TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

DATA COLLECTION IN HELICOPTER EMERGENCY MEDICAL SERVICES

Anssi Heino

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1542 | MEDICA – ODONTOLOGICA | TURKU 2021

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

DATA COLLECTION IN HELICOPTER EMERGENCY MEDICAL SERVICES

Anssi Heino

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1542 | MEDICA – ODONTOLOGICA | TURKU 2021

University of Turku

Faculty of Medicine
Anaesthesiology, Intensive Care, Emergency Care and Pain Medicine
Prehospital Emergency Care
Doctoral Programme in Clinical Research

Supervised by

Miretta Tommila, MD, PhD
Department of Perioperative Services,
Intensive Care Medicine and Pain Management,
Turku University Hospital and
University of Turku,
Turku, Finland

Timo Iirola, MD, PhD
Emergency Medical Services, Turku
University Hospital and
University of Turku,
Turku, Finland

Docent Jouni Nurmi, MD, PhD
Emergency Medicine and Services, Helsinki
University Hospital,
University of Helsinki,
Helsinki, Finland

Reviewed by

Professor Seppo Alahuhta, MD, PhD
Department of Anaesthesiology
Medical Research Center Oulu
Oulu University Hospital and
University of Oulu
Oulu, Finland

Docent Tuomas Brinck, MD, PhD
Department of Orthopedics and
Traumatology
Helsinki University Hospital and
University of Helsinki
Helsinki, Finland

Opponent

Associate professor Jens Aage Kølsen
Petersen, MD, PhD
Department of Anaesthesia and Surgery, section 2
Aarhus University Hospital
Aarhus, Denmark

ISBN 978-951-29-8412-1 (PRINT)
ISBN 978-951-29-8413-8 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama Oy, Turku, Finland 2021

To my family

UNIVERSITY OF TURKU
Faculty of Medicine
Anaesthesiology, Intensive Care, Emergency Care and Pain Medicine
ANSSI HEINO: Data collection in helicopter emergency medical services
Doctoral Dissertation, 126 pp.
Doctoral Programme in Clinical Research
May 2021

ABSTRACT

Prehospital critical care, especially helicopter emergency medical services (HEMS), is a costly but vital part in the chain of survival for a critically deteriorated patient. The quality assessment and outcome measures of this service are important for targeting the limited resources accurately. Clinical registries are a key element of this system follow-up and quality assurance. In addition, they are a vast resource for scientific objectives. Therefore, the data reliability in these clinical registries needs to be assured.

The aims of this thesis were to evaluate the accuracy and reliability of clinical data collection in a national HEMS service. In addition, to study the accuracy of prognostication based on prehospital patient classification and registry data. And finally, to revise a prehospital patient scoring system, the HEMS Benefit Score, to meet the modern standards of prehospital emergency medical services. This scoring is used in all Finnish HEMS units to evaluate the benefit of prehospital emergency medical services for patients treated on HEMS missions.

Inter-rater reliability was evaluated among HEMS clinicians as they registered written mission scenarios into the FinnHEMS database. Furthermore, the accuracy of prognostication was evaluated in a retrospective patient population of 6219 HEMS patients. Finally, a revision for the HEMS Benefit Score was performed with Delphi method.

The overall inter-rater reliability of data collected from the written mission scenarios was on an adequate level, however, vital signs documentation was shown to be poor. In addition, documentation of time-related parameters had a moderate inter-rater reliability. Patient scoring and classification indicated an overall poor inter-rater reliability among study participants. Prognostication in the HEMS setting had a moderate accuracy, and both futile and non-futile patients were treated with similar intensity. The revision of the HEMS Benefit Score resulted in a restructured and modernised version of a scoring for prehospital use, the EMS Benefit Score.

As a conclusion, the reliability and accuracy of data collection among Finnish HEMS clinicians is on an adequate level. The reliability of a prehospitally set futile prognosis is at least questionable, therefore, decisions to limit treatment in a prehospital setting should be made with caution. Delphi method was established as a suitable process for implementation of a prehospital scoring system.

KEYWORDS: clinical registries, prehospital critical care, scoring, classification, system quality

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Anestesiologia, tehohoito, ensihoito ja kivunhoito
ANSSI HEINO: Tiedon keruu ensihoidon helikopteritoiminnassa
Väitöskirja, 126 s.
Turun kliininen tohtoriohjelma
Toukokuu 2021

TIIVISTELMÄ

Lääkärijohtoiset ensihoidon helikopteriyksiköt (HEMS) ovat tärkeä osa kriittisesti sairastuneiden potilaiden hoitojärjestelmää. Jotta rajallisia resursseja voidaan kohdistaa oikealla tavalla, on tärkeää arvioida HEMS-toiminnan laatua ja vaikuttavuutta. Kliiniset laaturekisterit ovat olennainen osa toiminnan laadun arviointia, ja rekisterit toimivat myös tieteellisen tutkimuksen pohjana. Tästä syystä kliinisiin rekistereihin kerätyn tiedon luotettavuus tulee varmistaa.

Tämän väitöskirjan tavoitteena oli tutkia kansallisissa HEMS-yksiköissä toimivien ensihoitolääkärien ja ensihoitajien kirjauskäytäntöjen luotettavuutta ja yhtenäisyyttä. Lisäksi tutkittiin potilaiden luokittelun ja ennustearvion osuvuutta HEMS-tehtävissä. Väitöskirjan viimeisenä osaprojektina päivitettiin kansallisten HEMS-yksiköiden käyttämä pisteytysjärjestelmä, HEMS Benefit Score, vastaamaan nykyaikaisia ensihoidon käytäntöjä. HEMS Benefit Score on ensihoidon yksittäiselle potilaalle tuottamaa hyötyä arvioiva pisteytysjärjestelmä, joka on käytössä kaikissa suomalaisissa HEMS-yksiköissä.

Kirjausten luotettavuutta tutkittiin kuvitteellisten ensihoidon tehtävien avulla. Luotettavuutta arvioitiin erikseen sekä tehtäväkohtaisten muuttujien että potilasluokitus- ja pisteytysjärjestelmien osalta. Väitöskirjan kolmannessa osatyössä tutkittiin ennustearvion luotettavuutta 6219 potilaan retrospektiivisessä tutkimusasetelmassa. Viimeisessä osatyössä HEMS Benefit Score päivitettiin Delphi-menetelmää käyttäen.

Tulosten perusteella kirjaamisen luotettavuus oli kaiken kaikkiaan kohtalaisella, mutta peruselintoimintojen kirjaamisen osalta huonolla tasolla. Väitöskirjassa tutkittujen pisteytysjärjestelmien luotettavuus osoitettiin olevan vaihtelevaa vastaajien välillä. Ennustearvion teko onnistui kohtalaisen luotettavasti, ja sekä toivottomaksi arvioituja että todennäköisesti selviytyviksi arvioituja potilaita hoidettiin yhtä intensiivisesti. Väitöskirjan tulosten perusteella ensihoidossa asetetun toivottoman ennusteen osuvuus ei ole merkittävän korkea, joten päätöksiin rajoittaa hoitoa jo ensihoitotilanteessa tulisi suhtautua varovaisuudella. HEMS Benefit Score päivitettiin Delphi-menetelmällä vastaamaan nykyaikaisia hoitokäytäntöjä, ja nimettiin uudelleen EMS Benefit Scoreksi.

AVAINSANAT: potilasrekisterit, ensihoito, pisteytysjärjestelmät, potilasluokittelu, laatujärjestelmät

Table of Contents

Abbreviations

ALS	Advanced Life Support
AMPT	Air Medical Prehospital Triage Score
APACHE	Acute Physiology and Chronic Health Evaluation
ASA-PS	American Society of Anesthesiologist Physical Status
AVPU	Alert, Verbal, Pain, None
BLS	Basic Life Support
COPD	Chronic Obstructive Pulmonary Disease
CPR	Cardio-Pulmonary Resuscitation
CRB-65	Confusion, Respiratory Rate, Blood Pressure, Age
CURB-65	Confusion, Urea, Respiratory Rate, Blood Pressure, Age
CVI	Content Validity Index
DNAR	Do-Not Attempt Resuscitation
ECOG	Eastern Co-Operative Oncology Group
EEG	Electroencephalography
EMS	Emergency Medical Services
EMT	Emergency Medical Technician
EMT-B	Emergency Medical Technician - Basic
EMT-I	Emergency Medical Technician - Intermediate
EMT-P	Emergency Medical Technician - Paramedic
EUPHOREA	The European Prehospital Research Alliance
GCS	Glasgow Coma Scale
GP	Glykoprotein
HBS	HEMS Benefit Score
HCO3	Bicarbonate
HEMS	Helicopter Emergency Medical Services
HR	Heart Rate
HUS	Helsinki University Hospital
ICD-10	International Classification of Diseases Tenth Edition
ICP	Intracranial Pressure
ICPC-2	International Classification of Primary Care Second Edition
ISS	Injury Severity Score

MAP	Mean Arterial Pressure
MEDS	Mortality in Emergency Department Score
MODS	Multiple Organ Dysfunction Score
MEWS	Modified Early Warning Score
MPM	Mortality Probability Program
NACA	The National Advisory Committee of Aeronautics
NAS	Nursing Activities Score
NEMS	Nine Equivalents of Nursing Manpower Use Score
NEWS	National Early Warning Score
OHCA	Out of Hospital Cardiac Arrest
PCI	Percutaneous Coronary Intervention
PEA	Pulseless Electrical Activity
PIRO	Predisposition, Insult, Response, Organ Dysfunction
ROSC	Return of Spontaneous Circulation
SAPS	Simplified Acute Physiology Score
SCC	Simple Clinical Score
SD	Standard Deviation
SEWS	Standardised Early Warning Score
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Assessment
SpO_2	Blood Oxygen Saturation
TISS	The Therapeutic Intervention Scoring System
UK	United Kingdom
US	United States
VF	Ventricular Fibrillation
WHO	World Health Organization
ViEWS	Vitalpac Early Warning Score
WONCA	The World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians.

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Heino A, Iirola T, Raatiniemi L, Nurmi J, Olkinuora A, Laukkanen-Nevala P, Virkkunen I, Tommila M. The reliability and accuracy of operational system data in a nationwide helicopter emergency medical services mission database. BMC Emergency Medicine. 2019 Oct 15; 19 (1):53.

II Heino A, Laukkanen-Nevala P, Raatiniemi L, Tommila M, Nurmi J, Olkinuora A, Virkkunen I, Iirola T. Reliability of prehospital patient classification in helicopter emergency medical service missions. BMC Emergency Medicine. 2020 May 25; 20 (1):42.

III Heino A, Björkman J, Tommila M, Iirola T, Jäntti H, Nurmi J. Accuracy of prognostication in prehospital settings: a nationwide retrospective cohort study on helicopter emergency service patients. Manuscript

IV Heino A, Raatiniemi L, Iirola T, Meriläinen M, Liisanantti J, Tommila M. A score to measure the benefits of prehospital emergency medical services – A European Delphi study. Manuscript

1 Introduction

During the last decades, prehospital emergency medical services have evolved from a pure transportation service into an essential part of patient medical care already at the scene of the incident (Tanigawa et al. 2006, Roudsari et al. 2007, Bigham et al. 2015). Modern prehospital medical services can provide not only basic level medical care but also high level critical care like in-hospital intensive care units. One part of modern prehospital care is physician-staffed units, especially helicopter emergency medical service (HEMS) units, which nowadays are found in most high-income countries and their services. Because of the nature of this service, the cost-effectiveness and true benefit of prehospitally performed critical care is under ongoing debate, as is the value of HEMS units (Cairns et al. 1998, Van Schuppen et al. 2011, Rehn et al. 2014). Therefore, the importance of studies that focus on the quality, the performance and the outcome of prehospital emergency medical care is evident.

Clinical registries are a crucial part of quality assessment and follow-up in all medical care (Dreyer et al. 2009). As in other areas of medical care, also in prehospital care, recommendations to collect operational and patient data are found (Kruger et al. 2011). When quality of care is evaluated, this data collection is considered as the main instrument and provides a source for scientific purposes (Scmidt et al 2015, Hoque et al. 2018, Lysholm et al. 2019). However, the registry data itself needs to be reliable and coherent before it can be used for system development or scientific purposes (Pollock et al. 1995, Olthof DC et al. 2013, Coi A et al. 2016). In addition, the single parameters that are documented should provide clear value if used, and unnecessary data collection is merely a waste of limited resources. There is variation in the existing literature, and regarding single clinical registries, in terms of data reliability. In addition, there is a lack of clear indicators which should be followed when clinical registry reliability is evaluated (Arts et al. 2002).

Scoring and classification of patients is one aspect of follow-up in medical care. Furthermore, different patient scoring systems are used to predict outcome, to characterise disease severity and degree of organ dysfunction, and to evaluate resource use (Vincent et al. 2010, Ringdal et al. 2013, De Grooth et al. 2018, Jouffroy

et al. 2019). This scoring data is one aspect of system quality improvement also in prehospital care as patient characteristics are followed and system outcome measured (Kruger et al. 2011, Norwegian directorate of health 2018). However, when prehospital scoring and classification is performed, it should be considered that most of the prehospitally used scores and classifications are not originally built for prehospital use, and even further, no structured prehospital implementation exists in most of the scores used (Sankar et al. 2014, Ringdal et al. 2013). In the question of a single scoring system, the HEMS Benefit Score (HBS) was originally built for prehospital use, but only one reliability study existed in the literature concerning the use of the HBS prior to this thesis, and there were no studies on the predictive value of the HBS (Raatiniemi et al. 2017). This score aims to evaluate the benefit of the whole prehospital emergency medical services for a single patient. Essentially, the intrinsic benefit assessment is a vital part of system quality advancement.

To direct the limited and costly resources in prehospital critical care to patients who can truly benefit from it should be the focus of prehospital system evolution (Persad et al. 2009). Efforts to detect futile patients reliably in the prehospital phase is a crucial aspect of this (Ferrand et al. 2006, Kangasniemi et al. 2019). A directive to limit treatment is a heavy decision, with a strong risk of becoming a self-fulfilling prophecy if set based on limited knowledge of patient background and acute condition. Therefore, these decisions should be made with great caution. To target the resources in most practical manner, the prehospital critical care needs to be evaluated constantly. This can be performed by data collected in the prehospital clinical registries. However, the data quality in the registries has to be secured before it can be used for system development.

2 Review of the Literature

2.1 Prehospital emergency medical care

2.1.1 Prehospital Emergency Medical Services

Prehospital emergency medical services (EMS) consist of clinical examination, assessment of treatment need, treatment on the scene, and transportation of an injured or medically deteriorated patient. A marked international variation exists on the implementation of EMS systems (MacFarlane et al. 2005, Tanigawa et al. 2006, Vaitkaitis 2008, Sun et al. 2017). There are differences between developed and developing countries, but also among high income countries, on how these services are provided (Thomson 2005, Roudsari et. al 2007). Most often, the EMS system constitutes a nationally, or regionally, coordinated dispatch centre that receives emergency calls and dispatches the necessary EMS units. However, especially in low-income countries, there may be a lack of a coordinated emergency call and dispatch systems.

Most EMS systems rely on ambulance ground units, supported with rapid response cars, advanced critical care ground units, and helicopter emergency medical service (HEMS). In addition, some systems have motorcycle EMS units, or even fixed-wing aircrafts for certain prehospital missions (Evans et al. 2014). The greatest international variation is found in EMS staff training and the level of medical care they administer (Roudsari et al. 2007, Sun et al. 2017). EMS systems can rely on units administering basic life support (BLS) or advanced life support (ALS). In some systems and ambulances that focus solely on transportation and do not perform any treatments or medical interventions, there are also volunteer first responders.

Most western EMS systems have different level units, which operate based on their staff training and certifications for procedures and medications. For example, the United Kingdom (UK) EMS is based on ambulances that are staffed with paramedic and technician, in addition with solo responder and rapid response units, and ambulance officer units (Black et al. 2005). Depending on the training level and certification of UK EMS unit in question, these can perform tasks from basic level resuscitation and drug administration to tracheal intubation and administration of a wide range of critical care drugs. United States (US) has national variation in EMS

systems, but the staff operating in US EMS are first responders and ambulance emergency medical technicians (EMTs) (Pozner et al. 2004, Van Gelder et al. 2005). The EMTs are set on three levels, depending on their training: EMT-basic (EMT-B), EMT-intermediate (EMT-I), and EMT-paramedic (EMT-P). Of these, EMT-B level personnel can administer oxygen, extricate patients, and transfer patients, whereas EMT-P level can perform advance procedures like tracheal intubation, needle thoracostomy, and intravenous access and medication administration.

Despite the variance in EMS systems and how they are formed, there are few studies on the superiority of diverse systems (Nichol et al. 2008, Shin et al. 2012, Okubo et al. 2018). There is evidence on the beneficial aspect of staff training and geographical density of EMS units on patient outcome. In addition, cost-effectiveness studies on single systems do exist, again, providing data on the benefit of provider resources on patient outcome and system effectiveness (Taylor et al. 2012). However, there is a lack of studies comparing effectiveness and advantage among international EMS systems, especially between developed countries where EMS personnel are highly trained and resources are similar.

2.1.2 Prehospital Emergency Medical Services in Finland

In Finland, hospital districts provide emergency medical services in their areas together with joint municipal authorities. Emergency medical services are planned and implemented in cooperation with units providing emergency medical care to form a regionally coherent system. This can be produced by in-house personnel, in cooperation with the region's rescue services, by joint municipal authorities for other hospital districts, or by outsourcing the services to other service providers (Ministry of Social Affairs and Health, Finland, 2017). The services constitute six nationally coordinated emergency call centres, which receive public emergency calls and dispatch EMS units (Raatiniemi et al. 2015, Pappinen et al. 2018, Aitavaara-Anttila et al. 2020). Finnish EMS is formed by first responders, basic and advanced level ground ambulances, EMS supervisors, added with six nationally organised HEMS units and physician-staffed ground units. First responders can be volunteer laypersons trained for first aid, basic CPR, and use of automatic defibrillators, or fire department units dispatched on time and distance basis. Basic level ambulances perform basic life support, and some of them can administer intravenous resuscitation medication, whereas advanced ambulances are trained and licensed to a wider range of critical care medication and procedures, such as airway management and needle thoracostomy. In Finland, the basic training of bachelor level nurse-paramedics operating in ambulances is normally at least four and a half-year training in University of Applied Sciences (240 ECTS, European Credit Transfer and Accumulation System). (Länkimäki et al. 2015).

The actual dispatch criteria for medical incidents are based on a national risk assessment protocol which includes 40 medical keywords. These keywords are added with predetermined priority questions to guide the dispatcher on mission urgency. Finnish EMS systems have four level priority classes: A, B, C, and D; A being the highest priority. The class is set by the dispatch centre, and this class defines the dispatch criteria as well as the number and level of EMS units dispatched (Hoikka et al. 2016). In Finland, dispatch centres organise responses to all medical, police, fire and rescue calls. The people operating in the dispatch centres are laypersons who have undergone a national 18-month training programme.

