combination of mortality measures and multiple QIs measuring mainly process quality. Using multiple QIs is imperative to secure a broad and comprehensive assessment of quality.

## **ACKNOWLEDGEMENTS:**

We thank the following physician-staffed emergency medical services for participating in the data collection:

Vantaa HEMS, Turku HEMS, Tampere HEMS, Oulu HEMS, Kuopio HEMS, all Finland.

Skive HEMS, Billund HEMS, Ringsted HEMS, all Denmark.

Uppsala HEMS, Karlstad HEMS, both Sweden.

Lørenskog HEMS, Rygge SAR, Arendal HEMS, Stavanger HEMS, Trondheim HEMS, Ørland SAR, all Norway.

We thank Bjørn Henrik Moshuus, IT Manager at The Norwegian Air Ambulance Foundation, for developing the web-based database.

We thank Päivi Laukkanen-Nevala for statistical support and Jukka Tennilä for IT support at FinnHEMS Research and Development Unit, and Sasu Liuhanen at Absolute Imaginary Ltd for the adaption of the FinnHEMS database.

## **Funding:**

Helge Haugland holds a position as a PhD-student financed by The Norwegian Air Ambulance Foundation. We thank all the donators of The Norwegian Air Ambulance Foundation for the financial support that made this study possible.

## **CONFLICT OF INTEREST:**

The authors declare the following conflict of interests: HH and AK holds research positions in The Norwegian Air Ambulance Foundation, a non-commercial charity owning The Norwegian Air Ambulance Ltd, which is the contractor of the national air ambulance service in Norway.

## **AUTHOR`S CONTRIBUTION:**

HH and AK conceived the idea. All authors were part of the study design. HH, AO, LR and DO organized and administered the data collection in Norway, Finland, Denmark and Sweden, respectively. All authors have approved the final version of the manuscript.

## **LIST OF ABBREVIATIONS:**

EMS: Emergency Medical Service

P-EMS: Physician-staffed Emergency Medical Service

HEMS: Helicopter Emergency Medical Service

SAR: Search and rescue

QI: Quality Indicator

EQUIPE: Establishing Quality Indicators in Physician-staffed Emergency Medical Services

## REFERENCES:

1. Fevang E, Lockey D, Thompson J, Lossius HM, Torpo Research C. The top five research priorities in physician-provided pre-hospital critical care: a consensus report from a European research collaboration. *Scand J Trauma Resusc Emerg Med*. 2011;19:57. doi:10.1186/1757-7241-19-57.
2. Snooks H, Evans A, Wells B, Peconi J, Thomas M, Woollard M et al. What are the highest priorities for research in emergency prehospital care? *Emerg Med J*. 2009;26:549-50. doi:10.1136/emj.2008.065862.
3. Haugland H, Rehn M, Klepstad P, Kruger A, EQUIPE collaboration group. Developing quality indicators for physician-staffed emergency medical services: a consensus process. *Scand J Trauma Resusc Emerg Med*. 2017;25:14. doi:10.1186/s13049-017-0362-4.
4. Palmer RH. Using health outcomes data to compare plans, networks and providers. *Int J Qual Health Care*. 1998;10:477-83.
5. El Sayed MJ. Measuring quality in emergency medical services: a review of clinical performance indicators. *Emergency medicine international*. 2012:161630.
6. Borzecki AM, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. *Med Care*. 2010;48:1117-21. doi:10.1097/MLR.0b013e3181ef9d53.
7. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet*. 2004;363:1147-54. doi:10.1016/S0140-6736(04)15901-1.
8. Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf*. 2012;21:737-45. doi:10.1136/bmjqs-2011-001159.
9. Roudsari BS, Nathens AB, Cameron P, Civil I, Gruen RL, Koepsell TD et al. International comparison of prehospital trauma care systems. *Injury*. 2007;38:993-1000.
10. Horsted TI, Rasmussen LS, Lippert FK, Nielsen SL. Outcome of out-of-hospital cardiac arrest--why do physicians withhold resuscitation attempts? *Resuscitation*. 2004;63:287-93. doi:10.1016/j.resuscitation.2004.05.005.
11. Institute of Medicine. *Emergency Medical Services at a Crossroads*. The National Academies Press, Washington DC, 2006.
12. Hudson P. Applying the lessons of high risk industries to health care. 2003;12:i7-i12. doi:10.1136/qhc.12.suppl\_1.i7 %J Quality and Safety in Health Care.
13. Diley I, Badrinath P, Annon S. Is mortality a good indicator of the clinical quality of National Health Service hospitals? A cross-sectional study of outlier trusts for mortality indices using quality dashboards. *JRSM open*. 2014;5:2054270414533325-. doi:10.1177/2054270414533325.
14. Kruger AJ, Skogvoll E, Castren M, Kurola J, Lossius HM, ScanDoc Phase 1a Study G. Scandinavian pre-hospital physician-manned Emergency Medical Services--same concept across borders? *Resuscitation*. 2010;81:427-33. doi:10.1016/j.resuscitation.2009.12.019.
15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12:1495-9. doi:10.1016/j.ijsu.2014.07.013.
16. Haugland H, Olkinuora A, Rognas L, Ohlen D, Kruger A. Testing quality indicators and proposing benchmarks for physician-staffed emergency medical services: a prospective Nordic multicentre study. *BMJ Open*. 2019;9:e030626. doi:10.1136/bmjopen-2019-030626.
17. Kruger AJ, Lockey D, Kurola J, Di Bartolomeo S, Castren M, Mikkelsen S et al. A consensus-based template for documenting and reporting in physician-staffed pre-hospital services. *Scand J Trauma Resusc Emerg Med*. 2011;19:71. doi:10.1186/1757-7241-19-71.



18. Ken Lee KH, Matthew Austin J, Pronovost PJ. Developing a Measure of Value in Health Care. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2016;19:323-5. doi:10.1016/j.jval.2014.12.009.
19. Gruen RL, Gabbe BJ, Stelfox HT, Cameron PA. Indicators of the quality of trauma care and the performance of trauma systems. *Br J Surg*. 2012;99 Suppl 1:97-104. doi:10.1002/bjs.7754.
20. Gisvold SE, Fasting S. How do we know that we are doing a good job - can we measure the quality of our work? *Best Pract Res Clin Anaesthesiol*. 2011;25:109-22. doi:10.1016/j.bpa.2011.02.010.
21. Bøtker MT, Terkelsen CJ, Sørensen JN, Jepsen SB, Johnsen SP, Christensen EF et al. Long-Term Mortality of Emergency Medical Services Patients. *Annals of emergency medicine*. 2017;70:366-73.e3. doi:10.1016/j.annemergmed.2016.12.017.