Tampere University

JOONAS TAMMINEN

Prehospital Emergencies

Early Recognition, Response and Machine Learning in Risk Stratification

Tampere University Dissertations 457

Tampere University Dissertations 457

JOONAS TAMMINEN

Prehospital Emergencies

Early Recognition, Response and Machine Learning in Risk Stratification

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty Council of the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the Jarmo Visakorpi auditorium
of the Arvo Building, Arvo Ylpön katu 34, Tampere,
on 17 September 2021, at 12 o’clock.

ACADEMIC DISSERTATION
Tampere University, the Faculty of Medicine and Health Technology
Tampere University Hospital, Centre for Pre-Hospital Emergency Care
Finland

Responsible supervisor	Docent Sanna Hoppu Tampere University Finland	
Supervisor	Docent Antti Kämäräinen Tampere University Finland	
Pre-examiners	Associate Professor Timo Laitio University of Turku Finland	Assistant Professor Pekka Marttinen Aalto University Finland
Opponent	Professor Teijo Saari University of Turku Finland	
Custos	Professor Arvi Yli-Hankala Tampere University Finland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-2067-6 (print)
ISBN 978-952-03-2068-3 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN: 978-952-03-2068-3

PunaMusta Oy – Yliopistopaino
Joensuu 2021

To my family, friends and colleagues

ACKNOWLEDGEMENTS

This dissertation project was financially supported by the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital, The Finnish Medical Foundation and The Finnish Society of Anaesthesiologists. These grants are gratefully acknowledged.

The studies presented in this dissertation were carried out at the Faculty of Medicine and Health Technology, Tampere University and in the Emergency Medical Services, Tampere University Hospital.

I am extremely grateful to my supervisors Docent Sanna Hoppu (MD, PhD) and Docent Antti Kämäräinen (MD, PhD). Without your positive feedback, guidance and encouragement, I would have not completed this thesis. Sanna, you have always been available to help and supervise me despite your numerous other projects. Antti, it is no accident that you were rewarded as the supervisor of the year.

I want to thank Professor Arvi Yli-Hankala (MD, PhD), Docent Reetta Huttunen (MD, PhD) and Docent Jaakko Lånsjö (MD, PhD) who have comprised my follow-up group. Your high-quality insights have developed my dissertation greatly. I would like to acknowledge my official reviewers of this dissertation, Assistant Professor Pekka Marttinen (MSc, PhD) and Associate Professor Timo Laitio (MD, PhD). I am very thankful to statistician Heini Huhtala (MSc), for her invaluable guidance and support in statistical analysis. I am very grateful to Antti Kallonen (MSc) who introduced me to the fascinating world of machine learning. I also want to thank Erik Lydén and Jan Kurki for their hard work listening and transcribing the emergency calls.

I would also like to thank all my other co-authors: Jari Kalliomäki (MD), Jussi Pirneskoski (MD), Docent Jouni Nurmi (MD, PhD), Docent Markku Kuisma (MD, PhD) and Professor Klaus Olkkola (MD, PhD). Additionally, I want to thank my colleagues Verneri Hannula, Pekka Innanen, Erno Iso-Aho, Aleksi Lamminen and Miku Tuominen for their work collecting the data at the emergency department. I am very grateful to Joonas Tirkkonen, Maija-Liisa Kalliomäki and Ville Jalkanen for their insightful comments that helped me to improve the original manuscripts. I also wish to acknowledge Janne Virta for his contribution at the very beginning of this journey.

I would like to express my gratitude to all my friends in Tampere, Jyväskylä and Rovaniemi. The Finnish ultimate frisbee community has acted as a healthy counterbalance to this research project. A particular thankyou goes to ultimate frisbee friends Satu Sihvonen and Kaisa Vihervaara for their expertise in translating the trigger words. Additionally, I would like to thank Atte Nikkilä for his tips and peer support as a young researcher.

My deepest gratitude goes to my parents Päivi and Ari and to my little brothers Julius and Johannes for their compassion and support throughout my studies. I wish to thank my relatives, especially Paula and Elma, for supporting me through this process. Finally, the loving thankfulness goes to Emmaliisa.

Seinäjoki, August 2021

Joonas Tamminen

ABSTRACT

Early recognition of prehospital emergencies and the dispatch of emergency medical services (EMS) to the incident when appropriate is based on a structured risk assessment and the communication between the emergency dispatcher and the caller. Out-of-hospital cardiac arrest (OHCA) is an example of a medical emergency which requires early recognition and prompt treatment. Cessation of mechanical cardiac function, circulatory collapse and a subsequent loss of cerebral perfusion will result in an anoxic brain injury and ultimately death if spontaneous circulation is not restored. Early recognition of OHCA is the cornerstone link of the chain of survival as the emergency dispatcher may dispatch first-responding units and ambulances and direct the caller to initiate cardiopulmonary resuscitation (CPR) unless bystander CPR is already being performed. Additionally, the signs of impending cardiac arrest and disturbances in vital functions should be detected in all encountered prehospital patients. The aim of this thesis is to investigate the first links in the chain of survival and the prediction of short-term mortality among prehospital patients.

According to the International Liaison Committee on Resuscitation (ILCOR), the trigger words used by callers in association with cardiac arrest constitute a scientific knowledge gap as they may facilitate OHCA recognition. Study I was a retrospective cohort study which aimed to find laypeople's spontaneous trigger words in emergency calls. Of the 78 dispatcher-suspected OHCAs, 49 were confirmed to be cardiac arrests at the scene. The dispatcher had not suspected a later confirmed OHCA in two cases. A total of 291 trigger words were identified in the emergency calls. Trigger words 'is not breathing' (n = 9 in the confirmed cardiac arrest group vs n = 1 in the non-cardiac arrest group, odds ratio [OR] 6.00, 95% confidence interval [CI] 0.72–50.0), 'the patient is blue' (n = 9 vs n= 1, OR 6.00, 95% CI 0.72–50.0), 'collapsed or fallen down' (n = 12 vs n = 2, OR 4.15, 95% CI 0.86–20.1) and 'is wheezing' (n = 17 vs n = 5, OR 2.40, 95% CI 0.78–7.40) were frequently used to describe a true cardiac arrest. 'Is snoring' was associated with a false suspicion of cardiac arrest (n = 1 vs n = 6, OR 0.08, 95% CI 0.009–0.67).