As a part of emergency medical service, the nationally coordinated and financed HEMS begun in Finland in 2011. Before this, there were HEMS units provided by trusts and funded by donations. Five of the HEMS units cover the geographical areas of each five university districts, added with one HEMS unit covering the northern Lapland area of Finland (Raatiniemi et al. 2017, Saviluoto et al. 2020). Finnish HEMS bases and their geographical locations are presented on Figure 1. The HEMS unit operating in the Lapland area is staffed with two advanced paramedics and two pilots, while the other five units are staffed with prehospital critical care physician together with paramedic or firefighter, and a helicopter pilot. All HEMS units operate 24 hours per day, 365 days per year. In addition to a helicopter, all six HEMS crews can operate with a rapid response car, depending on potential time and distance advantages. The HEMS dispatch is done according to nationally set guidelines and criteria in emergency call centres, but EMS ambulances can also request HEMS assistance on scene. The current HEMS system covers most of the geographical areas, excluding some of the south-eastern and western areas of Finland (Pappinen et al. 2019). Typical HEMS missions include out-of-hospital cardiac arrests, major trauma, and comatose patients. Most often, HEMS units operate within the university hospital district area where they are located, but they can also be dispatched to missions located outside their primary area and on another HEMS unit's district. The Finnish HEMS system is similar to other Scandinavian countries, as it is mainly staffed by senior anesthesiologists (Krüger et al. 2010). However, Finnish HEMS seldom takes part in search and rescue missions, or in inter-hospital or neonatal transportation.

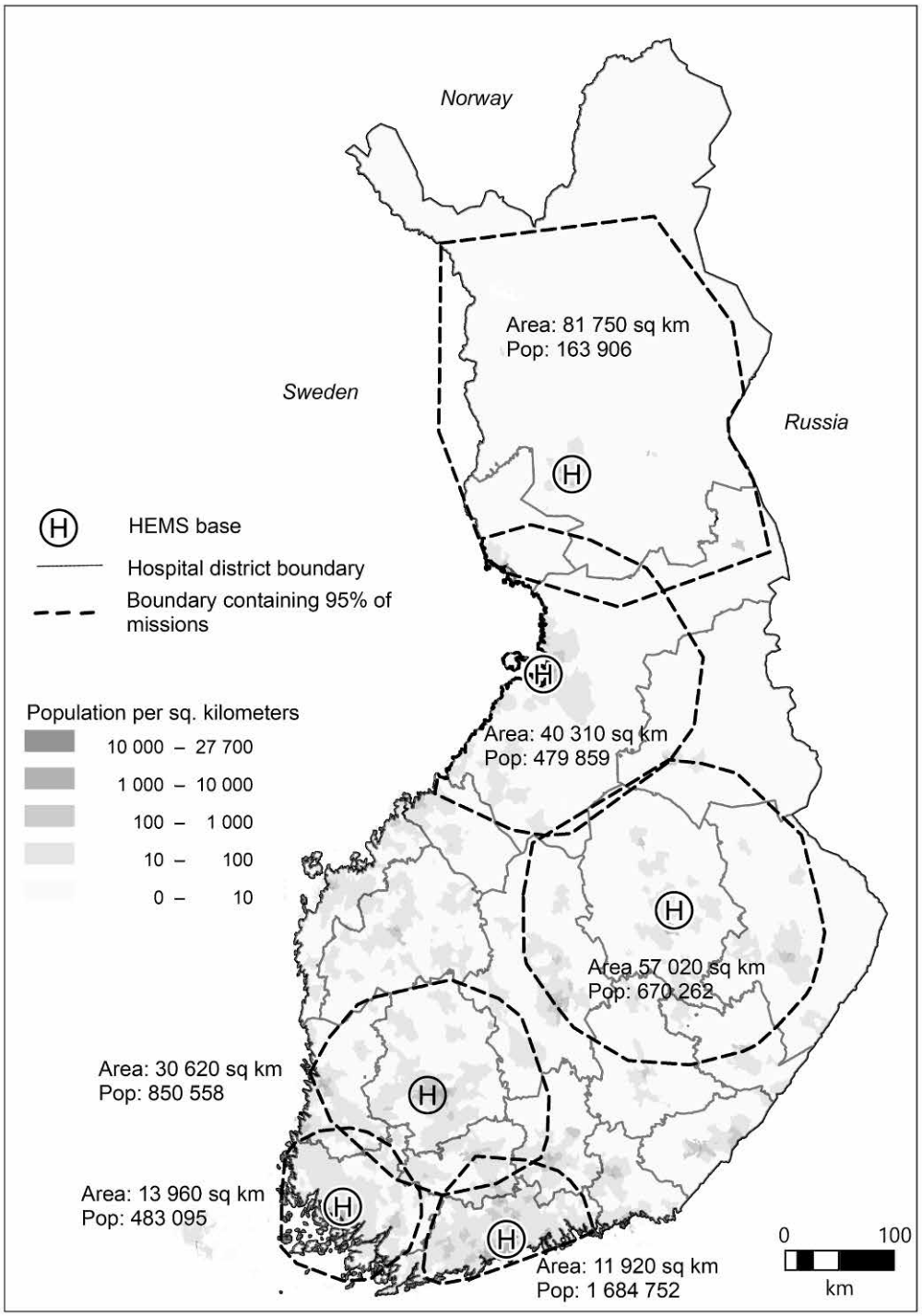

Figure 1. National helicopter emergency medical services base locations in Finland (Pappinen et al. 2019, Saviluoto et al. 2020)

2.1.3 Role of prehospital physicians

Prehospital physicians are a part of modern emergency medical service in most western prehospital systems. However, the true influence of a prehospitally operating physician is still under debate (Van Schuppen et al. 2011, Risgaard et al. 2020). Like physicians working in critical care inside hospital walls, prehospital physicians are also used for treating the most critically injured or deteriorated patients. In addition, they provide consultation for paramedics, take medical leadership in multi-patient operations, and operate as highest medical experts in the prehospital emergency medical service (Van Schuppen et al. 2015, Mikkelsen et al. 2017, Friberg et al. 2018). The undisputed benefit of a physician providing this expertise, compared to highly trained advanced paramedics, is yet to be solved (Schewe et al. 2019, Pakkanen et al. 2019). However, there is evidence with certain patient groups and interventions that prehospital physicians provide higher quality care than advanced paramedics alone, for example treatment of traumatic brain injuries or prehospital intubation (Pakkanen et al. 2017*).* The level of prehospital physicians' training varies between countries and systems. In Scandinavia, prehospital physicians are most often senior anesthesiologists. Whereas in central Europe, as in Germany or Austria, a prehospital physician can be a general practitioner with some minor additional training for prehospital critical care (Fullerton et al. 2011, Trimmel et al. 2017).

2.1.4 Impact of prehospital care

Modern prehospital care aims to deliver the emergency medical service to patient homes or scenes of the incidents, opposite to decades ago, when ambulances were solely modes of transportation for medically deteriorated patients (Bigham et al. 2015*).* The protocols and intervention capabilities in the prehospital phase are the same as, or similar to, those delivered in-hospital. There is evidence on the benefit of taking these interventions to the scene of the incident; but at the same time, there is evidence on the benefit of avoiding unnecessary time in the prehospital phase, when life-saving interventions are performed in-hospital (McClean et al. 2002, Cone et al 2007, Rehn et al. 2014). The balance between "stay and play" and "scoop and run" is essential when the impact of prehospital care is discussed (Smith et al. 2009). The dilemma between these two is still under continuous discussion, and studies supporting both lines of duty are being released constantly (Van der Velden et al. 2008, Harmsen et al. 2015, Mills et al. 2019). Most importantly, when the impact of prehospital care is evaluated, the patient characteristics and severity of medical condition present a major aspect together with the features and resources of the system under observation. There is clear evidence on the benefit of a single intervention in restricted patient groups, such as early defibrillation in prehospital

cardiac arrest (Koenraad et al. 2015). However, even in this example, the beneficial intervention can be performed by a layperson, not necessarily by a prehospital clinician or expert. The true impact of overall prehospital care on patient outcomes still remains unsolved.

2.1.5 Prognostication and limiting treatment in prehospital setting

Prognostication and decisions to limit treatment of patients who are at the end-of-life stage of their illnesses, or acutely and incurably deteriorated, are important parts of all intensive and critical care (Persad et al. 2009). To avoid prolonging of care in futile patients is ethical and a fundamental aspect of palliative care. A senior physician may limit any medical treatment considered futile, and the patient has the right to refuse treatments offered, even if they are demanded by the patient relatives (Ministry of Social Affairs and Health, Finland 1992, Kangasniemi et al. 2019). The patient can create an advance directive to limit and guide their medical care. These directives are stated in the patient's medical records. The most common advance directive is do-not attempt resuscitation (DNAR). Additionally, the limitations can regard intensive care, intubation, mechanical ventilation, invasive procedures, intravenous antibiotics, transferring the patient to a certain level hospital, and feeding or hydrating the patient. These decisions are often made in the hospital, or in primary care; thus, these decisions are also made in a prehospital setting (Kangasniemi et al. 2019). However, there is little research data on or scientific basis for this decision making at the prehospital phase of the care. Existing studies do present that these decisions are made already in prehospital critical care and are an important aspect of physician-staffed prehospital emergency medical service (Ferrand et al. 2006). A recent study on Finnish HEMS physicians presented that over 80% of the responders have set limitations to treatment in their duty (Kangasniemi et. al 2019). In addition, this study revealed that over 80% of HEMS physicians consider limiting treatment as a prehospital physician's obligation. Characteristics most often related to these decisions were severe comorbidities, disabled patients, nursing home or healthcare facility patients, and aged patients. Still, little is known on the adequacy of this decision making already in prehospital phase – especially, when there is often limited knowledge of the patient's underlying or pre-existing illnesses, performance status, living will, or the will of the relatives.

2.2 Clinical registries

2.2.1 Definition of clinical registries

Clinical registry is a collection of information and data on patients' health statuses, particular diseases, the healthcare that they receive as well as the outcomes of treatments; or it can be focused on a definite segment of these (Monash Clinical Registries portfolio 2018). Clinical registries are recommended and used in all fields of medicine (Framework for Australian clinical quality registries, 2014; Dreyer et al. 2009). The data can be collected for system development and quality control purposes, but also for scientific purposes.

There are national recommendations on how clinical registries are built and what variables to follow (Evans et al. 2011, Schmidt et al. 2015). For example, the Danish National Patient Registry follows administrative data, patient diagnoses, the treatments and examinations. In Australia, there are 10 strategic principles listed by the Health Minister that should be considered in the development of clinical quality registries (Framework for Australian clinical quality registries, 2014). These principles focus on regular reporting based on registered data, the infrastructure of the registry itself, the aim to improve healthcare with clinical registries, quality of the collected data, and safe collection and holding of the data.

The number of registries varies markedly depending on the type of registry and the country in which it is surveyed. Even nationally, there can be several registries for a single disease or patient group. In the case of United States, a total of 153 national clinical registries was identified in a literature review. Of these registries, 20 focused on cardiothoracic or cardiovascular data registration, and oncology had 18 different registries (Lyu et al. 2016). In the case of Sweden, 69 national quality registries were found in a recent study (Lysholm et al. 2019). In a literature review and international screening of registries, a total of 18 clinical registries were found when studying a single patient group and using colorectal cancer registries as an example (MacCallum et al. 2018). These examples reveal that a vast number of different types of clinical registries exist.

2.2.2 Scientific use and quality control with clinical registries

Development of clinical registries is often justified with quality control purposes and scientific use. Registry-based studies have increased their significance beside randomised controlled trials (Dreyer et al. 2009, Schmidt et al. 2015, Hoque et. al 2018, Lysholm et al. 2019). Data is collected in large quantities, or even multi-nationally, in the registries. In addition, registry-based studies enable reduced costs and methodological progress towards registry-based randomised controlled trials

(Armstrong et al. 2020). This constitutes an exceptional basis for scientific purposes and large population level outcome studies, as opposed to smaller scale experimental efficacy studies in a traditional randomised controlled setting (Mulder et al. 2019). For example, the clinical registries can be used in pharmaceutical safety, with plausible adverse events registered and studied; trauma registries can be used for trauma outcome and mortality studies; oncology registries enable large scale comparison studies between the effectiveness of different treatments; and data on rare diseases can be gathered in a multinational setting (Maret-Ouda et al. 2017). However, the quality and reliability of the registered data itself, concerning for example handling, storing, and completeness of data, remain major concerns in clinical registry studies.

In terms of quality control, the clinical registries are used especially in the drug, device, procedure, or treatment safety evaluation and follow-up (Mulder et al. 2019). Evident examples of these would be orthopedic registries that follow the outcome and plausible adverse events of joint prostheses, or pharmaceutical registries that follow adverse effects. However, the actual impact of clinical registries on healthcare outcomes is seldom evaluated (Hoque et al. 2017). Compared to the obvious benefit of data collection for scientific use, there is less evidence on quality improvement with use of clinical registries (Porter 2010, Hartmann-Johnsen et al. 2019, Lee et al. 2019). There seems to be a correlation with use of clinical registries and quality development in long-term follow-up, but a lack of clear evidence exists.

2.2.3 Data reliability in clinical registries

To achieve good quality data for scientific and quality control use, the registered data itself needs to be reliable and validated. This is an unambiguous statement, yet defined instruments and methods for quality control of the registry itself are not indisputably described (Arts et al. 2002). Reliability studies are performed for single clinical registries, but the follow-up or outcome on further data quality is not described (Pollock et al. 1995, Olthof et al. 2013, Coi et al. 2016, Horton et al. 2017*)*. In existing literature, there is a lack of wide international consensus on which quality indicators or markers should be used when quality of registries is evaluated. Most often used registry data reliability indicators include completeness of the registered data, correctness of the registered variables, and consistency of the data (Arts et al. 2002, Porgo et al. 2016). A Dutch study represented a clinical auditing process for registry data verification (Van der Werf et al. 2019). This process was based on external data managers, who followed the completeness and accuracy of data in clinical registries. In addition, the sign-up, sample size, and process of verification were presented. The study revealed that some registries had incomplete data, especially on severe complications or even deaths. The study speculated that this

could be due to a potentially negative effect of these incidents on a hospital's appearance, which would explain the incompleteness of the data. The accuracy of data was also addressed, and the importance of clear instructions for multi-interpretable variables underlined. A study on clinical surgical registries highlighted three points to achieve good quality data from clinical registries: auditing, validation, and active data follow-up (Stey et al. 2015); which all should be conducted by trained personnel designated for the task. As a conclusion, the data in a clinical registry needs to be complete in terms of each registered variable. Moreover, the data needs to be correct within tolerable range and consistent. The data quality should be secured by regular follow-up and auditing processes, preferably by personnel specifically designated for this.

2.2.4 Prehospital clinical registries

In recent years, some consensus reports have defined protocols on how clinical registries should be used in prehospital care, and what these prehospital clinical registries should include. In 2011, a European expert panel presented variables that should be documented and reported in physician-staffed prehospital services (Kruger et al. 2011). These variables were composed by sixteen prehospital experts working with the nominal group technique. Variables presented were divided in five groups: fixed system variables, event operational descriptors, patient descriptors, process mapping, and quality indicators and mission outcome. These five main groups were divided into 45 different mission- and patient-related variables that were recommended to be registered. A similar group technique was later used in 2017 to define specific quality indicators that should be followed and registered in physician-staffed prehospital services (Haugland et al. 2017). As a result, a total number of 26 quality indicators were presented, and these were divided into two groups: response-specific quality indicators and system-specific quality indicators. These two consensus processes focused on overall documentation in physician-staffed prehospital service. Recent literature and recommendations related to prehospital data collection are presented in Table 1.

Table 1. Recent literature and recommendations related to prehospital data collection.

Publication	Aim	Setting	Outcome	Clinical implementation
Ringdal et al. 2008	Development and update of Utstein Trauma template	Nominal group technique	31 variables	Trauma documentation
Ringdal et al. 2011	International use of Trauma template	Completeness and coherency of documentation	36 variables	Trauma documentation
Kruger et al. 2011	Development of a template for physician-staffed prehospital services documentation	Nominal group technique	45 variables	Prehospital documentation
Fattah et al. 2014	Literature review on prehospital major incident reporting	Systematic literature review	10 templates	None of the templates were tested for feasibility in real-life incidents
Murphy et al. 2016	Development of key performance indicators	Systematic literature review and Delphi consensus process	101 key performance indicators	Performance monitoring of prehospital emergency care in Ireland
Haugland et al. 2017	Development of quality indicators	Nominal group technique	26 quality indicators	International quality measurement in physician staffed emergency medical services
Sunde et al. 2018	Template for prehospital airway management reporting	Nominal group technique	38 variables	Prehospital airway management documentation

As early as 2008, an Utstein style template for in-hospital major trauma documentation was presented; of the 31 variables presented, a marked number of the documentation was related to the prehospital phase (Ringdal et al. 2008). In addition, an Utstein style template for prehospital airway management was released in 2009 (Sollid et al. 2009), and later updated in 2018 (Sunde et al. 2018). In 2013, a literature review on prehospital major incident reporting revealed that there are several templates for major incident reporting (Fattah et al. 2013). In 2016, a group process study presented 101 key performance indicators that should be followed and registered in overall prehospital emergency care, not just physician-staffed service (Murphy et al. 2016). These constituted 7 structure, 74 process, and 20 outcome indicators. However, despite all these templates for reporting prehospital incidents, there is a lack of follow-up studies presenting the actual real-life clinical use of these. No studies exist that present how widely these are implemented in EMS services, if at all.

A Scandinavian project on prehospital data collection and benchmarking was launched in 2014 and continued until 2018 (Norwegian directorate of health, 2018). This process revealed that a nationwide electronic prehospital emergency service registry is found only in Denmark. The current prehospital registries do not cover nationwide prehospital systems in other Scandinavian countries, even though they have registries covering some parts of their EMS service. The Scandinavian benchmark process recommended following existing guidelines on variables and quality indicators in nationwide prehospital clinical registries. It is the only benchmarking process on prehospital data collection described in the existing literature.

As presented earlier, no large-scale evaluation on prehospital clinical registries or data collection is found in the existing literature. However, studies on single parameters or templates do exist (Nishiyama et al. 2014). In addition, projects are launched to achieve multinational data collection also in prehospital emergency care, but again, only in single incidents or parameters like major trauma (Ringdal et al. 2011, Fattah et al. 2014). On a small scale and on a single nation level, a recent study was performed on the reliability of the Danish nationwide helicopter emergency services database, which focused on the data completeness of 26 above-listed quality indicators (Alstrup et al. 2019).

Like all clinical registries, prehospital clinical registries enable large data collection. Double-blinded randomized controlled trials are seldom possible in the prehospital setting due to the unpredictable nature of prehospital missions. Therefore registry-based studies are presented in definite incidents, such as mortality in a prehospital setting (Christensen et al. 2017), comparison of prehospital emergency care services among Scandinavia (Kruger et al. 2013), on-scene times in helicopter emergency service (Østerås et al. 2017), or in a single patient subgroup like paediatric drowning victims (Garner et al. 2015). In the most recent studies, a Japanese out-of-hospital cardiac arrest (OHCA) registry was used to evaluate the duration and location of pediatric OHCAs (Shida et al. 2019); and in another study setting, the same registry was used to study the benefits of a physician-staffed unit with OHCA following blunt trauma (Fukuda et al. 2018). In a Dutch study, prehospital and in-hospital trauma registry data was used to evaluate prehospital traumatic brain injury care in HEMS missions (Bossers et al. 2019).

System quality control is a widely used argument for prehospital clinical registries (Kruger et al. 2011, Haugland et al. 2017, Howard et al. 2018, Mowafi et al. 2019). Still, there is a lack of evidence of the true influence of these registries on system quality and patient outcome. The disadvantage is that no follow-up studies exist on the system quality changes after launching clinical registries, and as in all clinical registries, the quality indicators to follow are not clear or widely implemented in clinical use. A recent study on the Danish HEMS database indicates

that registered data is used to monitor, assess, and improve the quality of clinical care (Alstrup et al. 2019). The same arguments are used on the Finnish national HEMS database, the FinnHEMS database (Saviluoto et al. 2020). The collected data on mission numbers, the number of patient contacts, and the number of cancelled missions is used for intrinsic quality control in terms of dispatch criteria in the Finnish HEMS. However, there are no peer-reviewed and published studies on the true impact of prehospital clinical registries on service quality improvement and patient outcomes.

2.3 Patient classification and scoring in critical care

2.3.1 Different types of patient classification and scores

Detecting deteriorating patients is essential especially in critical care, but also in all medical care. There are several patient scores and classification systems used to identify these patients (Nannan et al. 2017, Haniffa et al. 2018). Listed examples of scores used in critical care are presented in Table 2. The scores are used for outcome prediction purposes, detection of patients in risk of high mortality, and detection of patients in risk for adverse events (Ringdal et al. 2013, De Grooth et al. 2018, Jouffroy et al. 2019). Depending on the score in question, it can rely on vital parameters, underlying diseases or acute medical condition, or a combination of these. In critical care, the scores are divided in three categories: scores that predict outcome, scores that characterise disease severity and degree of organ dysfunction, and scores that assess resource use (Vincent et al. 2010).

The Acute Physiology and Chronic Health Evaluation (APACHE) score evaluates the degree of acute illness together with chronic health status to predict patient outcome (Knaus et al. 1981, Salluh et al. 2014). APACHE includes 12 physiological variables and is considered the most widely used critical care scoring system. Another scoring system predicting the risk of death in critical care is Simplified Acute Physiology Score (SAPS) (Le Gall et al. 1984). SAPS is divided in three sub-scores, which focus on pre-admission health status, on the actual incident, and on the physiological derangement within the first hour after admission to critical care. In addition to APACHE and SAPS, a Mortality Probability Program (MPM) is a scoring system used for outcome prediction, and this system is based on admission variables complemented with 24-hour follow-up variables (Lemeshow et al. 1987). Nowadays all three scores: APACHE, SAPS, and MPM, are used in their third or fourth updated and revised versions.

Opposite to outcome prediction scores, organ dysfunction scores are used to describe the severity of organ dysfunction. The two most common scores in this category are Sequential Organ Failure Assessment (SOFA) and Multiple Organ

Dysfunction Score (MODS) (Marshall et al. 1995, Vincent et al. 1996). SOFA focuses on six different organ systems, which are respiratory, cardiovascular, renal, hepatic, central nervous, and coagulation. These six variables are scored from 0 (normal) to 4 (most abnormal function) (Lambden et al. 2019). In MODS, there is a seventh variable, gastrointestinal, which is added to the variables included in the score (Aarvold et al. 2017). Both SOFA and MODS are followed in 24-hour periods during the whole critical care phase of the patient.

The third category includes scores for resource use assessment in critical care. These scores evaluate the adequate staffing in intensive care, and mainly focus on nurse – patient ratio. The Therapeutic Intervention Scoring System (TISS) measures nursing workload and evaluates severity of illness to patient care; Nine Equivalents of Nursing Manpower Use Score (NEMS) is a simplified version of TISS; and Nursing Activities Score (NAS) is an extended version of TISS (Hoogendoorn et al. 2020). List of other scores used in critical care are presented on Table 2.

Patient scoring and classification are also used in other areas than critical care. One of the best-known classifications in pre-operative stage is American Society of Anesthesiologist Physical Status (ASA-PS), which is used to assess a patient's perioperative physical status based on prior medical status of the patient (American Society of Anesthesiologists 1941, Ihejirika et al. 2015). ASA-PS was designed to determine and evaluate the operative risk for a patient in surgery. Scoring and classification are not only used for risk evaluation, but also to categorise patients in larger populations and cohorts for documentation purposes. International Classification of Diseases tenth edition (World Health Organization ICD-10) is an internationally used classification system for diagnostic coding, and it has an extension for procedural coding (WHO 1990, Epstein et al. 2019). International Classification of Primary Care second edition (ICPC-2) is a classification system developed for primary care use, and it is used for diagnostic and symptom documentation (WONCA 1998, Basilio et al 2016).

Table 2. Examples of scores used in critical care (Nannan et al. 2017, Haniffa et al. 2018)

SCORE	GROUND PARAMETERS	PURPOSE
MEWS Modified Early Warning Score	Pulse, respiratory rate, temperature, urinary output, blood pressure, AVPU	To detect a deteriorating patient
SEWS Standardised Early Warning Score	Pulse, respiratory rate, temperature, blood pressure, SpO_2, AVPU	To detect a deteriorating patient
NEWS National Early Warning score	Pulse, respiratory rate, temperature, blood pressure, SpO_2, oxygen supplemental, AVPU	To detect a deteriorating patient
VIEWS Vitalpac Early Warning score	Pulse, respiratory rate, temperature, blood pressure, SpO_2, oxygen supplemental, AVPU	To detect a deteriorating patient
CURB-65 Confusion, Urea, Respiratory Rate, Blood pressure, Age	Mental status, urea, respiratory rate, blood pressure, age (> 65)	Mortality prediction in pneumonia
CRB-65 Confusion, Respiratory rate, Blood pressure, Age	Mental status, respiratory rate, blood pressure, age (> 65)	Mortality prediction in pneumonia
PIRO Predisposition, Insult, Response, Organ dysfunction	Combination of comorbidity, laboratory results, current physiological parameters	Mortality prediction is sepsis
MEDS Mortality in Emergency Department Score	Functional status, vital parameters, lab values	Mortality prediction in sepsis
SIRS Systemic Inflammatory Response Syndrome	Vital parameters and laboratory values	Mortality prediction in sepsis
SCC Simple Clinical Score	Based on ABCDEF parameters: A; age, airway, SpO_2 B; breathing (respiratory rate) C: blood pressure/pulse D: stroke, altered mental status, pulse, E: ECG (abnormal ECG) F: fever	Overall mortality prediction

Abbreviations: AVPU: Alert, Verbal, Pain, None; EEG: electroencephalography; GCS: Glasgow Coma Scale; HR: heart rate; MAP: mean arterial pressure; SpO_2: blood oxygen saturation

2.3.2 Reliability of patient classification and scoring

Patient scoring and classification are widely accepted and used in all fields of medicine. Reliability can be evaluated by studying the true accuracy of mortality prediction, how well the scores detect deteriorating patients, what is the diagnostic accuracy of a score, and what are the intra- and interrater reliabilities of these scores among the people who use them (Bouzat et al. 2016, Challen et al. 2016). As described earlier, the scores can be used for screening of deteriorating patients, but

also for overall categorisation of patients (Nannan et al. 2017). To detect the deteriorating patients at an early stage and to target the most critical interventions to these patients is essential, and the benefit for a single patient, but also to the whole system, is obvious. Scores like The Glasgow Outcome Scale (GOS), Cerebral Performance Category (CPC) and Modified Ranking Scale (mRS) are used precisely to determine outcomes following brain injury or cardiac arrest (Raina et al. 2008, Rittenberger 2011, McMillan et al. 2016). In addition, based on these scores, patients can be categorised in subgroups by severity of their medical condition. This collected scoring and classification data can be used for scientific and system quality control purposes.

Despite the wide use of these scores, there is still a lack of evidence in sensitivity, specificity, and inter-rater reliability (Sankar et. al 2014, Arabian et al. 2015, Challen et. al 2016, Tusgul et al. 2017). In single patient or medical incident groups, such as Japanese trauma patients, even excellent reliability for outcome prediction has been shown (Miymoto et. al 2019*).* In addition, a single centre US study presented a substantial inter-rater reliability for ASA-PS among orthopaedic patients (Ihejirika et al. 2015). Moreover, in an earlier study among Finnish anaesthesiologists, a wide inter-rater variation was shown with ASA-PS (Ranta et al. 1997). The major deficit is that no large reliability studies exist on patient scoring and classification, especially ones that would cover more than one score and patient group. Even fewer studies exist on efforts to update and revise these scores, and there are no follow-up studies on the possible improvement of reliability after revisions.

2.3.3 Patient classification and scoring in prehospital setting

Patient scoring and classification are recommended and used also in a prehospital setting (Patel et al. 2018), but there is no international consensus on how and which scores or classification systems should be used prehospitally. There are instructions on ASA-PS and ICPC-2 that these should already be registered in the prehospital phase (Kruger et al. 2011, Norwegian directorate of health 2018). However, these two recommended variables are not originally built for prehospital use. There are in-hospital studies that question the reliability of ASA-PS (Sankar et al. 2014), yet based on a single prehospital study, the ASA-PS was shown to be substantially reliable (Ringdal et al. 2013). There are no reliability studies on prehospital use of ICPC-2.

The National Advisory Committee of Aeronautics (NACA) severity is one of the most widely used severity score in prehospital critical care and the EMS (Tryba et al. 1980, Raatiniemi et al. 2013, Schneider et al. 2018). The NACA score is an eight-level scoring system describing the injury or disease severity. The score is based on a clinical and subjective evaluation of a prehospital clinician. NACA levels are

described in Table 3. There are reliability studies on the NACA score, and it seems to carry substantial reliability on single centre studies (Raatiniemi et al. 2017, Darioli et al. 2019).

Table 3. The National Advisory Committee of Aeronautics (NACA) score (Tryba et al. 1980, Raatiniemi et al. 2017).

NACA 0	No injury or disease
NACA 1	Injuries/diseases without any need for acute physician care
NACA 2	Injuries/diseases requiring examination and therapy by physician, but hospital admission is not indicated
NACA 3	Injuries/diseases without acute threat to life but requiring hospital admission
NACA 4	Injuries/diseases that can possibly lead to deterioration of vital signs
NACA 5	Injuries/diseases with acute threat to life
NACA 6	Injuries/diseases transported after successful resuscitation of vital signs
NACA 7	Lethal injuries or diseases (with or without resuscitation attempts)

In addition to ASA-PS, NACA, and ICPC-2, there are other scoring and classification systems used prehospitally (Sartorius et al. 2010, Bouzat et al. 2016, Patel et al. 2018, Hoikka et al. 2018, Wang et al 2019) and even reliability studies on these. These prehospitally used scores include the same scores that are listed earlier and used in critical or in-hospital emergency care, even though these have not originally been developed for prehospital use. In addition, some prehospitally modified scores and classifications do exist, such as: prehospital early sepsis detection score (PRESEP); the rapid acute physiology score (RAPS); prehospital modified shock index (preMSI); and prehospital shock index (preSI) (Williams et. al 2016, Jouffroy et al. 2018). Some studies do report moderate, even good reliability of these prehospitally used scores. However, a major heterogeneity exists, and larger reliability studies are needed before real conclusions can be made.

There are prehospital studies that rely on patient scoring. For example, in recent Finnish studies on prehospital airway management and traumatic brain injury, the patient input and categorisation were done based on the Glasgow Coma Scale and the outcome was evaluated using the Glasgow Outcome Scale (Pakkanen et al. 2015, 2016 and 2017). The Glasgow Coma Scale (GCS) is a widely used and accepted tool to triage patients, categorise patients by severity of the acute incident, and even to guide critical pre- and in-hospital interventions like tracheal intubation. However, there is heterogeneous evidence on the reliability and accuracy of the GCS itself, and some studies indicate poor accuracy with the score (Bledsoe at al. 2015). In another recent study, the overall prehospital care was evaluated among Norwegian trauma victims (Wisborg et al. 2017). Patient selection was done with the Injury Severity Score (ISS). Patients receiving ISS > 15 were considered severely injured and

included in the study. But as in GCS, also in the case of ISS, a question might be raised about the reliability and usability of ISS for any patient (Maduz et al. 2017). These two examples demonstrate the importance of patient scoring and classification systems in prehospital scientific use.

2.3.4 The HEMS Benefit Score

The HEMS Benefit Score was introduced and implemented in clinical prehospital practice in 1997 (Raatiniemi et al. 2017). Despite its name, the score evaluates the benefit of the whole prehospital system for a single patient on HEMS missions. The score is based on a subjective opinion of the treating physician or advanced paramedic. The HEMS Benefit Score is a nine-level score, and a description of each of the levels is presented in Table 4.

Table 4. The HEMS Benefit Score (Raatiniemi et al. 2017).

0	The patient was not seen
1	Prehospital care was not deemed necessary
2	Prehospital care apparently had no significance from the patient's standpoint (e.g., cannulation, no medication or fluid therapy) or despite prehospital care the patient died before reaching the hospital
3	Prehospital care apparently had no significance from the standpoint of the prognosis, but the patient's symptoms or pain was alleviated (e.g., injured patient's analgesia)
4	Prehospital care was administered; its significance from the patient's standpoint is unknown, difficult to assess or only assessable retrospectively (e.g., treatment of ischaemic chest pain, brief convulsions, mild breathing difficulty)
5	Without prehospital care (administered by the first response unit or the physician-staffed unit), the patient would have died before reaching the hospital, but he/she is assessed as having a poor prognosis (e.g., serious brain damage, coma caused by spontaneous cerebral haemorrhage, primary survival from cardiac arrest after lengthy response times, terminal phase of a malignant disease)
6	The patient was given prehospital care that can be assessed to reduce mortality or otherwise improve the prognosis
7	Without prehospital care (administered by the first response unit or the physician-staffed unit), the patient would have died before reaching the hospital, and he/she cannot be assessed as having a poor prognosis
8	Category 7 in situations where other emergency medical staff on site would not have been capable of administering the aforementioned life-saving treatment

NOTE

- prehospital care = speed **and/or** quality of treatment **and/or** transport
- the basis must be assessment of benefit to the patient, not the demandingness or duration of the treatment
- the assessment must be done immediately after the operation, using available information, and the assessment must not be changed on the basis of information received later

The HEMS Benefit Score is currently only used in Finland. It is used nationally for effectiveness evaluation of the HEMS system, for example: comparison of HEMS bases and units, control of HEMS dispatch criteria, and overall evaluation of accurate use of HEMS units. Despite the over two-decade clinical use of the score – in patient evaluation, but also in system quality control – there is only one study evaluating the reliability and accuracy of the HEMS Benefit Score (Raatiniemi et al. 2017). This study focused on the inter-rater reliability of the HEMS Benefit Score and the NACA score based on intra-class correlation coefficient in a written patient scenario. The study found substantial inter-rater reliability in both studied scores.

There are written instructions on how the HEMS Benefit Score should be used in clinical practice, and how to score different types of prehospital interventions. However, the score was presented as early as 1997 with no later revision or update, and the instructions are based on prehospital protocols that were in use over twenty years ago (Raatiniemi et al. 2017). In addition, there are no follow-up studies on the outcome of patients rated in certain score categories, or studies on the accuracy of the HEMS Benefit Score. The instructions for use of the HEMS Benefit Score are presented in Table 5.

Table 5. The HEMS Benefit Score, application guideline (Raatiniemi et al. 2017).

Disease / injury	Score
Uncomplicated ST elevation myocardial infarction	
• Prehospital GP inhibitor, heparinoid and PCI < 90 min from emergency call	**6**
• Thrombolysis 0–4 h from onset of pain	**6**
• Thrombolysis > 4 h from onset of pain	**4**
Complicated ST elevation myocardial infarction	
• Whenever prehospital care has a positive response in haemodynamics	**6–8**
Cardiac arrest	
• No ALS attempted	**1**
• ALS attempted but the patient died	**2**
• Primarily survived normothermic adult	
○ VF ongoing when the physician encounters the patient	**7–8**
○ Found with asystole / PEA regardless of delays	**5**
○ BLS > 10 min **or** ALS > 20 min **or** ROSC > 30 min	**5**
○ BLS < 10 min **and** ALS < 20 min **and** ROSC < 30 min	**7–8**
Breathing difficulty without an injury	
• Pulmonary oedema and aggravated COPD	
○ SpO_2 < 80 % on encounter and treatment has a positive response	**6–8**
○ Decreasing SpO_2 regardless of administered oxygen	**6–8**
• Status asthmaticus	
○ SpO_2 < 90 % on encounter and prehospital care has a positive response	**6–8**
○ Decreasing SpO_2 regardless of administered oxygen	**6–8**
• Other aetiology; no previous incurable disease	
○ SpO_2 < 85 % on encounter and treatment has a positive response	**6–8**
Injury	
• Treatment of hypovolaemia > 1500 ml or > 20 ml / kg	**6–8**
• Securing the airway by intubation	**6–8**
• Capnography controlled ventilation as treatment of elevated ICP	**6**
• Successful drainage of the pleural cavity because of desaturation	**6–8**
• Successful drainage of tension pneumothorax	**6–8**
• Time savings > 30 min with helicopter transport of a multiple-injured patient	**6–8**
• Crucial time savings > 10 min with helicopter transport in a case of critical hypovolaemia caused by a penetrating injury (-> emergency surgery)	**8**
Status epilepticus and hypoglycaemia	
• Duration > 30 min; the patient awoke after glucose infusion	**6**
• Status epilepticus; treated with general anaesthesia	**6**
Unconsciousness without injury	
• Suspected intracranial haemorrhage and coma (GCS 3–5 /15)	**5**
• Securing the airway by intubation and controlled ventilation, if there is no reason to consider the patient's status as having a poor prognosis	**6**

GP inhibitor, Glycoprotein IIb/IIIa Inhibitor; PCI, Percutaneous Coronary Intervention; ALS, Advanced Life Support; VF, Ventricular Fibrillation; PEA, Pulseless Electrical Activity; BLS, Basic Life Support; ROSC, Return of Spontaneous Circulation; COPD, Chronic Obstructive Pulmonary Disease; SpO_2, Oxygen saturation measured by pulse oximetry; ICP, Intracranial Pressure; GCS, Glasgow Coma Score.

In terms of scores for HEMS use, Air Medical Prehospital Triage Score (AMPT) was presented in 2016 (Brown et al. 2016). As the HEMS Benefit Score, AMPT is also a score that is specially developed for the HEMS. Yet AMPT focuses solely on the transportation aspect of the HEMS, compared to the HEMS Benefit Score, which evaluates both the clinical interventions and mode of transportation when a patient is treated in a prehospital setting. The AMPT score aims solely to detect the patients that benefit from helicopter transportation. As described earlier, the development and implementation of the HEMS Benefit Score did not include any structured study process that would have been described in peer-reviewed literature. The first study was presented no sooner than twenty years after the HEMS Benefit Score was launched in clinical use. In the case of the AMPT score, its development and implementation processes are described in the existing literature (Brown et al. 2016). There are also studies on the external validation and cost effects of the AMPT score (Brown et al. 2017 and 2018).

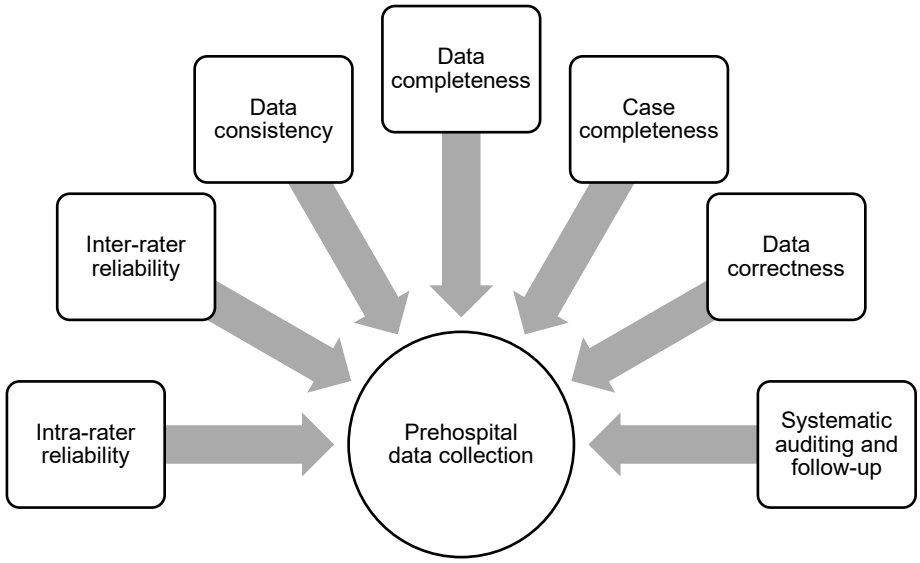

Figure 2. Factors influencing prehospital data collection.