First-responding units are widely used as a part of the emergency medical services response, especially in the Northern countries. Nevertheless, their impact on the emergency patient's care is unclear. Study II was a retrospective cohort study that

examined the emergency medical services missions that first-responding units attended during one year in the Pirkanmaa Hospital District. The first-responding units encountered 1,622 emergency patients, 1,015 of which were clinically evaluated. CPR was attempted in 83 OHCA missions and a first-responding unit initiated CPR in 42 (51%) patients at a median of 4 minutes prior to ambulance arrival.

Impending cardiac arrest has been traditionally predicted in early warning score systems (EWSs) which are based on logistic regression. However, all introduced EWSs have a limited capability to predict adverse outcomes in the prehospital setting. By contrast, modern machine learning models are able find unknown non-linear associations or interactions between the predictor variables, making them excellent for modelling complex physiological phenomena. In Studies III and IV, the traditional National Early Warning Score (NEWS) system was compared to a random forest machine learning algorithm trained with NEWS parameters and blood glucose for predicting short-term mortality in the prehospital setting. The material for Study III was retrospectively collected in the Helsinki and Uusimaa hospital district between 2008 and 2015, whereas the material for Study IV was prospectively collected in the Pirkanmaa Hospital district in June 2015. The 24-hour mortality in Study III was 1.0%, and a random forest model outperformed NEWS for predicting that outcome (an area under the receiver operating characteristic [AUROC] 0.868 [95% CI 0.843–0.892] vs 0.836 [95% CI 0.810–0.860], $p < 0.001$). Correspondingly, the 30-day mortality in Study IV was 3.4%, and a random forest was superior to NEWS for predicting that outcome (AUROC 0.758 [95% CI 0.705–0.807] vs 0.682 [95% CI 0.619–0.744], $p < 0.001$).

It can be concluded that the first-responding units shortened the delay from cardiovascular collapse to the initiation of CPR in half of the cases. No trigger word was associated with cardiac arrest, but 'is wheezing' was frequently used among the confirmed OHCA patients. Random forest machine learning algorithms showed better performance for predicting short-term mortality than the traditional NEWS system in two distinct hospital districts.

TIIVISTELMÄ

Sairaalan ulkopuolisten hätätilanteiden varhainen tunnistaminen ja asianmukaisen ensihoitoresurssin hälyttäminen tapahtumapaikalle perustuvat hätäkeskuspäivystäjän ja hätäpuhelun soittajan väliseen viestintään ja jäsenneltyyn riskinarvioon. Sairaalan ulkopuolinen sydänpysähdys on esimerkki hätätilanteesta, joka vaatii tilanteen nopeaa tunnistamista sekä välitöntä hoitoa eli elvytyksen aloittamista. Hoitamattomana verenkierron pysähtyminen aiheuttaa potilaalle aivovaurion ja johtaa tämän menehtymiseen. Kun sydänpysähdys on todettu, hätäkeskuspäivystäjä hälyttää ensivasteyksikön ja ambulanssin kohteeseen sekä tarvittaessa neuvoo soittajaa aloittamaan elvytyksen. Myös uhkaava sydänpysähdys ja tätä edeltävät peruselintoimintojen häiriöt tulisi tunnistaa kaikilta ensihoidon kohtaamista potilailta. Tämän väitöskirjan tarkoituksena on tutkia sydänpysähdyspotilaan selviytymisketjun (”the chain of survival”) ensimmäisiä vaiheita, ensivasteyksiköiden roolia osana hätätilapotilaan hoitoa sekä ensihoidon kohtaamien potilaiden lyhyen aikajänteen kuolleisuuden ennustamista.

Maailman elvytysneuvosto (the International Liaison Committee on Resuscitation, ILCOR) on nostanut hätäpuhelun sisältämät sydänpysähdykseen liittyvät avainsanat tärkeäksi tutkimuskohteeksi (”knowledge gap”). Avainsanojen avulla voidaan parantaa sydänpysähdyksen tunnistamista. Ensimmäinen osatyö on takautuva kohorttitutkimus, jonka tarkoituksena on löytää sydänpysähdykseen liittyviä avainsanoja, jotka esiintyvät hätäpuhelussa maallikon spontaanissa puheessa. Hätäkeskuspäivystäjä epäili sydänpysähdystä 78 tehtävässä, joista 49 osoittautui todellisiksi sydänpysähdyksiksi. Hätäkeskuspäivystäjä ei ollut epäillyt kahta kohteessa todettua elottomuutta. Puheluissa havaittiin yhteensä 291 avainsanaa. Avainsanoja ‘ei hengitä’ (n = 9 varmennettu sydänpysähdys vs n = 1 ei sydänpysähdystä, odds ratio [OR] 6.00, 95 prosentin luottamusväli [LV] 0.72–50.0), ‘sininen’ (n = 9 vs n = 1, OR 6.00, 95 %:n LV 0.72–50.0), ‘kaatunut’ (n = 12 vs n = 2, OR 4.15, 95 %:n LV 0.86–20.1) ja ‘korisee’ (n = 17 vs n = 5, OR 2.40, 95 %:n LV 0.78–7.40) käytettiin usein varmennetun sydänpysähdyksen yhteydessä. ‘Kuorsaa’ assosioitui väärään epäilyyn sydänpysähdyksestä (n = 1 vs n = 6, OR 0.08, 95 %:n LV 0.009–0.67).

Ensivasteyksiköitä käytetään laajasti osana ensihoitopalvelujärjestelmää erityisesti Pohjoismaissa. On kuitenkin epäselvää, mikä on ensivasteyksiköiden hätätilapotilaan

hoitoon osallistumisen merkitys. Toinen osatyö on takautuva kohorttitutkimus, jonka aineistona on yhden vuoden ensivastetehtävät Pirkanmaan sairaanhoitopiirin alueella. Tutkimuksessa kuvataan ensivasteyksiköiden tekemät hoitotoimenpiteet ja näiden vaste hätätilapotilaan hoidossa. Ensivasteyksiköt kohtasivat 1 622 potilasta ja arvioivat 1 015 potilasta. Elvytettyjä potilaita oli yhteensä 83, joista 42 (51 %) potilaan kohdalla ensivasteyksikkö oli aloittanut elvytyksen ennen ambulanssin saapumista kohteeseen (mediaani 4 minuuttia).