3 Aims

The aims of the thesis were:

1. To assess the reliability and accuracy of data collection in prehospital helicopter emergency services (studies I and II)
2. To assess the reliability of prehospital classification and prognosis among futile patients, and describe the features of these patients (study III)
3. To revise the HEMS Benefit Score to illustrate current prehospital operational environment and interventions performed (study IV)

4 Materials and Methods

4.1 Methods

4.1.1 Prehospital data collection (I and II)

Study I focused on the inter-rater reliability of mission coding, time variables, and vital parameters among Finnish HEMS clinicians. Whereas in the study II, the inter-rater reliability of patient scoring and classification were evaluated between these study participants. For studies I and II, a separate study registry was formed, which was identical with the actual FinnHEMS database used in the Finnish HEMS system. Studies I and II included 42 HEMS clinicians who were provided with written fictional scenarios including six HEMS missions with seven patient scenarios. All participants received identical material, which consisted of written descriptions of the mission duty, HEMS reports, EMS reports, and dispatch centre messages. These followed the real-life material used on HEMS mission documentation in Finland. The patient and mission scenarios were set in a form following the real-life nature of 24-hour duty in HEMS, and written descriptions were based on earlier user feedback on the usability of the database. The studies aimed to detect faults in the database, and the written scenarios were intentionally focused on the plausible problematic issues in registration habits.

Participants were asked to fill in the study registry based on the written materials provided, and in a similar way as they would do in their clinical work. As in the FinnHEMS database, also in the study database, all parameters included in the registry were documented and analysed. These parameters were time-related variables, patient-related vital parameters, intervention-related parameters, mission coding, patient scoring, and patient classifications. In addition, plausible adverse events in airway management are documented in the FinnHEMS database. The studied scores and classifications were: the American Society of Anesthesiologists physical status (ASA-PS) classification system, the HEMS Benefit Score (HBS), International Classification of Primary Care, second edition (ICPC-2), and Eastern Cooperative Oncology Group (ECOG) performance status. The scoring data was analysed within three individual patient descriptions, and on one patient that was described most severely injured in a multi-patient mission and was treated and

transported by HEMS. The three other patients in the multi-patient mission were triaged by the HEMS physician but transported by other EMS units. Thus, the following four patients were included in the study: a cardiac arrest patient, a major trauma patient, a seizuring paediatric patient, and an unconscious drug abuser.

4.1.2 Prehospital classification and prognosis (III)

Patient prognosis is evaluated by the treating HEMS clinician in all Finnish HEMS missions. The actual prognosis is based on observation and findings from clinical intervention by the HEMS clinician. Prognosis is registered in the FinnHEMS database by using the HEMS Benefit Score (HBS) (See Table 4.), which evaluates the benefit of EMS service for a patient and is used in Finnish HEMS units. In this nine-level grade, scores 5, 7, and 8 include a narration: patients would not have survived until hospital without prehospital interventions. The difference between these three values is that score 5 is defined as patients being futile, opposite to scores 7 and 8, where patients are considered non-futile. Scores 7 and 8 are demerged, as 8 is used when only the HEMS unit is capable to provide the life-saving intervention. Score 6 represents interventions where the impact of prehospital care was considered unclear, and therefore this category was not included in the study.

Study III included all patients with a HEMS Benefit Score of 5, 7, or 8 and registered in the FinnHEMS database between 1 January 2012 and 8 September 2019. This constitutes 6219 patients considered as not to survive without prehospital intervention. The patients were divided into two subgroups: patients considered non-futile (HBS 7 and 8) and patients considered futile (HBS 5) by the HEMS clinician. 30-day and 3-year survival were the primary endpoints. These endpoints were compared between non-futile and futile sub-groups.

4.1.3 Revision of the HEMS Benefit Score (IV)

Study IV was a three-round web-based Delphi study using an expert panel consensus process to define intervention examples for HBS categories. The technique involves a panel of experts who are asked to complete a series of questionnaires focusing on their opinions, predictions, or judgement about the topic of interest. The Delphi technique is widely used in research to obtain consensus in serial surveys, referred to as rounds (Polit et. al 2004 and 2007, Diamond et al. 2014). Key elements of the technique are 1) expert participants, 2) anonymity and individuality, and 3) a summary of results of the former round at the start of each round. Data collection, Delphi rounds, and data analysis were performed from 3 December 2018 to 19 November 2020. A pilot study was performed prior to the actual study to test the study setting. The pilot study participants consisted of Finnish and Danish

prehospital physicians, who did not participate in the planning of the study or the actual study.

The work of the expert panel and commentary board were executed in four Delphi rounds in the following manner:

1. The first Delphi round was performed by a Webropol software based electronic data collecting sheet. For ten complaint-based diagnoses, each expert was asked to list both common and rare examples of prehospital treatments and interventions and classify them, based on their current knowledge and personal experience, into categories from 8 to 3 in the HEMS Benefit Score (Kruger AJ et al. 2011). The answers were collected anonymously in an electronic data sheet by a data collector officer, who did not participate in example selection, but gathered suggestions in a common table.

   An additional commentary board gave their comments on the data gathered from the first Delphi round. Their comments were revealed for the expert panel and used on the second Delphi round to support the expert panel's work on Likert scaling the examples.

2. For the second Delphi round, identical suggestions from the first round were combined. The examples were set in a table and sent back to the panelists, who were asked to rate each example using a 5-point Likert scale (ranging from "strongly disagree" (1) to "strongly agree" (5)). Based on the Likert scale and data collected from this second Delphi round, a content validity index was calculated for each of the examples. At least 70% of the experts were required to rate a suggested example in the high agreement range of scores (4 or 5) for the example to be selected into the revised version of the HEMS Benefit Score.

3. In the third Delphi round, the examples which received at least 70% agreement in the high range of the Likert scale were analysed, and overlapping examples were removed. The remaining examples were then listed in their suggested HEMS Benefit Score categories. The expert panelists were asked to grade these remaining examples in each category as: "Accept", "Delete", or "Relocate to category number __". If an example received a higher than 70% rating on "Delete" or "Relocate to category number __", this suggested action was performed. The expert panelists were also offered the opportunity for free comments on each EBS category and on the whole process. The examples with acceptance rates below 70% were deleted or relocated to the category with the most "Relocate" suggestions – whichever had a higher percentage.

4. In the final Delphi round, the EBS was revealed to the prehospital expert panellists, who were offered an opportunity to comment on it or accept it in that form.

4.2 Study subjects

4.2.1 Prehospital data collection (I and II)

For studies I and II, a total of 59 Finnish helicopter emergency medical services operational clinicians were recruited to participate. Of these, 42 participants took part in the final study. The study population included 36 physicians and 6 advanced paramedics. Participants, who were active operational clinicians during the study period, were from all six helicopter emergency service bases in Finland. Study I and II participants are described in Figure 3.

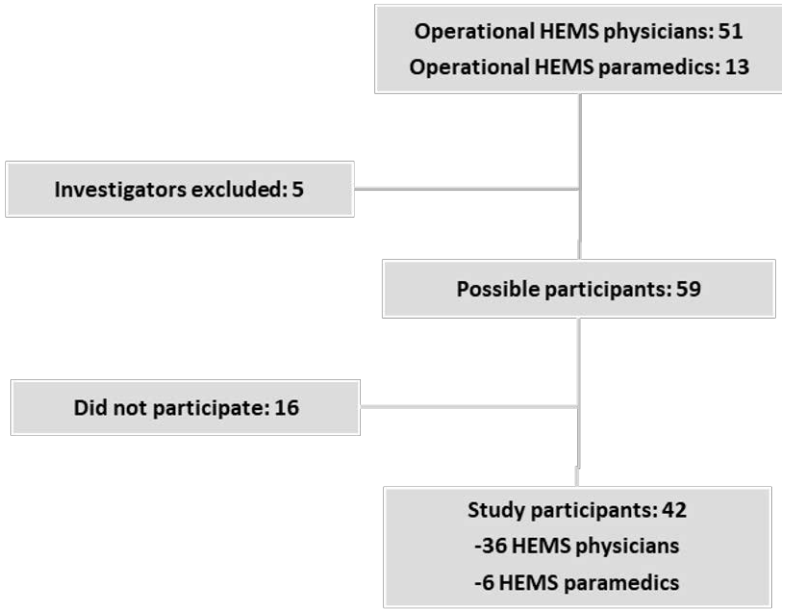

Figure 3. Participants of the studies I and II.

4.2.2 Prehospital classification and prognosis (III)

For study III, all patients met by Finnish helicopter emergency medical service units between 1 January 2012 and 8 September 2019 were screened. Patients who were evaluated on-scene by the treating clinician not to survive until hospital without

prehospital critical care were included in the analysing phase. This constitutes 6219 patients in total. Study III participants are described in Figure 4.

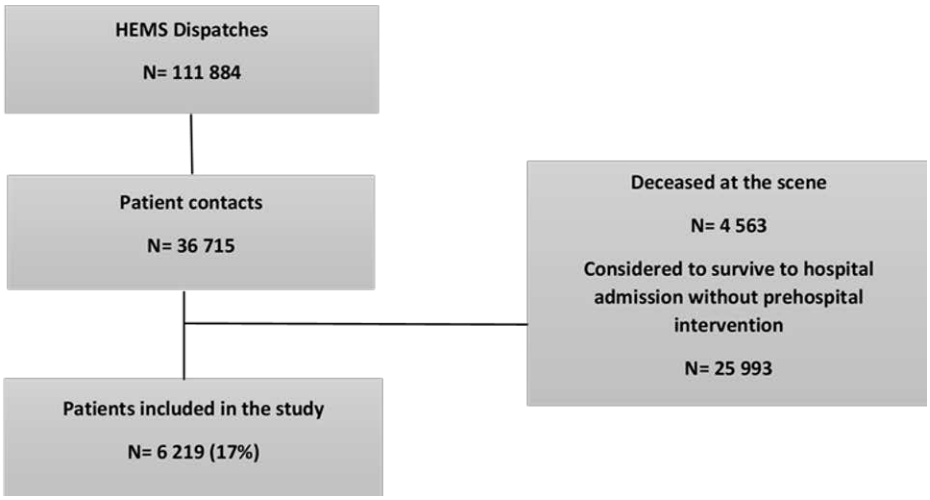

Figure 4. Participants of the study III.

4.2.3 Revision of the HEMS Benefit Score (IV)

For study IV, two expert groups were formed. For the first group, the European Prehospital Research Alliance (EUPHOREA) was used for the recruitment process. In the case of the second expert group, Finnish national specialty societies were used. Both expert groups were recruited based on individual clinical and scientific experience. The first group included 18 prehospital physicians from Scandinavia and Northern Europe, whereas the second group included 11 Finnish physicians from six different specialties. The total number of study subjects was 29. The separate commentary board containing physicians from different specialties was gathered to give their comments on diagnose groups related to their specialty. Physicians from traumatology, cardiology, neurology, neurosurgery, paediatrics, and obstetrics were recruited. Characteristics of the expert panelists are described in Table 6.

Table 6. Characteristics of 18 prehospital expert panelists.

	n	%
Speciality		
Anaesthesia	6	33%
Anaesthesia and intensive care	8	44%
Emergency medicine	3	17%
Anaesthesia, intensive care and emergency medicine	1	6%
Clinical experience in prehospital critical care		
5–10 years	1	6%
10–15 years	9	50%
15–20 years	4	22%
over 20 years	4	22%
Number of peer reviewed publications		
Less than five	4	22%
5–10	2	11%
10–20	5	28%
More than 20	7	39%

4.3 Statistical analysis

In studies I and II, inter-rater agreement was evaluated with percent agreement and free-marginal multi-rater kappa (Randolph 2005, McHugh et al. 2012). The number of equal variables among raters is divided with the total number of variables, which presents the percent agreement between raters. Unlike percent agreement, free-marginal multi-rater kappa considers the random agreement factor and is suitable for studies with free-marginal distributions. Free-marginal multi-rater kappa values vary from -1 to 1, as value 0 implies agreement equal with chance. Value 1 represents full agreement and values between 1 and 0 indicate a level of agreement better than chance, whereas -1 denotes full disagreement among raters. The analysis was performed with IBM SPSS Statistics 25 and with an online kappa calculator: http://justusrandolph.net/kappa/.

In study III, patients were divided into two groups according to the estimated prognosis by the prehospital clinician: (a) received life-saving prehospital treatment and evaluated as non-futile (HBS 7 to 8) and (b) those receiving life-saving prehospital treatment but estimated futile (HBS 5). Proportions are reported as n (%). Primary endpoints were mortality at 30-days and 3-years after the HEMS dispatch. Comparisons between the groups were evaluated with the Mann-Whitney U test for non-normally distributed numeric variables, a two-sample t-test was used for normally distributed variables, and chi^2 was calculated for categorical variables. A P value of < 0.05 was used for the analysis of statistical significance. Standard

deviations of the means and interquartile ranges were used for assessment of the null hypothesis. A Kapplan-Meier graph was used to illustrate the long-term survival rates of the studied groups. IBM SPSS Statistics 25 (IBM Corporation, Armonk, NY, USA) was used for analyses. The properties of the futile classification as a diagnostic test were estimated by calculation of accuracy with GraphPad Prism for Mac 8.41 (GraphPad Software, California, USA).

For study IV, a nominal group technique and Delphi method were used (Fink et al. 1984, Hasson et al. 2000, Junger et al. 2017). Data collection and data handling were performed using Webropol 3.0 by Webropol Group. Content validity index (CVI) was calculated for the collected data after the second Delphi round.

4.4 Study ethics

The ethical committees of each of the five Finnish university hospital districts were contacted for verification that no ethical approval was needed for studies I and II. All five university hospital districts gave their approval for the study. The study subjects participated voluntarily and gave their consent when filling in the study data. No personal data was collected, and data could not be associated with individual study subjects in studies I or II.

Study permission for study III was requested from and granted by all the participant hospital districts (Oulu University Hospital, Helsinki University Hospital, Turku University Hospital, Hospital District of Lapland, Kuopio University Hospital, and Tampere University Hospital). According to Finnish Law, ethical permission is not required for studies not involving patient contact. However, due to the large amount of data in study III, including sensitive patient data, ethical permission was requested from and granted by the Ethical Board of the University of Helsinki (HUS/3115/2019 §194).

In study IV, the Turku University Hospital ethical board was contacted for verification that no ethical approval was needed for this study, as there was no patient involvement included in the study setting. Turku University Hospital district approved study IV. Study IV subjects gave their consent when filling in the data. No personal data was collected, and data could not be associated with individual study subjects in study IV.

5 Results

5.1 Reliability of prehospital data collection in HEMS (I and II)

Study I included 42 HEMS clinicians from six national HEMS bases: 10 (24%) were from Vantaa, 9 (21%) from Turku, 7 (17%) from Tampere, 4 (10%) from Oulu, 6 (14%) from Kuopio, and 6 (14%) from Rovaniemi. Gender distribution was 13 (31%) female and 29 (69%) male participants.

Dispatch coding had the least inter-rater variability, whereas transport coding or mission cancellation varied more among study participants (Figure 5.). Most variation among raters was seen in the use of cancellation codes X-0 (mission denied, for example technical barrier or concurrent mission) and X-9 (mission cancelled). Inter-rater reliability of time-related variables is presented in Figure 6.

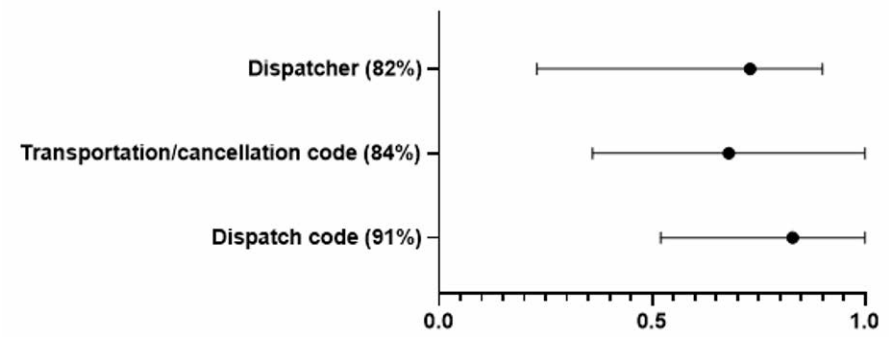

Figure 5. Free marginal multi-rater kappa distribution, with 95% confidence intervals, and percent agreements (n%) of mission coding among 42 HEMS clinicians.

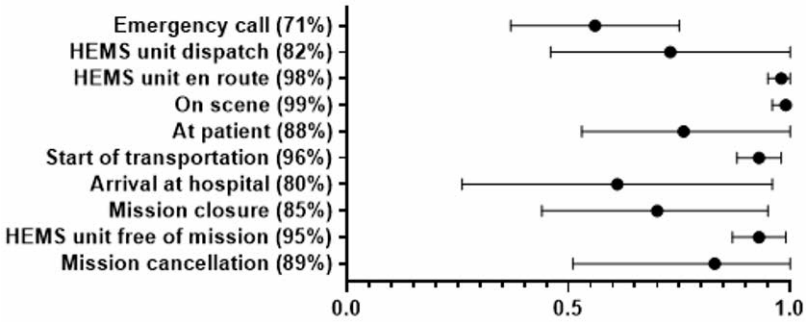

Figure 6. Free marginal multi-rater kappa distribution, with 95% confidence intervals, and percent agreements (n%) of time-related variable documentation among 42 HEMS clinicians.

Vital parameter values consisted of two separate time-points, excluding the Glascow Coma Scale that is registered only once. Inter-rater reliability of these parameters is presented in Table 7.

Table 7. Inter-rater reliability of vital parameter documentation among 42 HEMS clinicians.

Parameter[1] **1. measurement performed as patient is met** **2. measurement performed after medical intervention**	**Percent agreement**	**Free marginal multi-rater kappa, K_{free} [CI 95%]**
Cardiac rhythm	61 76	0.23 [-0.11, 0.56] 0.51 [0.17, 0.86]
Heart rate	60 82	0.19 [0.02, 0.37] 0.64 [0.34, 0.94]
Blood pressure	69 77	0.38 [0.04, 0.73] 0.54 [0.17, 0.91]
Respiration rate	65 72	0.30 [-0.11, 0.70] 0.43 [0.01, 0.86]
Blood oxygen saturation	68 91	0.36 [0.03, 0.69] 0.82 [0.72, 0.92]
Expiration carbon dioxide	62 76	0.24 [0.07, 0.42] 0.52 [0.22, 0.82]
Pain	73 73	0.46 [0.31, 0.60] 0.46 [0.17, 0.75]
GCS*	66	0.49 [0.16, 0.82]

[1]Vital parameter, registered in the FinnHEMS database by a participating prehospital clinician and based on the documents provided
*GCS is documented only once

In study I, the documentation of a multi-patient mission was evaluated, as one of the descriptions included four patients: of the 42 participants, 41 (98%) registered the

patient described as being the most severely injured in the multi-patient mission, whereas only 23 (55%) participants registered all four patients described therein.

Study descriptions included one patient receiving rapid sequence intubation, and 33 (79%) participants registered no adverse events for the airway management in the treatment of this patient. However, eight participants (19%) did register hypotension with hypoxia for this described incident, and one (2%) registered only hypotension as an adverse event.

In study II, the scoring and classification data of four patient descriptions were analysed. 42 participants evaluated the scoring and classification data on each of these patient descriptions, excluding one missing registration by one participant on a major trauma patient.

ASA-PS resulted in an overall agreement of 40.2% and Kfree of 0.28 [95% CI 0.12, 0.44]. ASA-PS distribution is described in Table 8.