Uhkaavaa elottomuutta on perinteisesti ennustettu logistiseen regressiomalleihin pohjautuvilla aikaisen varoituksen pisteytysjärjestelmillä, joista suomalaisille tutuin lienee National Early Warning Score (NEWS) -pisteytys. NEWS ei kuitenkaan sovellu erityisen hyvin riskinarvion apuvälineeksi ensihoidon potilasaineistossa. Sen sijaan modernit koneoppivat mallit kykenevät tunnistamaan algoritmille annetusta aineistosta entuudestaan tuntemattomia yhteyksiä ja riippuvuuksia. Tämä ominaisuus tekee niistä erityisen hyviä mallintamaan monimutkaisia fysiologisia ilmiöitä. Kolmannessa ja neljännessä osatyössä vertaamme NEWS-muuttujien ja verensokerin mittauksien avulla rakennettua random forest -algoritmia perinteiseen NEWS-pisteytykseen ensihoitopotilaan lyhyen aikajänteen kuolleisuuden ennustamisessa. Aineistot on kerätty takautuvasti Helsingin ja Uudenmaan sairaanhoitopiirin alueelta vuosien 2008–2015 ajalta sekä prospektiivisesti Pirkanmaan alueelta kesäkuussa 2015. Kolmannen osatyön yhden vuorokauden kuolleisuus oli 1,0 % ja random forest -algoritmi ennusti tätä NEWS-pisteytystä paremmin (receiver operating characteristics -käyrän alle jäävä pinta-ala [AUROC] 0.868 [95 %:n LV 0.843–0.892] vs 0.836 [95 %:n LV 0.810–0.860], $p < 0.001$). Vastaavasti neljännessä osatyössä 30 päivän kuolleisuus oli 3,4 % ja random forest -algoritmi ennusti tätä NEWS-pisteytystä paremmin (AUROC 0.758 [95 %:n LV 0.705–0.807] vs 0.682 [95 %:n LV 0.619–0.744], $p < 0.001$).

Väitöskirjan päätelminä todetaan, että ensivasteyksiköt lyhensivät viivettä sydänpysähdyspotilaan elvytyksen aloittamiseen puolessa tapauksista. Yksikään hätäpuheluiden avainsanoista ei ollut yhteydessä sydänpysähdykseen, mutta ’korisee’ esiintyi usein varmennetun sydänpysähdyksen tapauksessa. Random forest -algoritmi kykeni ennustamaan ensihoidon kohtaamien potilaiden lyhyen aikajänteen kuolleisuutta perinteistä NEWS-pisteytystä paremmin, mikä havaittiin kahden eri sairaanhoitopiirin alueella.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

AED	Automated external defibrillator
AHA	American Heart Association
ALS	Advanced life support
AVPU	Alert, verbal, pain, unresponsive
AUROC	Area under the receiver operating characteristic
BLS	Basic life support
BG	Blood glucose
CI	Confidence interval
GCS	Glasgow Coma Scale
CoSTR	Consensus on Science with Treatment Recommendations
CPC	Cerebral Performance Category
CPR	Cardiopulmonary resuscitation
DA-CPR	Dispatcher-assisted cardiopulmonary resuscitation
e.g.	Exempli gratia (lat.), 'for the sake of example', 'for example'
ED	Emergency department
ELS	EinsatzLeitSystem
EMS	Emergency medical services
EMT	Emergency medical technician
ERC	European Resuscitation Council
etc.	Et cetera (lat.) 'and the rest', 'and so forth'
EWS	Early warning score
FHQ	First hour quintet (cardiac arrest, severe respiratory failure, chest pain, severe trauma and stroke)
FRU	First-responding unit
GCS	Glasgow Coma Scale
HEMS	Helicopter emergency medical services
i.e.	Id est (lat.), 'that is'
ICPC	International Classification of Primary Care
ICU	Intensive care unit
IHCA	In-hospital cardiac arrest

ILCOR	International Liaison Committee on Resuscitation
IQR	Interquartile range
MA	Master of Arts
MEWS	The Modified Early Warning Score
MET	Medical emergency team
NEWS	The National Early Warning Score
NNT	Number needed to treat
OHCA	Out-of-hospital cardiac arrest
OR	Odds ratio
PAD	Public-access defibrillation
RCP	The UK's Royal College of Physicians
ROSC	Return of spontaneous circulation
RF	Random forest
RR	Risk ratio
SD	Standard deviation
STROBE	The Strengthening the Reporting of Observational studies in Epidemiology
TRIPOD	Transparent reporting of a multivariable prediction model for individual diagnosis or prognosis
VF	Ventricular fibrillation

ORIGINAL PUBLICATIONS

This thesis is based on four original publications, which will be referred to in the text as Studies I to IV. The publications are reprinted with the kind permission of the publishers Springer Nature (I), Wiley-Blackwell (II) and Elsevier (III and IV).

I Tamminen, J., Lydén, E., Kurki, J., Huhtala, H., Kämäräinen, A., & Hoppu, S. (2020). Spontaneous trigger words associated with confirmed out-of-hospital cardiac arrest: a descriptive pilot study of emergency calls. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *28*,1.

II Tamminen, J.I., Hoppu, S.E., & Kämäräinen, A.J.J. (2019). Professional firefighter and trained volunteer first-responding units in emergency medical service. *Acta Anaesthesiologica Scandinavica, 63*,111-116.

III Pirneskoski, J., Tamminen, J., Kallonen, A., Nurmi, J., Kuisma, M., Olkkola, K.T., & Hoppu, S. (2020). Random forest machine learning method outperforms prehospital National Early Warning Score for predicting one-day mortality: a retrospective study. *Resuscitation Plus, 4*, 100046.

IV Tamminen, J., Kallonen, A., Hoppu, S., & Kalliomäki, J. (2021). Machine learning model predicts short-term mortality among prehospital patients: a prospective development study from Finland. *Resuscitation Plus, 5*,100089.

AUTHOR'S CONTRIBUTION

I	Conceptualisation, Methodology, Formal analysis, Investigation, Writing and Visualisation
II	Conceptualisation, Formal analysis, Investigation and Writing
III	Conceptualisation, Methodology, Formal analysis and Writing
IV	Conceptualisation, Methodology, Formal analysis and Writing

Conceptualisation	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology; creation of models
Investigation	Conducting a research and investigation process, specifically performing the data collection
Formal analysis	Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesise study data
Writing	Preparation and creation of the published work, specifically writing the initial draft
Visualisation	Preparation and creation of the published work, specifically visualisation and data presentation

1 INTRODUCTION

Early recognition of high-risk emergency patients and the dispatch of ambulances and other resources to the scene when appropriate is based on a structured risk assessment and successful communication between the dispatcher and the caller. Out-of-hospital cardiac arrest (OHCA) is an example of medical emergency which requires prompt treatment (Olasveengen, et al., 2021). Cessation of mechanical cardiac function, circulatory collapse and a subsequent loss of cerebral perfusion will result in an anoxic brain injury and ultimately death if spontaneous circulation is not restored. The return of spontaneous circulation (ROSC) may be achieved by means of cardiopulmonary resuscitation (CPR). Despite progress in basic and clinical research, implementation of various emergency medical services (EMS) systems and education of citizens, OHCA has remained a major public health problem with poor patient outcomes worldwide (Berdowski, Berg, Tijssen, & Koster, 2010; Gräsner, Wnent, et al., 2020; Sasson, Rogers, Dahl, & Kellermann, 2010). The probability of survival after OHCA depends on a sequence of events which encompasses recognition of cardiac arrest and activation of EMS, early CPR and defibrillation, advanced life support (ALS) and post-resuscitation care. These key elements of successful resuscitation are collectively known as the chain of survival (Semeraro et al., 2021).

The first elements in the chain of survival are its strongest links (Deakin, 2018). Early recognition of OHCA is the cornerstone link, as the emergency dispatcher may activate an EMS response and direct the caller to initiate dispatcher-assisted CPR (DA-CPR) unless bystander CPR is already being performed. Dispatcher diagnosis of OHCA in Finland is based on a strict protocol that includes standardised questions regarding the patient's level of consciousness and breathing. The well-known clinical signs and symptoms of cardiac arrest are unresponsiveness and abnormal breathing, but it is unclear how laypeople interpret them, especially agonal breaths. Emergency calls may contain specific trigger words, and the International Liaison Committee on Resuscitation (ILCOR) has suggested that they form a scientific knowledge gap (Olasveengen et al., 2017). Identification of these hypothetical trigger words may further strengthen the first link in the chain of survival.

The first tier of the Finnish EMS response includes first-responding units (FRUs) staffed with professional firefighters, emergency medical technicians (EMTs) and trained volunteers. A FRU's main objective is to reach the OHCA patient first and shorten the delay to initiation of CPR and defibrillation. Although FRUs are widely dispatched to OHCA missions, especially in Northern countries, their performance in other medical emergencies and their contribution to an emergency patient's prehospital care in general is unknown.

The early signs of impending physiological deterioration should be detected and treated accordingly in the prehospital setting. This principle is called the chain of prevention as cardiac arrest can be predicted and prevented in hospital wards (Smith, 2010). The signs of threatening physiological deterioration are seen hours before in-hospital cardiac arrest (IHCA) (Schein, Hazday, Pena, Ruben, & Sprung, 1990), and various early warning score (EWS) systems have been developed and implemented in hospital wards and emergency departments (EDs). These EWS systems use easily accessible clinical data (e.g. physiological measurements) to predict mortality, cardiac arrest, intensive care unit (ICU) admission or sepsis. The use of the National Early Warning Score (NEWS) in the prehospital setting is advocated by the Royal College of Physicians, although its performance and predictive parameters in the prehospital setting could be improved (Royal College of Physicians, 2017). For instance, one possibility could be inclusion of a patient's blood glucose (BG) level in a prehospital EWS (Vihonen, Lääperi, Kuisma, Pirneskoski, & Nurmi, 2020). Additionally, modern machine learning methods may yield more precise estimates of short-term mortality than are obtained with traditional logistic regression models (Churpek et al., 2016). These estimates could facilitate risk stratification and help EMS personnel to recognise high-risk patients who might otherwise be left at the scene or transported to inappropriate destinations.

This study was undertaken to investigate the first links in the chain of survival and risk stratification of prehospital emergency patients. Professional firefighter FRUs and trained volunteer FRUs in the first tier of the EMS response and laypeople's spontaneous trigger words in emergency calls were examined. Study III and IV were development studies in which machine learning models and NEWS were compared for their ability to predict short-term mortality in prehospital patients.

2 REVIEW OF THE LITERATURE

2.1 Aetiology and pathophysiology of cardiac arrest

Cardiac arrest is characterised by an abrupt loss of heart function, together with circulatory collapse and loss of cerebral perfusion (Myat, Song, & Rea, 2018). According to updated Utstein-style reporting, the aetiologies for OHCA are classified as medical (e.g. presumed cardiac origin, asthma, anaphylaxis or no obvious cause), traumatic, drug overdose, drowning, electrocution and asphyxia (Perkins et al., 2015). Of these aetiologies, medical causes are attributed to 91% of all OHCAs in Europe, and a majority (70%) of OHCAs occur in private residences (Gräsner, Wnent, et al., 2020). Coronary artery disease is the leading cause of OHCA worldwide (Myat et al., 2018).

Mechanical cardiac function and circulatory collapse cause a loss of cerebral perfusion. This can result in an anoxic brain injury and will ultimately lead to death if spontaneous circulation is not restored. The pathophysiology of a shockable OHCA is suggested to have three distinct phases: the electrical phase (0 to 4 minutes after collapse), the circulatory phase (4 to 10 minutes after collapse) and the metabolic phase (after 10 minutes after collapse) (Weisfeldt & Becker, 2002). According to this model, defibrillation is the most critical intervention immediately after cardiac arrest, whereas prolonged CPR prior to defibrillation may have a beneficial effect in the second phase. However, a recent ILCOR review found that prolonged CPR does not improve resuscitation outcomes when compared with a short period of CPR before defibrillation (Gräsner, Mancini, et al., 2020). Beyond 10 minutes, the survival rates are poor due to global ischaemia and reperfusion injury.

2.2 Survival after out-of-hospital cardiac arrest

OHCA poses a major public health problem to the general community because the management of this unique medical emergency involves citizens, dispatch centres, an EMS response and hospital care (Ong, Perkins, & Cariou, 2018). The incidence of attempted resuscitation in OHCA victims is between 55 per 100,000 inhabitants per year globally (Berdowski et al., 2010), 56 per 100,000 population per year in Europe (Gräsner, Wnent, et al., 2020) and 47–51 per 100,000 inhabitants per year in Finland (Hiltunen et al., 2012; Setälä, Hoppu, Virkkunen, Yli-Hankala, & Kämäräinen, 2017), Table 1. The survival after OHCA is modest, although most predictors of survival are well-known (Berdowski et al., 2010; Gräsner, Wnent, et al., 2020; Sasson et al., 2010). These predictors are related to patient characteristics (e.g. age, gender, location, bystander response, first monitored rhythm), underlying cause for OHCA, an emergency dispatcher, an EMS system and the OHCA process, which are core elements in Utstein-style reporting (Perkins et al., 2015). According to the European Resuscitation Council (ERC) resuscitation guidelines, successful resuscitation contains a specific sequence of events, which are recognition of cardiac arrest and activation of EMS, early CPR, early defibrillation, early ALS and post-resuscitation care—collectively, these are known as the chain of survival (Olasveengen, et al., 2021).