Table 8. American Society of Anesthesiologists physical status (ASA-PS) distribution of four patients as recorded by 42 prehospital clinicians.

Patient case	American Society of Anaesthesiologists physical status						
	I	II	III	IV	V	Not known	Missing
Cardiac arrest	4	24	6	0	1	7	
Major trauma	32	5	0	0	1	3	1
Paediatric seizures	6	21	14	0	0	1	
Drug abuse, unconscious	15	15	4	1	0	7	

ASA I: ”A normal healthy patient”, ASA II: ”A patient with a mild systemic disease”, ASA III: ”A patient with a severe systemic disease”, ASA IV: ”A patient with a severe systemic disease that is a constant threat to life”, ASA V: ”A moribund patient who is not expected to survive without the operation”

HBS had an overall agreement of 44.7% and Kfree of 0.39 [95% CI 0.26, 0.51]. HBS distribution is described in Table 9.

Table 9. The HEMS Benefit Score distribution of four patients as recorded by 42 prehospital clinicians.

Patient case	The HEMS Benefit Score						
	3	4	5	6	7	8	Missing
Cardiac arrest	0	0	1	6	30	5	0
Major trauma	3	2	0	18	5	13	1
Paediatric seizures	3	26	0	13	0	0	0
Drug abuse, unconscious	1	1	0	13	27	0	0

ICPC-2 coding had an overall agreement of 51.5% and Kfree of 0.47 [95% CI 0.28, 0.67]. ICPC-2 distribution is described in Table 10.

Table 10. International Classification of Primary Care, second edition (ICPC-2) distribution of four patients as recorded by 42 prehospital clinicians.

Patient type	ICPC-2 Code n (%)				
Cardiac arrest	A96 Death **3[3] (7%)**	K80 Cardiac arrhytmia NOS **14 (33%)**	K99 Cardiovascular disease other **23 (55%)**	N79 Concusion **1 (2%)**	A80 Trauma/Injury NOS **1[6] (2%)**
Major trauma	A81 Multiple trauma/injuries **23[4] (55%)**	D80 Injury digestive system other **1 (2%)**	A80 Trauma/Injury NOS **12 (29%)**	A10 Bleeding/haemorhage NOS **5[7] (12%)**	Code missing **1 (2%)**
Paediatric seizures	N88 Epilepsy **5 (12%)**	N07 Convlusion/Seizure **37[2] (88%)**			
Drug abuse, unconscious	P19 Drug abuse **25[1] (60%)**	A84 Poisoning by medical agent **16 (38%)**	N79 Concusion **1[5] (2%)**		

NOS: Not Otherwise Specified

Study CQR enables an additional ICPC-2 with the primary code:

[1] two participants additionally coded **A84**

[2] two participants additionally coded **N88**

[3] one participant additionally coded **K99**

[4] one participant additionally coded **N80 (head injury other)** and one participant additionally coded **A10**

[5] one participant additionally coded **P19**

[6] one participant additionally coded **A88 (adverse effect physical factor)**

[7] one participant additionally coded **A80**

ECOG had an overall agreement of 49.6% and Kfree of 0.40 [95% CI 0.11, 0.68]. ECOG distribution is described in Table 11.

Table 11. Eastern Cooperative Oncology Group (ECOG) performance status distribution of four patients as recorded by 42 prehospital clinicians.

Patient case	**Eastern Cooperative Oncology Group performance status**						
	0	**1**	**2**	**3**	**4**	**Not known**	**Missing**
Cardiac arrest	22	5	1	0	0	14	
Major trauma	38	0	0	0	0	3	1
Paediatric seizures	18	9	4	2	1	8	
Drug abuser, unconscious	29	5	0	0	0	8	

0 Fully active, able to carry on all pre-disease performance without restriction
1 Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2 Ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours
3 Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours
4 Completely disabled; cannot carry on any selfcare; totally confined to bed or chair

5.2 Reliability of prehospital classification and prognosis (III)

A total of 6219 patients were classified as not surviving until hospital admission without prehospital interventions and were included in study III (Figure 7). Of these patients, 4166 (67%) were classified as non-futile (HBS 7 or 8) and 2053 (33%) were classified as futile (HBS 5) in the clinical registry.

The survival rates of the patients are presented in Figures 7 and 8. At 30 days, 2803 (67.3% / 95% CI 65.8% to 68.7%) of the patients classified as non-futile and 713 (34.7% / 95% CI 32.7% to 36.8%)) classified as futile were alive. At 3 years, 2 356 (56% / 95% CI 55.0% to 56.6%) and 522 (26% / 95% CI 31.3% to 76.8%) of the non-futile and futile patients were still alive.

The intensity of prehospital critical care was similar in non-futile and futile patients: drug-assisted endotracheal intubation was performed in 2850 (68.6%) and 1305 (65.6%) non-futile and futile patients, when vasoactive drugs were used in 2275 (54.6%) and 994 (48.4%) non-futile and futile patients, respectively. The overall accuracy for futile prognostication by a HEMS clinician was 68.8% (95% CI, 67.6% to 69.9%) in study III.

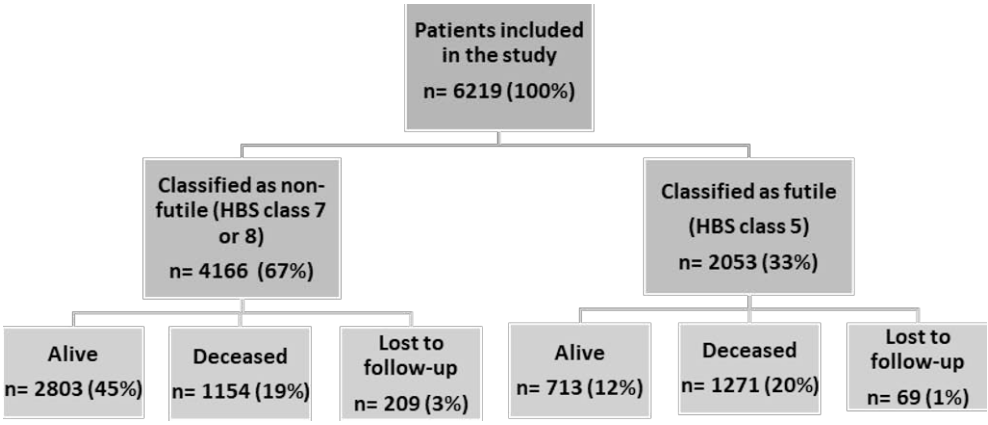

Figure 7. Study III patients.

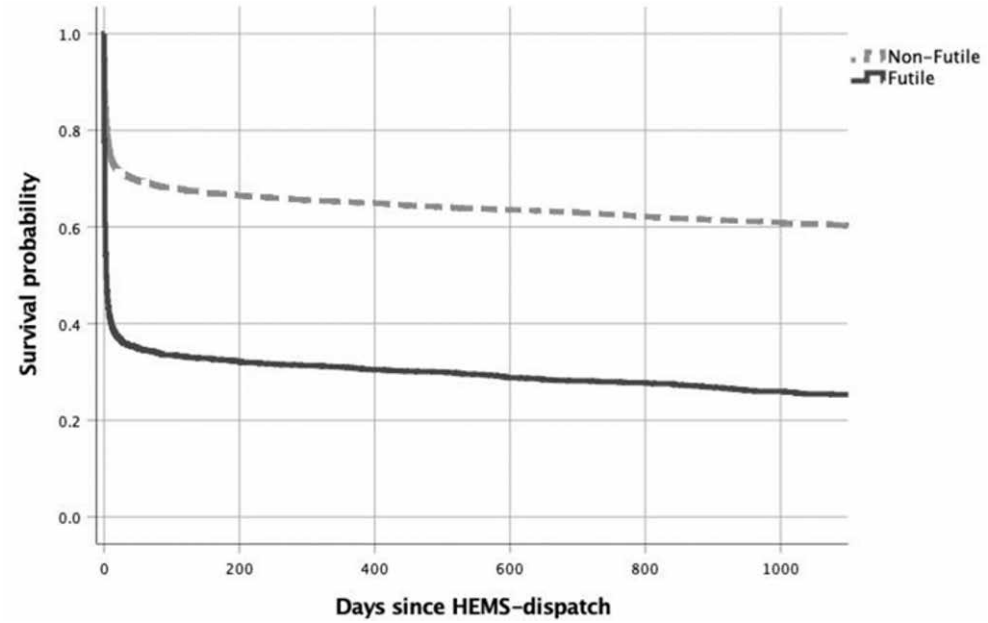

Figure 8. Kaplan-Meier survival comparison of patients considered futile or non-futile by prehospital clinician.

5.3 Revision of the HEMS Benefit Score (IV)

A total of 1484 examples from 18 expert panelists were received in the first Delphi round of study IV. These were divided in the HEMS Benefit Score categories from 3 to 8, and into subdivisions "Acute neurology excluding stroke", "Breathing difficulties", "Cardiac arrest", "Chest pain", "Infection", "Obstetrics including childbirth", "Other", "Psychiatry including intoxication", "Stroke", and "Trauma". Furthermore, seven participants complemented their answers with free-form comments.

The overlaps between examples on separate HEMS Benefit Score categories and between participants were analysed, and data was pooled. This produced 413 examples allocated in HEMS Benefit Score categories from 3 to 8. For this data, the independent commentary board gave their expert opinions on the relevance of suggestions related to their specialty and expertise.

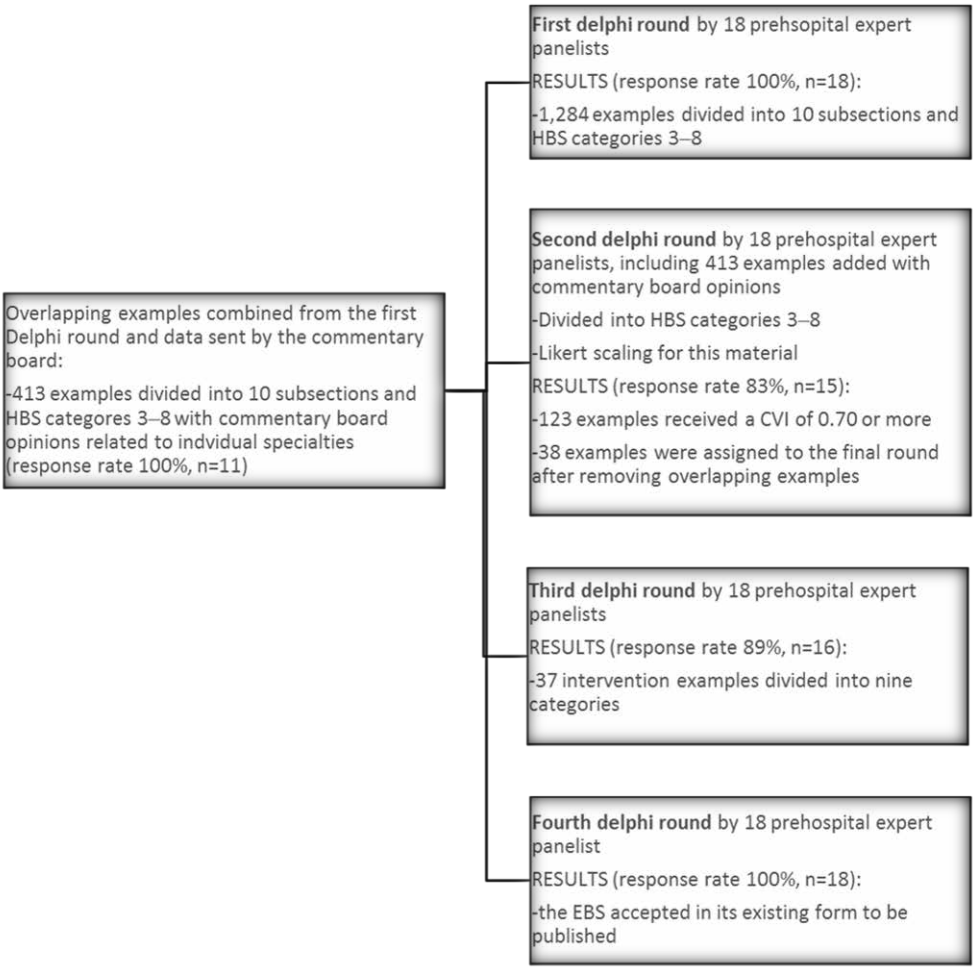

Figure 9. The course of the Delphi rounds in study IV.

During the second Delphi round, the 413 examples were graded on the Likert-Scale by prehospital expert panelists, and examples obtaining content validity index of 0.70 or more were assembled. This represented 123 examples. After pooling and eliminating of overlapping examples, a total of 37 examples were taken to the third Delphi round, and in the fourth Delphi round a consensus acceptance was received. The final form of the score is presented in Table 12.

Table 12. The EMS Benefit Score.

EBS	DESCRIPTION	EXEMPLARY INTERVENTIONS*
0	**The patient was not seen**	
1	**Prehospital care was not deemed necessary**	
2	**Prehospital care apparently had no significance from the patient's standpoint** (e.g., cannulation, no medication or fluid therapy) or, despite prehospital care, the patient died before reaching the hospital	
3	**Prehospital care apparently had no significance from the standpoint of the prognosis, but the patient's symptoms or pain was alleviated** (e.g., injured patient's analgesia)	○ Administration of analgesics ○ Administration of antihistamines to treat an allergic reaction ○ Antiemetic medication
4	**Prehospital care was administered; its significance from the patient's standpoint is unknown**, difficult to assess or only assessable retrospectively (e.g., treatment of ischaemic chest pain, brief convulsions and mild breathing difficulty)	○ Trauma patient immobilisation (cervical collar, back board, etc.) ○ Administration of inhaled bronchodilators for COPD or pneumonia ○ Administration of oxygen in moderate breathing difficulty
5	**Without prehospital care** (administered by the first response unit or the physician-staffed unit), **the patient would have died before reaching the hospital, but he/she was assessed as having a poor prognosis** (e.g., serious brain damage, coma caused by spontaneous cerebral haemorrhage, primary survival from cardiac arrest after lengthy response times and terminal phase of a malignant disease)	○ Patient treated but due to severe symptoms and/or underlying diseases has a poor prognosis (e.g., severe trauma or traumatic cardiac arrest, severe hypoxic insult, prolonged resuscitation and cardiac arrest due to severe traumatic brain injury or subarachnoid haemorrhage)
6	**The patient was given prehospital care that can be assessed to reduce mortality or otherwise improve the prognosis**	○ Administration of physician-staffed EMS-level medication (medication not allowed in other units) followed by relief of signs and symptoms ○ Administration of tranexamic acid ○ Medication for circulatory support (i.v. ephedrine, i.v. noradrenaline or norepinephrine, etc.) ○ Treatment of prolonged seizures by first- or second-line i.v. medication (bentsodiazepines, phosphenytoin, etc.) ○ Treatment of hypoglycaemia-induced coma or seizures by i.v. glucose or s.c./i.m. glucagon ○ Treatment of hypoglycaemia by i.v. glucose or s.c./i.m. glucagon when patient is disoriented but not in coma ○ Reduction and stabilisation of fractures or luxations ○ Triage and patient selection to dedicated centre and rapid transportation (major trauma, TBI,

EBS	DESCRIPTION	EXEMPLARY INTERVENTIONS*
		need of thrombectomy, need for re-implantation in traumatic amputation, etc.) o Treatment of opioid or benzodiazepine poisoning by antidotes o Maternal positioning in case of prolapsed umbilical cord o Thrombolysis for STEMI in cases with long transportation times o Rapid transportation to PCI
7	**Without prehospital care** (administered by the first response unit or the physician-staffed unit), **the patient would have died before reaching the hospital, and he/she cannot be assessed as having a poor prognosis**	o Mass casualty incident leadership and triage o Treatment and stabilising of a multi-trauma patient in shock by i.v. fluid administration and/or vasoactive medication o Isolated severe trauma managed with simple manoeuvres (e.g., direct compression and tourniquet) o Needle thoracocentesis followed by a relief of signs and symptoms o Cardioversion or cardiac pacing o Medication (adrenalin/epinephrine) in anaphylactic shock and relief of signs and symptoms o Successful resuscitation with reasonable prognosis o Transfer to ECMO, bypass or angiography during CPR o Manual opening of an obstructed airway and bag-mask ventilation o Use of a supraglottic device and bag-mask ventilation
8	**Category 7 in situations where other emergency medical staff on site would not have been capable of administering the aforementioned life-saving treatment** (use of physician-staffed EMS unit or advanced trained paramedic unit in systems where licenced to perform)	o Thoracotomy or tamponade release with other manoeuvres o Thoracostomy or pleural drainage followed by relief of signs and symptoms o ECMO initiation in prehospital phase (ECPR) o Management of complicated childbirth (shoulder dystocia, malposition, etc.) o Prehospital Caesarean section (resuscitative hysterectomy) o Resuscitation of a newborn by bag-mask ventilation or by more advanced procedures o Rapid sequence intubation or surgical airway management and mechanical ventilation o Blood product transfusions

*Interventions are listed as examples for each category; determination of the correct EBS is made by the prehospital clinician in charge of patient care.

COPD: chronic obstructive pulmonary disease, ECMO: extracorporeal membrane oxygenation, ECPR: extracorporeal cardiopulmonary resuscitation, CPR: cardiopulmonary resuscitation, STEMI: ST-elevation myocardial infarction

Table 13. Summary of the studies in the thesis.

	Aims	Study population	Methods	Results
Study I	Inter-rater reliability of data collection in national HEMS	42 Finnish HEMS clinicians	Free marginal multi-rater Kappa and percent agreement	Poor inter-rater reliability was found on vital parameter data.
Study II	Inter-rater reliability of patient classification and scoring in national HEMS	42 Finnish HEMS clinicians	Free marginal multi-rater Kappa and percent agreement	Inter-rater reliabilities of ASA-PS, ECOG, ICPC-2 and HBS were on a poor level.
Study III	Accuracy of prognostication in national HEMS	6219 patients considered not to survive without HEMS intervention	30-day and 3-year survivals between patients considered futile or non-futile	Futility was similar between studied groups.
Study IV	Revision of a prehospital scoring system	18 prehospital experts and 11 in-hospital commentary board members	Delphi methodology	A revised form of a EMS Benefit Score was formed

ASA-PS American Society of Anesthesiologist Physical Status, ECOG Eastern Co-Operative Oncology Group, EMS Emergency Medical Services, HEMS helicopter emergency medical services, HBS Hems Benefit Score, ICPC-2 International Classification of Primary Care Second Edition

6 Discussion

6.1 Reliability of prehospital data collection

Despite the findings of study I, where the inter-rater reliability was found to be on an adequate level in most of the analysed data, there were marked discrepancies in vital parameter documentation, and inter-rater reliability was poor for vital parameters. This is a marked liability in the documentation and should be focused on when HEMS data collection is improved. Vital parameters are a critical part of quality control and post-assessment of patient management; they also form the basis for clinical research. As the information on vital parameters was delivered in written form, the most likely reason for this inadequacy is that the correct time-point for data registration is unclear. For example, whether blood pressure should be registered immediately after an intervention, or just before the start of patient transportation. Even though instructions for registration were found, there seems to be incoherency as to how these are put into practice.

As vital parameters, time-related variables are also an important part of prehospital data collection. For instance, for patient groups like sudden cardiac arrest or major trauma, treatment delay (from the start of the emergency call to the hospital) is one of the most significant elements in measuring and improving the use of the EMS and the HEMS (Harmsen et al. 2015, Perkins et al. 2018). However, when time-related variables were studied in study I, the time of the emergency call had the lowest inter-rater reliability. In addition, the documentation of HEMS unit dispatch, mission end and arrival at hospital times showed only moderate inter-rater reliability. The correctness of vital parameter and time-related data documentation could benefit from automate data collection by monitoring devices, even though these devices have limitations as well, especially when used in volatile prehospital settings (Liu et al. 2015, Nakada et al. 2016). The accuracy of prehospitally registered data is essential, as time delays influence decision making for example on cardiac arrest patients, and vital signs are a key element when deteriorating trauma patients are detected.