A recently published study covering 28 countries in Europe reported 25,171 CPR attempts during the three-month study period (Gräsner, Wnent, et al., 2020). Of these patients, 58% received bystander CPR, ROSC was achieved in 33% of the cases and 8% of patients survived to hospital discharge. Survival to hospital discharge among patients with an initial shockable rhythm and witnessed OHCA was 28%. Patients with initial shockable rhythm were more likely to survive to hospital discharge (24% in the shockable group vs 3% in the non-shockable group).

In the Finnresusci study, CPR was attempted in 671 patients in the year 2010. Bystander CPR was provided in 47% of these patients, ROSC was obtained in 44% of the cases, 20% of the patients were discharged alive and the overall survival at one year was 13% (Hiltunen et al., 2012). Survival to hospital discharge among patients with initial shockable rhythm and witnessed OHCA (the Utstein comparator group) was 46%. In another Finnish study, conducted in the Pirkanmaa Hospital District, CPR was attempted in 280 OHCA victims during a 12 month study period (Setälä et al., 2017). Of these patients, a bystander had initiated CPR in 54% of the cases, ROSC was obtained in 36% of the patients and survival to hospital discharge was observed in 14% of the cases. Survival to hospital discharge among patients with witnessed OHCA and an initial shockable rhythm was 33%. A favourable neurological outcome (Cerebral Performance Category [CPC] 1 or 2) at hospital discharge was reported in 29 (10%) of all resuscitated patients. In a third Finnish study, survival to hospital discharge in patients with ventricular fibrillation (VF) as the initial rhythm and witnessed OHCA was 35% (Kuisma et al., 2005).

Table 1. Epidemiology of out-of-hospital cardiac arrest in the world, in Europe and in Finland.

	Berdowski 2010	Gräsner 2020	Kuisma 2005	Hiltunen 2012	Setälä 2017
Design	Systematic review of 67 prospective studies	Prospective observational	Prospective observational	Prospective observational	Prospective observational
Region	Europe, North America, Asia, Australia	28 European countries	Helsinki EMS, Finland	The southern, central and eastern parts of Finland	The Pirkanmaa Hospital District, Finland
Data collection	1984–2008	Oct 1, 2017–Dec 31, 2017	Jan 1, 1997–Dec 31, 2002	Mar 1–Aug 31, 2010	Jun 1, 2013–May 31, 2014
Inclusion criteria	EMS attended adult OHCA	EMS attended OHCA	Bystander witnessed VF, CPR attempted, medical aetiology	Suspected or EMS attended OHCA	EMS attended OHCA
Location residence, %	-	70	80	59	58
Confirmed OHCA	-	37,054	-	1,042	314
Confirmed OHCA per 100,000 inhabitants per year	62	89	-	78	52
CPR attempts	-	25,171	373	671	280
CPR attempts per 100,000 inhabitants per year	55	56	6	51	47
Bystander CPR	-	58%	-	317/671 (47%)	118/280 (54%)
Dispatcher-assisted CPR	-	53%	-	53/164 (32%)	-
ROSC	-	33%	249/373 (68%)	292/671 (44%)	100/280 (36%)
Survival to discharge or 30-day survival	7%	8%	-	133/671 (20%)	39/280 (14%)
Utstein comparator group*	-	28%	132/373 (35%)	64/140 (46%)	19/57 (33%)
Good neurological outcome at discharge**	-	-	-	-	29/280 (10%)

CPR = cardiopulmonary resuscitation, EMS = emergency medical services, OHCA = out-of-hospital cardiac arrest, ROSC = return of spontaneous circulation. VF = ventricular fibrillation. *Bystander witnessed cardiac arrest and initial shockable rhythm, **Cerebral Performance Category 1 or 2.

2.3 Resuscitation guidelines and existing knowledge gaps

The ILCOR's aim is to improve poor survival rates after OHCA by reviewing resuscitation science literature and regularly publishing consensus statements and treatment recommendations (Perkins et al., 2017). These international publications are the basis of the ERC resuscitation guidelines. ILCOR's first international resuscitation guidelines were published in 2000, with subsequent Consensus on Science with Treatment Recommendations (CoSTR) summaries published in 2005, 2010, 2015 and 2020 and continuous evidence evaluation CoSTRs documents in 2017, 2018 and 2019. In addition to the treatment recommendations, ILCOR's CoSTR publications include knowledge gaps which the researchers in the field of resuscitation should address in future studies (Kleinman et al., 2018).

2.4 First links of the chain of survival

2.4.1 Recognition of out-of-hospital cardiac arrest

Recognition of OHCA is the first link and the cornerstone of the chain of survival, as it starts the sequence of events that attempts to restore spontaneous circulation. When the emergency dispatcher has identified OHCA, the dispatcher will activate an adequate EMS response and will instruct the caller to commence CPR unless the caller is already attempting CPR. Nevertheless, the current emphasis in clinical practice has been on ALS procedures and post-resuscitation care, rather than on recognition and communication between the dispatcher and the caller (Ong et al., 2018). ILCOR has noticed this controversy, and the essential role of the emergency dispatcher as the director of an EMS response is underlined in the 2020 CoSTR for basic life support (BLS) and the ERC guidelines 2021 (Olasveengen et al., 2020; Olasveengen et al. 2021). According to the current resuscitation guidelines, the emergency dispatcher should identify OHCA if the caller describes the patient as both unresponsive and breathing abnormally (Olasveengen, et al., 2021).

The diagnostic performance of emergency dispatchers is evaluated in ILCOR's recently published 2020 CoSTR summary for BLS (Olasveengen et al., 2020). The review reported that the median sensitivity for adult OHCA recognition is 0.79 (interquartile range [IQR] 0.69–0.83) in 46 observational studies and the median specificity for adult OHCA recognition is 0.99 (IQR 0.93–1.00) in 12 observational studies. The authors were unable to compare different algorithms or criteria due to large heterogeneity between the studies. Another systematic review found a sensitivity of 0.74, ranging from 0.14 to 0.97, for dispatcher recognition of OHCA (Viereck, Møller, Rothman, Folke, & Lippert, 2017). The large variations observed for the reported sensitivities and specificities may reflect the different definitions of recognised OHCA (Viereck, Møller, Rothman, Folke, & Lippert, 2017).

A recent study undertaken in Finland examined 2,054 emergency calls in which OHCA was witnessed by a bystander (Syväoja, Salo, Uusaro, Jäntti, & Kuisma, 2018). The authors found a sensitivity of 0.80 for the dispatcher to recognise OHCA. Other smaller studies conducted in Finland showed sensitivities ranging from 0.79 to 0.83 for the dispatcher to recognise OHCA (Hiltunen et al., 2015; Kuisma et al., 2005; Nurmi et al., 2006).

Early recognition of OHCA has a marked impact on patient survival to hospital discharge and neurological outcome (Berdowski, Beekhuis, Zwinderman, Tijssen, & Koster, 2009; Nichol et al., 2016). Even a 30 second reduction in call processing may increase the absolute survival rate by 0.7 percentage points (Nichol et al., 2016). In Finland, survival to hospital discharge was higher among recognised OHCA victims (23 % in recognised OHCA group vs 16% in unrecognised OHCA group; odds ratio [OR], 1.56 [95% confidence interval [CI] 1.17–2.08]; number needed to treat [NNT] 15) (Syväoja et al., 2018).

One might state that agonal breathing is the pitfall for OHCA recognition, as misinterpretations of abnormal breathing results in delays in initiations of CPR (Fukushima et al., 2015). The emergency dispatcher should pay more attention to the assessment of patient's breathing, and they should repeat the question regarding breathing if the caller's answer is other than a simple 'yes'. Agonal breathing in OHCA patients is common: approximately 30% of OHCA victims show signs of abnormal breathing (Bång, Herlitz, & Martinell, 2003; Berdowski et al., 2009; Lewis, Stubbs, & Eisenberg, 2013; Riou, Ball, Williams, et al., 2018). In a retrospective study conducted in Japan, the authors concluded that agonal breathing was present in as many as 60% (169/283) of OHCA patients whose breathing pattern was known

(Fukushima et al., 2015). Table 2 summarises the descriptions and frequencies of breathing patterns determined in previous studies.

Table 2. Frequencies and proportions (%) of different breathing patterns in emergency calls among out-of-hospital cardiac arrest patients.

	Bång et al. 2003 n = 100	Fukushima et al. 2015 n = 659	Berdowski et al. 2009 n = 267	Riou et al. 2018 n = 176
Not asked	17 (18)	-	65 (24)	19 (11)
Not available	5 (5)	376 (57)	-	-
Normal breathing	16 (17)	-	-	63 (36)
Not breathing	24 (26)	114 (17)	119 (45)	44 (25)
Abnormal breathing*	50 (53)	107 (16)	67 (25)	50 (28)
Undefined breathing	-	62 (9)	18 (7)	-

*Abnormal breathing included the following descriptions: difficulties, poorly, weak, gasping, wheezing, impaired, occasional and snoring.

Equally to various definitions of recognised OHCA, a miscellaneous number of descriptions of agonal breathing might occur in different EMS systems. Some institutions have tried to construct lists of words which are indicative of agonal breathing. The Medical Priority of Dispatch Systems has suggested that the dispatcher should consider agonal breathing if the following words are present in an emergency call: "barely breathing", "can't breathe at all", "fighting for air", "gasping for air", "just a little", "making funny noises", "not breathing", and "turning blue/purple" (Riou, Ball, Williams, et al., 2018). In a Dutch system, the following words have been regarded as agonal breathing: "occasional breathing," "barely/hardly breathing," "heavy breathing," "laboured or noisy breathing," "sighing," and "strange breathing" (Berdowski et al., 2009). These breathing-related words, together with other caller's descriptions of the emergency patient (e.g. level of consciousness, facial colour or a history of the present illness), can be collectively considered "trigger words" (Berdowski et al., 2009). These hypothetical trigger words could prompt immediate suspicion of OHCA and the dispatching of assets, resulting in shorter EMS response times, initiation of DA-CPR and, ultimately, improved patient outcomes.

ILCOR suggested the following top 3 knowledge gaps for BLS in 2017: (1) 'What is the optimal instruction sequence for coaching callers in dispatch-assisted CPR?', (2) 'What are the identifying key words used by callers that are associated with cardiac arrest?', (3) 'What is the impact of dispatch-assisted CPR instructions on cardiac arrests from noncardiac causes such as drowning, trauma, or asphyxia in adult and paediatric patients?' (Olasveengen et al., 2017). The ILCOR 2020 CoSTR also highlighted several knowledge gaps regarding dispatcher diagnosis of OHCA (Olasveengen et al., 2020). Which algorithm is the most accurate remains unclear, as does the relation between algorithms and time to OHCA recognition or initiation of DA-CPR. The efficacy and the effectiveness of adjunct technologies, such as pulse detection technologies via caller's mobile phone and artificial intelligence to recognise OHCA, should be explored.

2.4.2 Bystander cardiopulmonary resuscitation

The concept of modern CPR—the lifesaving actions that combine both mouth-to-mouth breathing and external cardiac compressions—was introduced in 1960, although various techniques of artificial ventilation had been known hundreds of years earlier and external chest compressions had been successfully used to revive two chloroform-anaesthetised patients in the early 1890s (Chamberlain, 2004; Hurt, 2005; Kouwenhoven, Jude, & Knickerbocker, 1960). The first mass citizen training in CPR was started in Seattle, Washington in the early 1970s, and the first DA-CPR program was launched in King County, Washington in 1982 (American Heart Association, 2018).

Bystander CPR should be started promptly, before an ambulance has arrived at the scene, when OHCA is suspected. Current estimates indicate that the probability of survival will decrease 10% for every minute of delay between the onset of cardiac arrest and the initiation of CPR or defibrillation among patients with an initial shockable rhythm (Valenzuela, Roe, Cretin, Spaite, & Larsen, 1997). The mean EMS response time (time interval from incoming call to the time EMS has reached the scene) varies naturally among EMS systems; for example, it is approximately 10 minutes (SD 4.5) for an attempted CPR in the Pirkanmaa Hospital District in Finland (Setälä et al., 2017). Thus, initiation of CPR before EMS arrival is crucial, as this may prevent the complete cessation of perfusion and oxygenation to the patient's vital organs, such as the heart and brain. Bystander-initiated CPR is worth attempting, as it increases the probability of 30-day survival twofold (adjusted OR, 2.15 [95% CI 1.88–2.45]; NNT 16) and is associated with improved long-term neurological outcome (Hasselqvist-Ax et al., 2015; Kragholm et al., 2017).