Documenting possible adverse events is a vital part of prehospital critical care, as in all healthcare (Hagiwara et al. 2019). Recording complications in prehospital airway management is recommended, and a platform for this is presented in the existing literature (Sunde et al. 2018). Study I focused on overall documentation,

with no major focus on adverse events; therefore, no complete conclusions on their reliability can be made. A similar imperfection applies to the data gathered from multi-patient missions.

The goal of study II was to assess the inter-rater reliability of the prehospital ASA-PS, HBS, ICPC-2, and ECOG. The results of study II reveal that the ICPC-2 has moderate, and the ASA-PS, HBS, and ECOG poor, inter-rater agreement among studied prehospital clinicians. This could be explained by limited patient history that is accessible in the prehospital phase of patient care, or because the ASA-PS, ICPC-2, and ECOG were not originally formed for prehospital use. ASA-PS is a classification system to be used in assessment of pre-anesthesia medical co-morbidities, and for perioperative risk evaluation. As most HEMS units are staffed with senior anesthesiologists who are familiar with perioperative use of ASA-PS, the explanation for inter-rater unreliability is not an unfamiliar scoring system, but most likely the unsuitability for a prehospital setting. In addition, in an earlier study among Finnish anaesthesiologists, a wide inter-rater variation was shown with ASA-PS also in preoperative evaluation and in hospital use (Ranta et al. 1997). Still, both ASA-PS and ICPC-2 are recommended to be followed and to be registered already in the prehospital stage (Kruger et al. 2011, Norwegian Directorate of Health 2018). In addition, the HBS demonstrated poor inter-rater reliability in study II, suggesting that the HBS needed to be updated and revised systematically.

Based on the findings of the thesis and as the results on data reliability were revealed, actions to correct the existing faults were launched in the Finnish HEMS. For example, more detailed instructions were constructed for the use of the database and on how each variable should be registered. The focus has been especially on the sections which had the most inter-rater variation in the thesis study I. In addition, an independent data handling officer has been appointed for verification and follow-up of the registered data and its coherency and accuracy. These measures have been taken to address the inaccuracies revealed. Yet data quality assurance is an ongoing process, where reliability evaluation needs to be performed regularly; and corrective procedures are required if more inaccuracies are revealed.

Study II in the thesis focused on the reliability of the scoring systems used in the Finnish HEMS. The thesis presented that all the registered scores had major insufficiencies when used in a prehospital setting, and all the scores had marked inter-rater unreliability. Therefore, the focus has since been on more detailed guidance on how prehospital clinicians should score their patients. It has even been speculated if any of these scores should be registered in the prehospital phase at all. For example, ICPC-2 has been replaced with ICD-10 coding, and ICPC-2 is no longer registered or gathered in the Finnish HEMS. However, this maneuver does not solve the initial problem: structured implementation for prehospital use, described for instance in case of AMPT score (Brown et al 2016). Therefore, an

ICPC-2 code including precise instructions for prehospital use is under development and will hopefully solve the inter-rater unreliability of the prehospital diagnostic patient coding. Similar proposals have also been revealed on ECOG, but as there are no substitutes, the focus remains on the guidance for and accurate use of the score. In terms of ASA-PS, this coding is no longer registered on the Finnish HEMS missions, due to its inaccuracy when used prehospitally. The same shortfalls and inaccuracies were found in the use of the HEMS Benefit Score. However, corrective acts were started in study IV; and these measures will continue when the revised HEMS Benefit Score and the EMS Benefit Score are studied in a clinical setting with actual prehospital patients.

6.2 Patient classification and prognosis in prehospital setting

Study III focused on the reliability and accuracy of prehospital patient classification and prognosis by the HEMS Benefit Score. The study reveals that the overall accuracy of the HEMS Benefit Score based prognosis of critically ill patients is moderate. Yet one-third of the patients considered futile survived to 1 month and one-fourth of patients considered futile survived for 3 years, respectively. Based on the findings in study III, both patients considered futile and non-futile were treated with a similar intensity. In addition, there was no difference in critical interventions between patients considered futile and those considered non-futile. Even if assessed as futile, the futile group established no difference in the number of endotracheal intubations or in the use of vasoactive drugs compared to the non-futile group. Explanation for this equally intense care could be that HEMS clinicians are aware of the uncertainty of prehospital prognostication. Nonetheless, a study performed on the same national HEMS services showed that prehospital physicians and paramedics do limit treatment already in the prehospital phase if a patient is considered futile due to a serious medical incident and has deteriorated physical and/or mental performance (Kangasniemi et al. 2019). In addition to awareness of the uncertain nature of a prehospital setting, one explanation for the study III outcome of equally intense care in futile and non-futile subgroups might be the possibility that some of the patients were treated actively as potential organ donors. Moreover, patients like OHCA or trauma have well-described variables which influence the patient outcome, such as: time delays, patient age, primary rhythm, and type of trauma (Zwingmann et al. 2012, Martinell et al. 2017). Opposite to this, identifying futile neurological patients is difficult because of the heterogeneity in the causes of neurological symptoms. In addition, patients given a futile prognosis initially survive, but their neurological recovery and performance status might be heavily compromised. The main outcome in study III was that a prehospital

prognosis made by clinical evaluation should be handled with great caution. Nevertheless, it is also important to recall that the study data was gathered in a single country. Therefore, different perspectives, cultural characteristics, and clinical practices could disturb the generalisability of the results.

6.3 Revision of the HEMS Benefit Score

The HEMS Benefit Score was presented as early as 1997 and has since been used in the Finnish HEMS units. In study IV, this scoring system was revised to meet the modern standards of prehospital emergency service. The initial configuration of nine-level scoring categories with original descriptions was kept intact, but the examples of interventions included in each category were completely renewed by using the Delphi method. During this process, the name of the scoring system was modernised from the HEMS Benefit Score to the EMS Benefit Score to encourage non-HEMS units to use it, even though the HBS has always measured the benefit of the entire EMS system to a patient despite the original name. The revised name, the EBS or the EMS Benefit Score, signifies this central aspect of this scoring system, and encourages non-HEMS units to exploit the EBS in their service.

The revised EBS includes prehospitally performed intervention examples, which are divided into EBS categories. These examples do not cover the entire field or constitute a complete description of all interventions performed by paramedics and prehospital physicians, but the examples include the most typical cases. These examples are not listed to order or impose how a single patient should be scored, but to guide and assist the clinician in the process of selecting the most adequate EBS category for a patient. As is stated in the EBS chart: “Interventions are listed as examples for each category, decision of the correct EBS is made by prehospital clinician in charge of the patient care”.

The EBS emphasises the interventions that are accomplished prehospitally and assesses the impact of these procedures for the patient treated. There are other scores and classifications used in a prehospital setting, but these are seldom originally developed for prehospital use, or their reliability in a prehospital setting is questioned. The most often used score in a prehospital setting is the NACA score, which has been shown to present reliability (Raatiniemi et al. 2013 and 2017). However, the NACA score does not evaluate the benefit of the EMS or actual interventions performed prehospitally; instead, it is a score describing and categorising the medical state of the patient. A single scoring system does not solve the lack of process control in EMS systems. But when added with other quality assessment tools, it can reinforce the intrinsic quality improvement. For instance, data on EMS unit dispatch codes and criteria can be compared between EMS Benefit Scores, and the information can be used to evaluate the actual operations on a

specific mission. In addition, the data from the EBS can be used to enhance the accurate dispatching of proper level, and appropriate number, of EMS units. Even further, the EBS data from certain geographical areas could be used for planning EMS system coverage and locating units in these areas (Pappinen et al. 2018, Røislien et al. 2018). An important aspect of locating EMS units and bases is the type and number of missions historically presented in the areas under observation.

The quality and true impact of prehospital medical care have been discussed for decades (Cairns et al. 1998, Risgaard et al. 2019), especially since prehospital care has evolved from solely patient transportation into true critical care provided in the prehospital phase. To evaluate the effectiveness of prehospital care, different quality indicators and measurement protocols have been adopted (McLean et al. 2002, Murphy et al. 2016, Haugland et al. 2017). Despite the years of discussion on quality control of prehospital emergency medical service, there is still a lack of evidence on the actual impact of prehospital care itself. In addition, only a few studies published in the 1990s focus on the implementation and outcome of this quality control (Rehn et al. 2014). The revised EMS Benefit Score could be one of the solutions, when the impact of prehospitally performed interventions is evaluated. Nevertheless, there is no consensus on the true benefit of these listed interventions in the EMS Benefit Score and when they are performed prehospitally (Fullerton et al. 2011, Bigham et al. 2015, Hagiwara et al. 2019). Even though the Delphi panelists considered these crucial, further studies should focus on the prognostic influence of the prehospitally performed medical interventions. For these studies, the data gathered in the form of the EMS Benefit Score could establish a firm basis.

Despite limitations related to Delphi method, this reduces the risk of dominant or high-profile members of the group giving extra credence over other members in an expert panel (Barret et al. 2020). The Delphi method gives panellists substantial time to express their ideas, reflect on their answers and make changes, and it avoids geographical constraints. In addition, the EBS revision process included a separate in-hospital commentary board, to control the plausible bias of solely prehospital experts determining the benefit of interventions within their expertise. As pointed out, there is lack of systematically developed prehospital scoring systems, and this EBS process presented a structured manner by Delphi method for the development and revision of a scoring system for prehospital use.

6.4 Data collection and overall system quality assurance in helicopter emergency medical services

This thesis focused on data collection and patient classification in helicopter emergency services. These are one aspect of system quality assessment and

assurance. There are several ways to evaluate and secure system quality, for example: use of quality indicators, system data collection, adverse event follow-up, customer feedback, personnel feedback, formation of standard operational protocols, and use of methods for circumstance awareness (Kruger et al. 2011, Fattah et al. 2013, Christensen et al. 2017, Howard et al. 2018, McGettigan et al. 2019). As a Nordic Emergency Medical Services project on data collection and benchmarking listed in 2018: there is a need for collection of nationwide reliable and valid data in the field of the EMS, there is a lack of common classification systems to describe reason for care in the EMS, and the criteria for urgency and priority in the emergency medical dispatch centres varies among Scandinavia (Norwegian Directorate of Health 2018). Denmark has concentrated on these tasks with their HEMS database, and the first quality study of the collected data was released in 2019 (Alstrup et al. 2019).

This thesis studied the reliability of data collection and the reliability of patient scoring and classification among Finnish HEMS operative personnel. All these topics are essential parts of system quality assessment and quality improvement in the Finnish HEMS. The fourth and final section of this thesis focused on a scoring system that evaluates the clinical content and interventions performed in EMS and HEMS missions. Scoring systems like this can be used as one of the tools for quality improvement and system control in prehospital service. However, a scoring system needs to be implemented and revised adequately before clinical use (Brown et al. 2016, 2017 and 2018). The revision was performed as part of this thesis.

The data gathered in HEMS missions is used for scientific purposes and for overall assessment of the national HEMS system; therefore, the data itself needs to be reliable. In addition, the adequate use of HEMS units for patients who can benefit from them can be considered the foundation of all HEMS actions and the basis for good quality prehospital critical care. On the contrary, using a HEMS unit when there is no added value for the patient is a waste of limited and costly resources. To improve the quality and appropriate use of the HEMS, finding effective instruments for the assessment of HEMS dispatches and the actual content of the missions is crucial.

The results of this thesis present, how the quality of prehospitally collected data has to be secured (Van der Werf et al. 2019). Whether it is operational system data, or patient related data. In the Finnish HEMS, it is the prehospital clinician responsible for the data quality. However, the findings of this thesis present how vulnerable this existing policy is, and suggest that alternative options should be at least discussed. For example, if a trained data handling officer could be used already at the registration phase, not merely for follow-up. Or if the collected variables could be gathered in a more structured or categorized form, for instance, blood pressures within certain limits, not as precise numbers. Moreover, a prehospital clinical

registry would definitely benefit from an automatized data collection, where most of the measured variables would be registered automatically, not by a person logging them into a database. This could be performed for vital parameters and time-related variables, as is already done in many in-hospital registries (Armstrong et al. 2020).

The registration of time points may be considered as simple, but the documentation is based on the individual interpretation of national guidelines, and therefore be exposed to varying personal conceptions of the definitions of specific time points which may lead to inaccurate documentation. In addition, documentation of mission coding is based on the physicians or paramedics interpretation of given definitions and instructions, which increases the risk of individual variation also in mission coding. Similar individual deviation in the understanding of guidelines could explain the variation in vital parameter documentation. For example, an instruction to perform a measurement when a patient is met, is a clear time point, but there can be marked differences in whether the second vital parameter of a patient is measured at the beginning of a prehospital intervention, at the end of it or just before transportation of a patient. Moreover, this variation among HEMS personnel how guidelines are perceived could at least partially explain the variation in prehospital patient coding. For some may have scored based on the patients' past medical history, and others based on the patients' acute status.

One important factor influencing the data collection and data reliability, when performed by operational HEMS personnel, is stress and workload. As described in a report by Finnish Institute of Occupational Health, these factors play an important role when the performance of personnel is evaluated (Finnish Institute of Occupational Health 2019). Even though the report concentrated on overall performance in HEMS operational duties, the stress and workload must have a major role in the quality of data registration and collection. This also underlines the fact, whether a separate data handling officer could be used to assure a good quality prehospital data, instead of a prehospital clinician working under significant amount of stress.

6.5 Limitations

Studies I and II were based on written fictional mission scenarios. This setting can never match a real-life prehospital mission, which contains a true patient contact. The written description of the scenarios might have been influenced by individual perceptions as described earlier. In addition, registrations based on written material in several documents may lead to more errors related to interpretation of the materials than the actual accuracy of registration. As the primary hypothesis was that there is individual variation in the data collection, the missions and patient descriptions may have been written in a way that causes more disparities. Still, it can

be assumed that studies I and II still relieved most of the inaccuracies in the registration, even though there might not be as many inaccuracies in real-life clinical registry.

In study III, the research data was not originally registered for study purposes. The analyses used recorded patient data from daily HEMS missions, and the mortality data was gathered by the Population Register Center. Still, the FinnHEMS database serves as a profuse source of real-life mission data covering all Finnish HEMS bases for several years. Furthermore, study III aimed to explore the capability of patient classification to prognosticate futility in a prehospital setting, by identifying the futility among patients considered to receive marked benefits from prehospital intervention. The study clinical registry, however, did not include data on the limitations of treatments made either beforehand or during HEMS missions. In addition, study III did not to include patients deceased before hospital arrival, and therefore no conclusions can be made on how many of these patients were estimated to be futile by the prehospital clinician. Therefore, to study prognostication ability expansively, the patients evaluated as futile but treated only in a palliative manner due to treatment limitations should be included.

Study IV was based on the Delphi method and an expert panel technique. Obvious limitations related to this method are vulnerability with respect to who is considered an expert, bias in participant selection, and the setting being utterly dependent on questionnaire design. In addition, obliviousness to reliability measurement and scientific validation of findings does exist. The intervention examples found in the study were based on the data gathered from the prehospital expert panelists. They represent therefore only the opinions of the panelists and a minor part of the prehospital field, even though the panel consisted of widely known and extremely experienced European participants. The EMS score itself was revised by the Delphi process, and no conclusions can be made on the outcome prediction capabilities of the scoring system based on study IV. Moreover, the actual scoring process with the EBS relies entirely on a subjective opinion of a person determining the score for an individual patient, and no objective evaluation for the adequacy of this decision making process is included in the EBS.

6.6 Future considerations

Prehospital prognostication of futility based on clinical evaluation and classification was one of the themes in this thesis. To identify futility as early as possible is essential when the aim is to avoid unnecessary suffering of a patient with poor or no prognosis. Yet is it also crucial in directing limited resources to those who can truly benefit from them. On the other hand, it is equally important to beware the limited information on patient background, patient will, and prior medical condition, and

thoroughly consider if futility should be identified to any extent in prehospital care. This thesis presented some realities behind this process in prehospital critical care, but future studies should focus on this topic with wider perspectives and with an international study population, and most importantly by controlling the limitations in the study setting of this thesis.

As this thesis revealed, there are major problems when scoring systems are used in a prehospital setting – especially if these scores or classifications are not initially even developed for prehospital use. One example of proper revision of a prehospital scoring system was presented in the form of the EMS Benefit Score. Still, future studies are needed, and already planned, for evaluation of this score in real-life prehospital missions. Consequently, projects have been launched to implement the EMS Benefit Score to non-HEMS units and abroad, for the score has only been used in Finnish HEMS units. As the EBS has been used on a wider scale, genuine trials on its accuracy and reliability can be launched multi-nationally and with real prehospital patient contacts.

When prehospital prognostication and outcome evaluation are combined with patient classification, the most important goal for future studies is revealed: a proper prehospital tool for performance status evaluation of prior medical incident, and how this can be used for quality-of-life prognosis already in a prehospital setting. This thesis studied ECOG in prehospital performance status evaluation, and the HBS together with clinical evaluation in prognostication, which were both shown to be poor, or at least questionable, in use. Future studies should focus on investigating and finding clinically significant prehospital factors influencing patient prognosis, and not only survival, but also quality of life. As in-hospitally, also prehospitally, the patients' prior performance status must be the key element of the prognosis, and therefore the prior performance status should be estimated prehospitally as well. Thus, the tool for prehospital prior medical incident status evaluation needs to be accurate and reliable.

Prehospital patient data can be used and combined with other registry data, when prehospital factors are screened and their relevance for patient outcome is evaluated. For instance, comorbidities registered already in the prehospital phase could be combined with patient age and severity of the incident, and used to perform a follow-up on patient survival and quality of life. This way, factors influencing post-incident quality of life could be identified and further used in prehospital decision making.

Finally, the true influence and impact of prehospital interventions are still under debate, and reliably registered data is a critical component when these interventions are studied. Even further, when data from the now revised and released EMS Benefit Score is gathered and combined with outcome follow-up, the impact of prehospital interventions can be studied in larger categories. For instance, patients receiving and EBS of 8 or 7 have undergone major prehospital interventions, and their quality of

life could be studied within this subgroup after the prehospital incident. In addition, the focus with the EBS should be on the international implementation of this prehospital benefit score. This way, the influence and performance of different EMS services could be appraised.

7 Conclusions

Conclusions of this thesis were:

1. The findings on reliability and accuracy of data collection in prehospital helicopter emergency services revealed deficits that need corrective maneuvers and re-evaluation. These discrepancies were found on vital parameter registration, ASA-PS, ECOG, ICPC-2, and the HBS. Minor inter-rater variation was found on time-related variables and mission coding.
2. The reliability of prehospital classification and prognosis among futile patients was found to be debatable, as futility was similar in both groups: those considered futile and those considered non-futile by the prehospital clinician. Even further, both groups received identical prehospital critical care. Therefore, prehospital prognostication and limitation of treatments should be considered carefully and not be based on subjective clinical evaluation.
3. Revision of the HEMS Benefit Score was performed by using the Delphi technique, which included a prehospital expert panel added with an in-hospital commentary board to obtain objectivity. As a result, the EMS Benefit Score was released.

Kiitokset

Työskentely tämän väitöskirjan parissa suoritettiin Turun yliopiston anestesiologian, tehohoidon, ensihoidon ja kivunhoidon oppiaineessa sekä Turun Yliopistollisessa keskussairaalassa vuosina 2017–2021.

Haluan kiittää kaikkia kolmea ohjaajaani tuesta tämän prosessin aikana.