According to current ERC resuscitation guidelines, bystander CPR should include chest compressions and, optionally, rescue breaths at a ratio of 30:2 if the bystander is trained and able to provide mouth-to-mouth ventilation (Olasveengen, et al., 2021). The bystander CPR rates have increased gradually in Europe in the 2000s (Hasselqvist-Ax et al., 2015; Wissenberg et al., 2013). The recent EuReCa TWO trial found that a bystander CPR rate of 58% (range of 28 country values 13%–82%) and a DA-CPR rate of 53% (range of 24 country values 3%–88%) for OHCA patients in Europe in 2017 (Gräsner, Wnent, et al., 2020). The bystander CPR rate is similar in Finland, where approximately half of the OHCA patients receive bystander CPR (Hiltunen et al., 2012; Setälä et al., 2017). The Finnresusci study showed that a dispatcher provided instructions to start CPR in 32% of cases with EMS-confirmed OHCA (Hiltunen et al., 2015).

An ILCOR systematic review and meta-analysis examined whether the emergency dispatcher should give CPR instructions via telephone (Nikolaou et al., 2019). The one-month survival and survival with good neurological outcome at hospital discharge were greater among patients with DA-CPR than in patients with no bystander CPR (adjusted OR, 1.63 [95% CI 1.32–2.01] and adjusted OR, 1.54 [95% CI 1.35–1.76], respectively). The authors also compared the DA-CPR group to patients who had received unassisted bystander CPR. Survival at one month was slightly higher in the DA-CPR group (adjusted OR, 1.13 [95% CI 1.06–1.20]) but no significant difference was observed regarding survival with good neurological outcome at hospital discharge (adjusted OR, 1.12 [95% CI 0.94–1.34]). The review concluded that provision of DA-CPR is associated with improved patient survival when compared with no attempted bystander CPR. Similarly, a recent retrospective study that compared audio-instructed DA-CPR to video-instructed DA-CPR in Korea found no difference in survival to hospital discharge (adjusted OR, 1.20 [95% CI 0.74–1.94]) (Lee, Song, Shin, Hong & Kim, 2020).

EMS systems should provide CPR instructions when appropriate, but the optimal instruction sequence for DA-CPR is unclear (Olasveengen et al., 2017). The communication between the caller and the dispatcher during the emergency call was examined in two Australian linguistic studies (Riou, Ball, Whiteside, et al., 2018; Riou et al., 2017). In the first study, the authors noticed that the phrasing of the question "Tell me exactly what's happened?" rather than "Tell me exactly what happened?" resulted in fewer narrative answers (42% vs 57%), and this influenced the length of the caller's answer (9 s vs 18 s), thereby reducing the time to dispatch an ambulance (50 s vs 58 s). The second study assessed the commencement of bystander CPR in cases with EMS-confirmed OHCA. The bystanders were more likely to attempt CPR if the dispatcher referred to CPR in terms of futurity or obligation rather than willingness (97%, 84% and 43%, respectively).

Correspondingly, another important knowledge gap in bystander CPR concerns the provision of mouth-to-mouth ventilation and its timing. Current guidelines suggest that only trained lay rescuers should give rescue breaths to OHCA patients (Olasveengen, et al., 2021). A significant exception to that rule are paediatric patients and cases in which the aetiology of OHCA is drowning or asphyxial. In these circumstances, a combination of chest compressions and rescue breaths is recommended. Among OHCA victims with VF as a cardiac rhythm, a three-phase pathophysiological time-sensitive model has been proposed (Weisfeldt & Becker, 2002). In that model, the first four minutes are called the electrical phase, the following minutes between four and ten minutes are termed the circulatory phase, and this is followed by the metabolic phase. The authors argued that artificial ventilation becomes more relevant in the last metabolic phase, whereas continuous chest compressions are essential in the first and the second phases.

An ILCOR review regarding the compression to ventilation ratio in DA-CPR showed that continuous chest compressions resulted in similar survival rates with good neurological outcome when compared with compression and ventilation at a ratio of 15:2 (unadjusted risk ratio [RR], 1.25 [95% CI 0.94–1.66]), with no substantial benefit for survival (unadjusted RR, 1.20 [95% CI 1.00–1.45]) (Ashoor et al., 2017). An ongoing randomised, controlled trial is addressing a controversial research question regarding whether trained bystanders should also provide rescue breaths at a ratio of 30:2 rather than compressions-only CPR (TANGO2, ClinicalTrials.gov identifier: NCT03981107).

2.4.3 Public-access defibrillation

BLS consists of CPR and defibrillation (Olasveengen, et al., 2021), and the use of an automated external defibrillator (AED) is possible even with no prior experience or training (Yeung, Okamoto, Soar, & Perkins, 2011). VF and pulseless ventricular tachycardia can be defibrillated, whereas pulseless electrical activity and asystole are non-defibrillatable rhythms. The proportion of initial shockable rhythms is approximately 20% of all OHCAs (Gräsner, Wnent, et al., 2020). The first monitored cardiac rhythm is a core patient parameter in the Utstein template, since it has a marked effect on patient outcomes and treatment protocol (Perkins, Jacobs, et al., 2015).

The first public-access defibrillation (PAD) programs were launched in the 1990s (American Heart Association, 2018). Scientific evidence regarding the implementation of AEDs used by non-medical personnel emerged in 2000 (Page et al., 2000; Valenzuela et al., 2000). These studies, which were conducted in special circumstances (i.e. in casinos and airplanes), were followed by the PAD trial (Hallstrom et al., 2004) that showed a doubling of the survival rate at hospital discharge when early defibrillation was performed by trained volunteers rather than volunteers with public CPR training only.

Several PAD programs have subsequently been implemented and large OHCA registries in Japan and in USA have confirmed their effectiveness (Blom et al., 2014; Kitamura et al., 2016; Ringh, Jonsson, et al., 2015; Weisfeldt et al., 2010). The survival with good neurological outcome (CPC 1 or 2) at one month was significantly higher with PAD than without PAD (38.5% vs 18.2%; adjusted OR, 1.99 [95% CI 1.80–2.19]; NNT 5) (Kitamura et al., 2016). The ILCOR 2020 CoSTR for BLS also evaluated the effectiveness of PAD and found a greater effect size for the same endpoint (OR, 6.60 [95% CI 3.54–12.28]) (Olasveengen et al., 2020).