Miretta Tommilan visio toimi pohjana väitöskirjakokonaisuudelle. Ilman häntä tätä väitöskirjaa ei olisi koskaan edes aloitettu. Miretan kannustava ja positiivinen asenne on kantanut osatöiden suunnitteluvaiheesta aivan projektin loppumetreille asti. Miretan apuun on voinut luottaa aina, ja ajoittain lähes epätoivoisiinkin avunpyyntöihini olen aina saanut vastauksen viipymättä.

Timo Iirolan panos tälle väitöskirjalle on ollut niin ikään erittäin tärkeä. Timon vankka kokemus ensihoidosta – sekä kliinisestä, tieteellisestä että hallinnollisesta näkökulmasta – on luonut vankan pohjan osatöille ja väitöskirjan kokonaisuudelle. Timon pedantti ja yksityiskohdatkin huomioiva asenne on toiminut tärkeänä laadunvarmistuksena ja tasapainottavana tekijänä huomioiden ohjattavan herkkä taipumus kirimiseen viimeistelyvaiheissa. Myös Timon apu on ollut aina saatavilla, ja hän on vastannut viesteihin ripeästi jopa heinäkuun helteiden keskellä viikonpäivään tai vuorokauden aikaan katsomatta.

Dosentti Jouni Nurmen kokemus ensihoidon tutkimustyöstä on kiistaton. Kun tähän yhdistetään pitkä kokemus kliinisestä ensihoitotyöstä, saadaan erinomainen väitöskirjan ohjaaja. Olen kiitollinen, että pääsin Jounin ohjaukseen, ja matkan varrella olen saanut häneltä runsaasti tukea ja oppinut paljon.

Varsinaisten ohjaajien lisäksi väitöskirjan tutkimuksiin on osallistunut monia muita tärkeitä henkilöitä. Ilman *dosentti Lasse Raatiniemen* kokemusta ja näkemystä kolme neljästä osatyöstä olisi jäänyt valmistumatta. Lasse on aktiivisesti auttanut kaikissa ongelmakohdissa ja kannustanut matkan eri vaiheissa. *Päivi Laukkanen-Nevalalla, Anna Olkinuoralla ja dosentti Ilkka Virkkusella* oli merkittävä osa kahdessa ensimmäisessä osatyössä, ja sain heiltä tärkeää oppia myös muita osatöitä varten. Päivin ohjauksessa pääsin perehtymään tilastotieteen saloihin, Annan avulla artikkelien kielellinen sujuvuus parani merkittävästi – kielitieteellinen näkemys oli

äärettömän tärkeä jo tutkimusasetelman laatimisvaiheessa. Ilkan ensihoito-osaaminen oli tärkeää tutkimusten toteutumiselle, ja hän toimi ystävällisesti myös väitöskirjan ohjausryhmän jäsenenä.

Haluan kiittää *Johannes Björkmania, Helena Jänttiä, professori Janne Liisananttia ja Merja Meriläistä*. Johannes ja Helena olivat mukana kolmannessa osatyössä, joka ei olisi toteutunut ilman heitä. Neljännen osatyön tutkimusasetelma oli haastava, ja Merjan panos oli äärettömän tärkeä kaikissa osatyön vaiheissa. Olen iloinen siitä, että saimme Jannen mukaan ryhmään; hänen laaja kokemuksensa takasi, että tutkimus saatiin julkaistavaan muotoon. Webropol-velhomme *Jarmo Määttäsen* asiantuntemus oli niin ikään edellytyksenä neljännen osatyön valmistumiselle.

Apulaisprofessori Teijo Saari toimi väitöskirjan ohjausryhmän jäsenenä. Teijo teki laajan tieteellisen ja kliinisen ensihoitokokemuksensa pohjalta tärkeitä huomioita, jotka veivät väitöskirjaa eteenpäin. Haluankin kiittää sekä Ilkkaa että Teijoa aktiivisesta osallistumisesta ohjausryhmän kautta. Haluan kiittää myös esitarkastajia, *professori Seppo Alahuhtaa* ja dosentti *Tuomas Brinckiä*, joilta sain tärkeää tukea väitöskirjan viimeistelyyn.

Haluan kiittää *FinnHEMS*:iä, jonka tutkimus- ja kehitysyksikön rahoituksen turvin sain keskittyä täysipäiväisesti tutkimuksen tekoon sen muutamissa kriittisissä vaiheissa. Haluan kiittää erikseen jokaista tukikohtaa, joiden lääkärit ja ensihoitajat osallistumisellaan mahdollistivat ensimmäisen ja toisen osatyön. Erityisesti haluan kiittää FH20 lääkäreitä, sillä heidän avullaan saimme pilotoitua neljännen osatyön tutkimusasetelman. Tämän pilotoinnin osalta haluan kiittää myös *Morten Overgaardia* ja *professori Erika Christensenia*. Mortenin kanssa suoritin lisäksi yhtä aikaa SSAI:n lasten anestesiologian ja tehohoidon lisäkoulutuksen.

Vuosina ennen väitöskirjan valmistumista, jakautui innostukseni sen ja lasten anestesiologian saloihin perehtymisen välille. Haluan kiittää *dosentti Tuula Manneria, erikoislääkäri Sanna Viloa* ja *dosentti Markku Taittosta* siitä, että olen voinut yhdistää kliinisen työn ja tutkimuksen tekemisen mielekkäällä tavalla. Kaikki kolme ovat esimiehiäni, mutta ennen kaikkea ystäviä ja lastenanestesiologikollegoita. Tuulan valvovien silmien alla olen päässyt asentamaan ensimmäistä kertaa sentraalisen kanyylin imeväiselle – tuolloin vielä erikoistuvan lääkärin vapisevin käsin. Erikoislääkäriksi valmistuttuani pääsin Sannan ja Markun ohjauksessa syventymään lasten anestesiologiaan tarkemmin. Sannan vinkkejä on tarvittu myös kotioloissa, ja niiden avulla olen löytänyt sisältäni kääpiövillakoiran kouluttajan. Markun kliiniset kädentaidot ovat omaa luokkaansa, mutta olen saanut häneltä myös kullanarvoisia vinkkejä laadukkaista skandinaavisista TV-sarjoista.

Olli Vänttinen ja *Mari Fihlman* ovat olleet niin ikään tärkeitä henkilöitä polullani lastenanestesiologiksi. Molempien tukeen voi luottaa kaikissa tilanteissa, ja olen oppinut heiltä paljon. Erityisesti arvostan molempien äärimmilleen viritettyä taitoa

asiakeskeiseen palautteen antoon ilman turhia korulauseita, ja tästä palautteenannosta saankin nauttia päivittäin. Koko perheeni on saanut kokea Ollin vieraanvaraisuuden Vänttisten mökillä, jonne odotamme kutsua myös tulevana kesänä (omia ruokia emme tuo tälläkään kertaa). Ilman Marin apua en olisi selvinnyt väitöskirjan loppuun saattamisesta, ja hänen tukensa tohtorintutkinnon käytännön asioiden hoitamisessa on ollut korvaamaton.

Marin lisäksi *Kosti Koivisto-Kokko* on ollut tärkein innoittajani lähteä erikoistumaan lastenanestesiologiaan. Kosti on ollut tärkeä tukihenkilö ulkomaan klinikkavierailuillani, ja näyttänyt miten ulkomailla työskentelystä ei tarvitse pelkästään haaveilla; haaveet voi myös toteuttaa. Kosti ja Olli osoittavat lisäksi esimerkillään, miten ensihoidon ja lastenanestesiologian osaamisen voi yhdistää kliinisessä työssä.

Monna Myllykangas on tuorein lastenanestesiologikollega, jonka kanssa olen päässyt ventiloimaan väitöskirjan valmistumisen tuskaa. Olemme lasten leikkausosastolla usein tukeutuneet Monnan korvaamattomaan osaamiseen hengitysvajauspotilaiden hoidossa, ja häneltä olen saanut tärkeitä työkaluja myös lasten tehoosastolla työskentelyyn.

Tomi Vainiolle haluan varata oman kappaleen. Tomin kokemus ja näkemys anestesiologiasta on toiminut innoittajana useille klinikkamme nuoremman polven kollegoille. Itse kuulun myös tähän "Tomin talliin", ja vuonna 2010 tullessani keltanokkana taloon sain aivan ensimmäiset oppini Tomilta. Edelleen vaikeissa tilanteissa ja niihin valmistautuessa tulee usein konsultoitua Tomia.

Haluan osoittaa kiitokset muillekin anestesiakollegoilleni Tyksissä, kuten myös sekä Totek-toimialueen muille ammattilaisille, toki erityisesti lasten leikkausosaston ja lasten teho-osaston henkilökunnalle. Haluan kiittää myös talomme lastenkirurgeja ja pediatrian yksiköiden henkilökuntaa. Erityiskiitokset *Ulla Ahlmen-Laiholle*, joka varmisti äidinkielen sujuvuuden ennen väitöskirjan painoon lähettämistä.

Kiitos *Suomen Anestesiologiyhdistykselle, Ensihoidon Tukisäätiölle* ja *Turun Yliopistolle* apurahoista, joiden avulla tämä väitöskirja on valmistunut.

Lopuksi haluan kiittää perhettäni – *Kirsiä, Siiriä ja Emiliaa*. Ilman heidän tukeaan tämä väitöskirja ei olisi valmistunut. He ovat pyyteettömästi tukeneet minua niin Kemijärven kaamoksessa kuin Kapkaupungin auringonpaisteessa. Suomalainen mies ei turhia tunteistaan puhu, mutta nyt suurin kiitos on lausuttava heille ääneen. Kiitos myös vanhemmilleni ja sisaruksilleni, joilta olen saanut tukea kaikissa elämäni käänteissä.

Turussa maaliskuussa 2021
Anssi Heino

References

Aarvold ABR, Ryan HM, Magee LA, Von Dadelszen P, Fjell C, Walley KR. Multiple Organ Dysfunction Score Is Superior to the Obstetric-Specific Sepsis in Obstetrics Score in Predicting Mortality in Septic Obstetric Patients. *Crit Care Med*. 2017;45(1):e49–e57.

Aitavaara-Anttila M, Liisanantti J, Ehrola A, Spalding M, Ala-Kokko T, Raatiniemi L. Use of prehospital emergency medical services according to income of residential area. *Emerg Med J.* 2020 Online ahead of print.

Alstrup K, Petersen JAK, Barfod C, Knudsen L, Rognås L, Møller TP. The Danish helicopter emergency medical service database: high quality data with great potential. *Scand J Trauma Resusc Emerg Med.* 2019;27(1).

American Society of Anaesthesiologists, ASA House of Delegates/Executive Commitee. ASA physical Status Classification System. https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system.

Arabian SS, Marcus M, Captain K, et al. Variability in interhospital trauma data coding and scoring: A challenge to the accuracy of aggregated trauma registries. *Journal of Trauma and Acute Care Surgery.* Vol 79. Lippincott Williams and Wilkins; 2015:359–363.

Armstrong RA, Mouton R, Hinchliffe RJ. Routinely Collected data and patient-centred research in anaesthesia and peri-operative care: a narrative review. *Anaesthesia.* 2020 Online ahead of print

Arts DGT, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: A literature review, case study, and generic framework. *J Am Med Informatics Assoc.* 2002;9(6):600–611.

Australian Commission on Safety and Quality in Health Care. Framework for Australian clinical quality registries. Sydney: ACSQHC; 2014.

Barret D and Heale R. What are Delphi studies? *Evidence Based Nursing* 2020; 23: 3.

Basílio N, Ramos C, Figueira S, Pinto D. Worldwide Usage of International Classification of Primary Care. *Rev Bras Med Família e Comunidade*. 2016;11(38):1–9.

Bigham B, Welsford M. Applying hospital evidence to paramedicine: Issues of indirectness, validity and knowledge translation. *Can J Emerg Med.* 2015;17(3):281–285.

Black JJM, Davies GD. International EMS Systems: United Kingdom. *Resuscitation.* 2005; 64:21–29

Bledsoe BE, Casey MJ, Feldman J, et al. Glasgow Coma Scale Scoring is Often Inaccurate. *Prehosp Disaster Med.* 2015;30(1):46–53.

Bossers SM, Boer C, Greuters S, et al. Dutch Prospective Observational Study on Prehospital Treatment of Severe Traumatic Brain Injury: The BRAIN-PROTECT Study Protocol. *Prehospital Emerg Care.* 2019;23(6):820–827.

Bouzat P, Legrand R, Gillois P, et al. Prediction of intra-hospital mortality after severe trauma: Which pre-hospital score is the most accurate? *Injury*. 2016;47(1):14–18.

Brown JB, Gestring ML, Guyette FX, et al. Development and Validation of the Air Medical Prehospital Triage Score for Helicopter Transport of Trauma Patients. *Ann Surg.* 2016;264(2):378–385.

Brown JB, Gestring ML, Guyette FX, et al. External validation of the Air Medical Prehospital Triage score for identifying trauma patients likely to benefit from scene helicopter transport. *Journal of Trauma and Acute Care Surgery*. Vol 82. Lippincott Williams and Wilkins; 2017:270–279.

Brown JB, Smith KJ, Gestring ML, et al. Comparing the air medical prehospital triage score with current practice for triage of injured patients to helicopter emergency medical services: A cost-effectiveness analysis. *JAMA Surg*. 2018;153(3):261–268.

Cairns CB, Garrison HG, Hedges JR, Schriger DL, Valenzuela TD. Development of New Methods to Assess the Outcomes of Emergency Care. *Ann Emerg Med*. 1998;31(2):166–171.

Challen K, Roland D. Early warning scores: A health warning. *Emerg Med J*. 2016;33(11):812–817.

Christensen EF, Bendtsen MD, Larsen TM, 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(8).

Coi A, Santoro M, Villaverde-Hueso A, et al. The quality of rare disease registries: Evaluation and characterization. *Public Health Genomics*. 2016;19(2):108–115.

Cone DC. Knowledge Translation in the Emergency Medical Services: A Research Agenda for Advancing Prehospital Care. *Acad Emerg Med*. 2007;14(11):1052–1057.

Darioli V, Taffé P, Carron PN, et al. Evaluation of the discriminative performance of the prehospital National Advisory Committee for Aeronautics score regarding 48-h mortality. *Eur J Emerg Med.* 2019;26(5):366–372.

de Grooth HJ, Girbes AR, Loer SA. Early warning scores in the perioperative period: applications and clinical operating characteristics. *Curr Opin Anaesthesiol*. 2018;31(6):732–738.

Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol*. 2014;67(4):401–409.

Dreyer NA, Garner S. Registries for Robust Evidence. *JAMA*. 2009; 19;302(7):790–1.

Epstein RH, Dexter F, O'Neill L. Development and Validation of an Algorithm to Classify as Equivalent the Procedures in ICD-10-PCS That Differ only by Laterality. *Anesth Analg.* 2019;128(6):1138–1144.

Evans R, McGovern R, Birch J, Newbury-Birch D. Which extended paramedic skills are making an impact in emergency care and can be related to the UK paramedic system? A systematic review of the literature. *Emerg Med J*. 2014;31(7):594–603.

Evans SE, Scott IA, Johnson NP, Cameron PA and McNeil JJ. Development of clinical-quality registries in Australia: the way forward. *Medical Journal of Australia.* 2011; 194: 360–363.

Fattah S, Rehn M, Lockey D, Thompson J, Lossius HM, Wisborg T. A consensus based template for reporting of pre-hospital major incident medical management. *Scand J Trauma Resusc Emerg Med.* 2014;22(1).

Fattah S, Rehn M, Reierth E, Wisborg T. Systematic literature review of templates for reporting prehospital major incident medical management. *BMJ Open.* 2013;3: e002658.

Ferrand E, Marty J. Prehospital withholding and withdrawal of life-sustaining treatments. The French LATASAMU survey. *Intensive Care Med*. 2006;32(10):1498–1505.

Fink A, Kosecoff J, Brook RH. Consensus Methods: Characteristics and Guidelines for Use. *Am J Public Health.* 1984; 74(9): 979–983.

Finnish Institute of Occupational Health report 2019. Well-being at work among helicopter emergency medical service personnel in Finland.

Friberg ML, Rognås L. Patient-tailored triage decisions by anaesthesiologist-staffed prehospital critical care teams: A retrospective descriptive study. *BMJ Open*. 2018;8(7).

Fukuda T, Ohashi-Fukuda N, Kondo Y, Hayashida K, Kukita I. Association of prehospital advanced life support by physician with survival after out-of-hospital cardiac arrest with blunt trauma following traffic collisions Japanese registry-based study. *JAMA Surg*. 2018;153(6).

Fullerton JN, Roberts KJ, Wyse M. Should non-anaesthetists perform pre-hospital rapid sequence induction? An observational study. *Emerg Med J.* 2011;28(5):428–431.

Garner AA, Barker CL, Weatherall AD. Retrospective evaluation of prehospital triage, presentation, interventions and outcome in paediatric drowning managed by a physician staffed helicopter emergency medical service. *Scand J Trauma Resusc Emerg Med*. 2015;23(1).

Hagiwara MA, Magnusson C, Herlitz J, et al. Adverse events in prehospital emergency care: A trigger tool study. *BMC Emerg Med*. 2019;19(1).

Haniffa R, Isaam I, De Silva AP, Dondorp AM, De Keizer NF. Performance of critical care prognostic scoring systems in low and middle-income countries: A systematic review. *Crit Care*. 2018;22(1).

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(4):602–609.

Hartmann-Johnsen OJ, Kåresen R, Schlichting E, Naume B, Nygård JF. Using clinical cancer registry data for estimation of quality indicators: Results from the Norwegian breast cancer registry. *Int J Med Inform.* 2019; 125:102–109.

Hasson FB, Keeney SB and McKenna H. Research Guidelines for the Delphi Survey Technique. *Journal of Advanced Nursing.* 2000; 32(4): 1008–1015.

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(1).

Hoikka M, Länkimäki S, Silfvast T, Ala-Kokko TI. Medical priority dispatch codes-comparison with National Early Warning Score. *Scand J Trauma Resusc Emerg Med*. 2016;24(1).

Hoikka M, Silfvast T, Ala-Kokko TI. Does the prehospital National Early Warning Score predict the short-term mortality of unselected emergency patients? *Scand J Trauma Resusc Emerg Med*. 2018;26(1).

Hoogendoorn ME, Margadant CC, Brinkman S, Haringman JJ, Spijkstra JJ, de Keizer NF. Workload scoring systems in the Intensive Care and their ability to quantify the need for nursing time: A systematic literature review. *Int J Nurs Stud.* 2020;101.

Hoque DME, Kumari V, Hoque M, Ruseckaite R, Romero L, Evans SM. Impact of clinical registries on quality of patient care and clinical outcomes: A systematic review. *PLoS One*. 2017;12(9).

Hoque DME, Ruseckaite R, Lorgelly P, Mcneil JJ, Evans SM. Cross-sectional study of characteristics of clinical registries in Australia: A resource for clinicians and policy makers. *Int J Qual Heal Care.* 2018;30(3):192–199.

Horton EE, Krijnen P, Molenaar HM. Are the registry data reliable An audit of a regional trauma registry in the Netherlands. *Int J Qual Health Care*. 2017; 1;29(1):98–103.

Howard I, Cameron P, Wallis L, Castren M, Lindstrom V. Quality Indicators for Evaluating Prehospital Emergency Care: A Scoping Review. *Prehosp Disaster Med*. 2018;33(1).

Ihejirika RC, Thakore R V., Sathiyakumar V, Ehrenfeld JM, Obremskey WT, Sethi MK. An assessment of the inter-rater reliability of the ASA physical status score in the orthopaedic trauma population. *Injury.* 2015;46(4):542–546.

Jouffroy R, Saade A, Ellouze S, et al. Prehospital triage of septic patients at the SAMU regulation: Comparison of qSOFA, MRST, MEWS and PRESEP scores. *Am J Emerg Med*. 2018;36(5):820–824.

Jouffroy R, Pierre Tourtier J, Gueye P, et al. Prehospital shock index to assess 28-day mortality for septic shock. *Am J Emerg Med*. 2020;38(7):1352–1356.

Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliat Med*. 2017;31(8):684–706.

Kangasniemi H, Setälä P, Huhtala H, et al. Limitation of treatment in prehospital care – The experiences of helicopter emergency medical service physicians in a nationwide multicentre survey. *Scand J Trauma Resusc Emerg Med*. 2019;27(1).

Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE—Acute physiology and chronic health evaluation: A physiologically based classification system. *Crit Care Med*. 1981;9:591–597.

Koenraad GM, Nolan JP, Bossaet LL, et al. European Resuscitation Council Guidelines for Resuscitation 2015 Section 1. Executive summary. *Resuscitation.* 95 (2015): 1–80.

Krüger AJ, Skogvoll E, Castren M, et al. Scandinavian pre-hospital physician-manned Emergency Medical Services – Same concept across borders? *Resuscitation.* 81 (2010): 427–433.

Krüger AJ, Lossius HM, Mikkelsen S, Kurola J, Castrén 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):1175–1185.

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.

Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score – Development, utility and challenges of accurate assessment in clinical trials. *Crit Care*. 2019;23(1):1–9.

Lee P, Chin K, Liew D, et al. Economic evaluation of clinical quality registries: a systematic review. *BMJ Open*. 2019; 15;9(12):e030984.

Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Crit Care Med.* 1984; 12:975–977.

Lemeshow S, Teres D, Avrunin JS, Pastides H. A comparison of methods to predict mortality of intensive care unit patients. *Crit Care Med.* 1987; 15:715–722.

Lysholm J, Lindahl B. Strong development of research based on national quality registries in Sweden. *Ups J Med Sci*. 2019;124(1):9–11.

Lyu H, Cooper M, Patel K, Daniel M, Makary MA. Prevalence and Data Transparency of National Clinical Registries in the United States. *J Healthc Qual*. 2016;38(4):223–34.

Länkimäki S, Alahuhta S, Silfvast T, et al. Feasibility of LMA Supreme for airway management in unconscious patients by ALS paramedics. *Scand J Trauma Resusc Emerg Med* 2015, 23(1).

MacCallum C, Skandarajah A, Gibbs P, Hayes I. The value of clinical colorectal cancer registries in colorectal cancer research a systematic review. *JAMA Surg.* 2018;153(9):841–849.

MacFarlane C, Van Loggerenberg C, Kloeck W. International EMS systems in South Africa – Past, present, and future. *Resuscitation*. 2005;64(2):145–148.

Maduz R, Kugelmeier P, Meili S, Döring R, Meier C, Wahl P. Major influence of interobserver reliability on polytrauma identification with the Injury Severity Score (ISS): Time for a centralised coding in trauma registries? *Injury.* 2017;48(4):885–889.

Maret-Ouda J, Tao W, Wahlin K, Lagergren J. Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. *Scand J Public Health*. 2017;45(17_suppl):14–19.

Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Crit Care Med*. 1995;23(10):1638–52.

Martinell L, Nielsen N, Herlitz J, et al. Early predictors of poor outcome after out-of-hospital cardiac arrest. *Crit Care*. 2017;21(1).

McGettigan P, Alonso Olmo C, Plueschke K, et al. Patient Registries: An Underused Resource for Medicines Evaluation: Operational proposals for increasing the use of patient registries in regulatory assessments. *Drug Saf*. 2019;42(11):1343–1351.

McHugh ML. Lessons in biostatistics interrater reliability : the kappa statistic. *Biochem Medica*. 2012;22(3):276–282.

McLean SA, Maio RF, Spaite DW, Garrison HG. Emergency medical services outcomes research: Evaluating the effectiveness of prehospital care. *Prehospital Emerg Care*. 2002;6(2 SUPPL.).

McMillan T, Wilson L, Ponsford J, Levin H, Teasdale G, Bond M. The Glasgow Outcome Scale – 40 years of application and refinement. *Lancet Neurol*. 2014 Aug;13(8):844–54.

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;7:14383.

Mills EHA, Aasbjerg K, Hansen SM, et al. Prehospital time and mortality in patients requiring a highest priority emergency medical response: A Danish registry-based cohort study. *BMJ Open*. 2019;9(11).

Ministry of Social Affairs and Health, Finland N.B. No. 785/1992 Act on the status and rights of patients. Legally valid only in Finnish and Swedish.

Ministry of Social Affairs and Health, Finland N.B. No. 1326/2010 Health Care Act Issued in Helsinki on 30 December 2010. Legally binding only in Finnish and Swedish.

Miyamoto K, Shibata N, Ogawa A, Nakashima T, Kato S. Prehospital quick sequential organ failure assessment score to predict in-hospital mortality among patients with trauma. *Am J Emerg Med.* 2019;37(12):2165–2170.

Monash Clinical Registries Portfolio 2018, Registry Science and Research Epidemiology and Preventive Medicine Monash Public Health and Preventive Medicine. Monash University, Australia. https://www.monash.edu/__data/assets/pdf_file/0018/2220840/monash-clinical-registries-portfolio-2018.pdf.

Mowafi H, Ngaruiya C, O'Reilly G, et al. Emergency care surveillance and emergency care registries in low-income and middle-income countries: Conceptual challenges and future directions for research. *BMJ Glob Heal*. 2019;4.

Mulder DS, Spicer J. Registry-Based Medical Research: Data Dredging or Value Building to Quality of Care? *Ann Thorac Surg*. 2019;108(1):274–282.

Murphy A, Wakai A, Walsh C, Cummins F, O'Sullivan R. Development of Key performance indicators for prehospital emergency care. *Emerg Med J.* 2016;33(4):286–292.

Nakada TA, Masunaga N, Nakao S, et al. Development of a prehospital vital signs chart sharing system. *Am J Emerg Med*. 2016;34(1):88–92.

Nannan Panday RS, Minderhoud TC, Alam N, Nanayakkara PWB. Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): A narrative review. *Eur J Intern Med.* 2017;45:20–31.

Nichol G, Thomas E, Callaway CW, et al. Regional variation in out-of-hospital cardiac arrest incidence and outcome. *JAMA*. 2008;300(12):1423–1431.

Nishiyama C, Brown SP, May S, et al. Apples to apples or apples to oranges? International variation in reporting of process and outcome of care for out-of-hospital cardiac arrest. *Resuscitation.* 2014;85(11):1599–1609.

Okubo M, Schmicker RH, Wallace DJ, et al. Variation in Survival after Out-of-Hospital Cardiac Arrest between Emergency Medical Services Agencies. *JAMA Cardiol*. 2018;3(10):989–999.

Olthof DC, Luitse JSK, De Groot FMJ, Goslings JC. A Dutch regional trauma registry: Quality check of the registered data. *BMJ Qual Saf.* 2013 Sep;22(9):752–8.

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(1).

Pakkanen T, Nurmi J, Huhtala H, Silfvast T. Prehospital on-scene anaesthetist treating severe traumatic brain injury patients is associated with lower mortality and better neurological outcome. *Scand J Trauma Resusc Emerg Med.* 2019;27(1).

Pakkanen T, Virkkunen I, Kämäräinen A, et al. Pre-hospital severe traumatic brain injury – comparison of outcome in paramedic versus physician staffed emergency medical services. *Scand J Trauma Resusc Emerg Med.* 2016;24(1).

Pakkanen T, Virkkunen I, Kämäräinen A, et al. Severe traumatic brain injury- physician provided pre-hospital care and early neurosurgical treatment are associated with improved survival. *Scand J Trauma Resusc Emerg Med.* 2015;23(2).

Pappinen J, Laukkanen-Nevala P, Mäntyselkä P, Kurola J. Development and implementation of a geographical area categorisation method with targeted performance indicators for nationwide EMS in Finland. *Scand J Trauma Resusc Emerg Med*. 2018;26(1).

Pappinen J, Olkinuora A, Laukkanen-Nevala P. Defining a mission-based method to determine a HEMS unit's actual service area. *Scand J Trauma Resusc Emerg Med*. 2019;27(1).

Patel R, Nugawela MD, Edwards HB, et al. Can early warning scores identify deteriorating patients in pre-hospital settings? A systematic review. *Resuscitation*. 2018;132:101–111.

Perkins GD, Olasveengen TM, Maconochie I, et al. European Resuscitation Council Guidelines for Resuscitation: 2017 update. *Resuscitation*. 2018;123:43–50.

Persad G, Wertheimer A, Emanuel EJ. Principles for allocation of scarce medical interventions. *Lancet*. 2009;373(9661):423–431.

Polit D and Beck C: Nursing Research – principles and methods. Lippincott, Williams& Wilkins (2004) A Wolters Kluwer Company.

Polit D, Beck C, Owen S, et al: Focus on Research Methods. Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing & Health*. 2007;30:459–467.

Pollock AM, Vickers N. Reliability of Data of the Thames Cancer Registry on 673 Cases of Colorectal Cancer: Effect of the Registration Process. *Qual Health Care*. 1995;4(3):184–9.

Porgo TV, Moore L, Tardif PA. Evidence of data quality in trauma registries: A systematic review. *J Trauma Acute Care Surg.* 2016;80(4):648–658.

Porter ME. What Is Value in Health Care? *NEJM.* 2010: 363;26.

Pozner CN, Zane R, Nelson SJ, Levine M. International EMS Systems: The United States: Past, present, and future. *Resuscitation*. 2004;60(3):239–244.

Raatiniemi L, Liisanantti J, Tommila M, et al. Evaluating helicopter emergency medical missions: a reliability study of the HEMS benefit and NACA scores. *Acta Anaesthesiol Scand.* 2017;61(5):557–565.

Raatiniemi L, Mikkelsen K, Fredriksen K, Wisborg T. Do pre-hospital anaesthesiologists reliably predict mortality using the NACA severity score? A retrospective cohort study. *Acta Anaesthesiol Scand*. 2013;57(10):1253–1259.

Raatiniemi L, Liisanantti J, Niemi S, et al. Short-term outcome and differences between rural and urban trauma patients treated by mobile intensive care units in Northern Finland: A retrospective analysis. *Scand J Trauma Resusc Emerg Med*. 2015;23(1).

Raina KD, Callaway C, Rittenberger JC, Holm MB. Neurological and functional status following cardiac arrest: method and tool utility. *Resuscitation*. 2008 Nov;79(2):249–56.

Randolph JJ. Free-marginal multirater kappa (multirater kfree): an alternative to Fleiss' fixed-marginal multirater kappa. Joensuu: Joensuu University Learning and Instruction Symposium; 2005.

Ranta S, Hynynen M, Tammisto T. A survey of the ASA physical status classification: significant variation in allocation among Finnish anaesthesiologists. *Acta Anaesthesiol Scand*. 1997;629–632, 41.

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(1).

Ringdal KG, Skaga NO, Hestnes M, et al. Abbreviated Injury Scale: Not a reliable basis for summation of injury severity in trauma facilities? *Injury.* 2013;44(5):691–699.

Ringdal KG, Skaga NO, Steen PA, et al. Classification of comorbidity in trauma: The reliability of pre-injury ASA physical status classification. *Injury*. 2013;44(1):29–35.

Ringdal KG, Coats TJ, Lefering R, et al. The Utstein template for uniform reporting of data following major trauma: A joint revision by SCANTEM, TARN, DGU-TR and RITG. *Scand J Trauma Resusc Emerg Med.* 2008;28;16:7.

Ringdal K, Lossius H, Jones JM, et al. Collecting core data in severely injured patients using a consensus trauma template: an international multicentre study. *Crit Care*. 2011;15(5):R237.

Risgaard B, Draegert C, Baekgaard J, et al. Impact of Physician-staffed Helicopters on Pre-hospital Patient Outcomes: A systematic review. *Acta Anaesthesiol Scand.* 2020; 691–704, 64(5).

Rittenberger JC, Raina K, Holm MB, Kim YJ, Callaway CW. Association between Cerebral Performance Category, Modified Rankin Scale, and discharge disposition after cardiac arrest. *Resuscitation.* 2011;82(8):1036–40.

Roudsari BS, Nathens AB, Arreola-Risa C, et al. Emergency Medical Service (EMS) systems in developed and developing countries. *Injury.* 2007;38(9):1001–1013.

Røislien J, van den Berg PL, Lindner T, et al. Comparing population and incident data for optimal air ambulance base locations in Norway. *Scand J Trauma Resusc Emerg Med*. 2018;26(1).

Salluh JIF, Soares M. ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care.* 2014;20(5):557–565.

Sankar A, Johnson SR, Beattie WS, Tait G, Wijeysundera DN. Reliability of the American Society of Anesthesiologists physical status scale in clinical practice. *Br J Anaesth*. 2014;113(3):424–32.

Sartorius D, Le Manach Y, David JS, et al. Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure (MGAP): A new simple prehospital triage score to predict mortality in trauma patients. *Crit Care Med*. 2010;38(3):831–837.

Saviluoto A, Björkman J, Olkinuora A, et al. The first seven years of nationally organized helicopter emergency medical services in Finland – The data from quality registry. *Scand J Trauma Resusc Emerg Med.* 2020;28(1).

Schewe JC, Kappler J, Dovermann K, et al. Diagnostic accuracy of physician-staffed emergency medical teams: A retrospective observational cohort study of prehospital versus hospital diagnosis in a 10-year interval. *Scand J Trauma Resusc Emerg Med*. 2019;27(1).

Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National patient registry: A review of content, data quality, and research potential. *Clin Epidemiol.* 2015;7:449–490.

Schneider F, Martin J, Schneider G, Schulz CM. The impact of the patient's initial NACA score on subjective and physiological indicators of workload during pre-hospital emergency care. *PLoS One.* 2018;13(8).

Shida H, Matsuyama T, Kiyohara K, et al. Prehospital cardiopulmonary resuscitation duration and neurological outcome after out-of-hospital cardiac arrest among children by location of arrest: A Nationwide cohort study. *Scand J Trauma Resusc, Emerg Med*. 2019;27(1).

Shin S Do, Hock Ong ME, Tanaka H, et al. Comparison of emergency medical services systems across Pan-Asian countries: A web-based survey. *Prehospital Emerg Care*. 2012;16(4):477–496.

Smith RM, Conn AK. Prehospital care – Scoop and run or stay and play? *Injury.* 2009;40(SUPPL. 4)

Sollid SJM, Lockey D, Lossius HM, et al. A consensus-based template for uniform reporting of data from pre-hospital advanced airway management. *Scand J Trauma, Resusc Emerg Med.* 2009; 17(58).

Stey AM, Russell MM, Ko CY, Sacks GD, Dawes AJ, Gibbons MM. Clinical registries and quality measurement in surgery: A systematic review. *Surg (United States)*. 2015;157(2):381–395.

Sun KM, Song KJ, Shin S Do, et al. Comparison of Emergency Medical Services and Trauma Care Systems Among Pan-Asian Countries: An International, Multicenter, Population-Based Survey. *Prehospital Emerg Care.* 2017;21(2):242–251.

Sunde GA, Kottmann A, Heltne JK, et al. Standardised data reporting from pre-hospital advanced airway management – a nominal group technique update of the Utstein-style airway template. *Scand J Trauma Resusc Emerg Med*. 2018;26(1).

Tanigawa K, Tanaka K. Emergency medical service systems in Japan: Past, present, and future. *Resuscitation.* 2006;69(3):365–370.

Taylor C, Jan S, Curtis K, et al. The cost-effectiveness of physician staffed Helicopter Emergency Medical Service (HEMS) transport to a major trauma centre in NSW, Australia. *Injury.* 2012;43(11):1843–1849.

The Nordic Emergency Medical Services, project on data collection and benchmarking, vol. IS-2750. Helsedirektoratet: Norwegian Directorate of Health, Report, Ordering NR; 2014–2018. https://www.landlaeknir.is/servlet/file/store93/item36845/Nordic%20EMS%20benchmarking%20report%202014-2018.pdf.

Thomson N. Emergency medical services in Zimbabwe. *Resuscitation.* 2005;65(1):15–19.

Trimmel H, Beywinkler C, Hornung S, Kreutziger J, Voelckel WG. In-hospital airway management training for non-anesthesiologist EMS physicians: A descriptive quality control study. *Scand J Trauma Resusc Emerg Med.* 2017;25(1).

Tryba M, Briiggemann H, Echtermeyer V. Klassifm- ierung von Erkrankungen und Verletzungen im Notarz- trettungssystem. *Notfalhned*. 1980; 8: 725.

Tusgul S, Carron PN, Yersin B, Calandra T, Dami F. Low sensitivity of qSOFA, SIRS criteria and sepsis definition to identify infected patients at risk of complication in the prehospital setting and at the emergency department triage. *Scand J Trauma Resusc Emerg Med.* 2017;25(1)

Vaitkaitis D. EMS systems in Lithuania. *Resuscitation*. 2008;76(3):329–332.

Van der Velden MWA, Ringburg AN, Bergs EA, Steyerberg EW, Patka P, Schipper IB. Prehospital interventions: Time wasted or time saved? An observational cohort study management in initial trauma care. *Emerg Med J*. 2008;25(7):444–449.

Van der Werf LR, Voeten SC, van Loe CMM, Karthaus EG, Wouters MWJM, Prins HA. Data verification of nationwide clinical quality registries. *BJS open*. 2019;3(6):857–864.

Van Gelder CM, Frantz R, Bogucki S. Emergency medical services in Connecticut. *Prehospital Emerg Care.* 2005;9(2):219–226.

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(6):384–390.

Van Schuppen H, Bierens J. Understanding the prehospital physician controversy. Step 1: Comparing competencies of ambulance nurses and prehospital physicians. *Eur J Emerg Med.* 2011;18(6):322–327.

Wang IJ, Bae BK, Park SW, et al. Pre-hospital modified shock index for prediction of massive transfusion and mortality in trauma patients. *Am J Emerg Med*. 2020;38(2):187–190.

WHO, The World Health Organization. 2010. International Statistical Classification of Diseases and Related Health Problems. 10th edition. Geneva.

Williams TA, Tohira H, Finn J, Perkins GD, Ho KM. The ability of early warning scores (EWS) to detect critical illness in the prehospital setting: A systematic review. *Resuscitation.* 2016;102:35–43.

Vincent JL, Moreno R, Takala J, Willatts S, Mendonca A, Bruining H. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996;22(7):707–10.

Vincent J, Moreno R. Scoring systems in the critically ill. *Crit Care*. 2010;14:1–9.

Wisborg T, Ellensen EN, Svege I, Dehli T. Are severely injured trauma victims in Norway offered advanced pre-hospital care? National, retrospective, observational cohort. *Acta Anaesthesiol Scand*. 2017;61(7):841–847.

WONCA: The World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians. International Classification of Primary Care, second edition (ICPC-2). https://www.globalfamilydoctor.com/groups/WorkingParties/wicc.aspx.

Zwingmann J, Mehlhorn AT, Hammer T, Bayer J, Südkamp NP, Strohm PC. Survival and neurologic outcome after traumatic out-of-hospital cardiopulmonary arrest in a pediatric and adult population: A systematic review. *Crit Care.* 2012 Jul 6;16(4):R117.

Østerås Ø, Heltne JK, Vikenes BC, Assmus J, Brattebø G. Factors influencing on-scene time in a rural Norwegian helicopter emergency medical service: A retrospective observational study. *Scand J Trauma Resusc Emerg Med.* 2017;25(1).

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama Oy, Turku, Finland 2021

ISBN 978-951-29-8412-1 (PRINT)
ISBN 978-951-29-8413-8 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)