Unfortunately, the on-site use of AEDs has many notable limitations. One limitation is that public-access AEDs are rarely used. Of the 43,762 bystander-witnessed VF OHCAs, AED was applied only in 4,499 (10.3%) cases in Japan between 2005 and 2013 (Kitamura et al., 2016). Similar observations have been reported in Sweden and Denmark, even though AEDs are widely distributed in Scandinavia (Malta Hansen et al., 2014; Ringh, Jonsson, et al., 2015). A second limitation is that public knowledge of and confidence in defibrillation is poor. Only a minority of citizens know where to find the nearest AED and even fewer would actually use it (Brooks et al., 2015). Another limitation is that the majority (70%) of OHCAs in Europe occur at private residences or homes, so PAD may not be applicable in these circumstances (Gräsner, Wnent, et al., 2020). A fourth limitation is that AEDs are accessible only at certain times of day, and this variable availability has a negative influence on bystander defibrillation rate and patient survival (Malta Hansen et al., 2013; Karlsson et al., 2019).

2.5 First responders

2.5.1 Professional first responders in the chain of survival

FRUs, along with ambulance units, are a part of a dual-dispatch system in EMS responses. Their main objective is to reach the emergency patient before an ambulance and to reduce the delay in the initiation of potentially lifesaving procedures. In the case of OHCA, FRUs aim to shorten the time between collapse and a CPR attempt and defibrillation. Firefighters, police officers, rescue squads, life-saving crews and home care providers can be equipped with AEDs, and their integration into EMS responses may be reasonable, especially in rural and sparsely populated residential areas (Berdowski et al., 2011; Hansen et al., 2015; Høyer & Christensen, 2009; Malta Hansen et al., 2015; Nehme, Andrew, Bernard, Haskins, & Smith, 2019; Nordberg et al., 2014; Saner, Morger, Eser, & von Planta, 2013; van Alem, Vrenken, de Vos, Tijssen, & Koster, 2003; Zijlstra et al., 2018). The characteristics and performance of various professional FRUs and their contribution to EMS responses are presented in Tables 3–4.

Robust evidence for the value of professional FRUs as a part of an EMS response in OHCA missions emerged in the early 2000s (van Alem et al., 2003). In that randomised, controlled trial, the time interval from collapse to first shock was shorter in the intervention area covered by police and firefighter FRUs than in the control area (11.1 min vs 12.8 s), and this may have resulted in increased rates of ROSC (57% vs 48%), hospital admission (42% vs 33%) and hospital discharge (18% vs 15%). A similar observation of a reduced EMS response time (1.2 minutes) and a shortened time from the emergency call to the first shock (2.5 minutes) has also been described after implementation of professional FRUs (Berdowski et al., 2011; Nordberg et al., 2014). Subsequent published observational studies have also demonstrated the contribution of FRUs to an EMS response, as improved survival rates after OHCA have been attributed to increases in the number of bystander CPR and PAD, as well as increased rates of FRU-initiated CPR and use of dispatched AED (Malta Hansen et al., 2015; Nehme et al., 2019; Zijlstra et al., 2018).

A large register study conducted in North Carolina showed that the combination of bystander-initiated CPR and subsequent defibrillation by a professional FRU improved the odds for survival with good neurological outcome when compared with EMS-initiated CPR and defibrillation (79/343 [23%] vs 29/198 [15%]; adjusted OR, 1.64 [95% CI 1.02–2.65]; NNT 12) (Malta Hansen et al., 2015). A large retrospective study of trends in non-EMS defibrillation reported an increase in the proportion of FRU-defibrillated patients from 11% to 38% in the state of Victoria, Australia between 2000–2002 and 2015–2017 (Nehme et al., 2019). That study also found that the chance for survival to hospital discharge was greater among patients who were defibrillated by FRU than by EMS personnel (adjusted OR, 1.40 [95% CI 1.18–1.67]).

Table 3. Characteristics of professional first-responding units in dual-dispatch systems for out-of-hospital cardiac arrest.

	van Alem 2003	Berdowski 2011	Saner 2013	Nordberg 2014	Malta Hansen 2015	Zijlstra 2018	Nehme 2019
Design	Randomised controlled	Prospective observational	Prospective observational	Prospective observational	Prospective, observational	Prospective, observational	Retrospective observational
Region	Amsterdam, The Netherlands	North Holland province, the Netherlands	Olten, Switzerland	Stockholm County, Sweden	North Carolina, USA	Copenhagen, Stockholm, Western Sweden, Amsterdam	State of Victoria, Australia
Population, million	1.6	2.4	0.018	2,0	2.7	6.8	6.4
Area, km^2	885	-	190	6,519	-	34,267	227,000
Population density, per km^2	1,800	-	95	310	-	200	30
Data collection	Jan 1, 2000 – Jan 31, 2002	Jan 1, 2006 – Mar 31, 2009	2001–2008	2006–2009	2010–2013	2008–2013	2000–2017
FRU type	Police, firefighter	Firefighter, police	Firefighter	Firefighter	Police, firefighter, rescue squad, life-saving crew	Firefighter, police	Firefighter
Number of responders	1,649	67	500	43 stations	-	-	-

FRU = first-responding unit.

Table 4. Performance of professional first-responding units in dual-dispatch systems for out-of-hospital cardiac arrest.

	van Alem 2003	Berdowski 2011	Saner 2013	Nordberg 2014	Malta Hansen 2015	Zijlstra 2018	Nehme 2019
Alerts	234	-	1,334	-	-	-	-
CPR attempts	469	2,833	238	1,961	4,961	22,453	10,451
Initial shockable rhythm	308/469 (66%)	1,372/2,833 (48%)	85/238 (36%)	25%	1,119/4,961 (23%)	6,530 (29%)	10,451
FRU first on scene	74/243 (30%)	-	1,166/1,334 (87%)	41%	-	-	-
EMS response time, median (IQR), min	-	9.0 (6.0–11.0)	12 (8–15)	7.9 (5.6–11.4)	8.0 (6.1–10.1)	7 (5–10)	8 (6–10)
FRU response time, median (IQR), min	-	-	6 (4–8)	-	-	-	-
EMS and FRU response time, median (IQR), min	-	-	-	6.7 (5.0–8.9)	-	-	-
Time from call to first shock, median (IQR), min	11.1 (8.8–15.7)	10.3 (8.0–13.1)	-	-	-	-	10 (8–13)
FRU-initiated CPR	-	-	164/238 (69%)	23%	2,012/4,956 (41%)	-	-
Defibrillation by FRU	-	478/1,372 (35%)	93/124 (75%)	-	774/1,648 (47%)	1,291/6,530 (20%)	797/10,451 (8%)
ROSC	139/234 (57%)	-	47/124 (38%)	29%	1,438/4,948 (29%)	-	5,586/10,451 (53%)
Survival to discharge or 30-day survival	44/243 (18%)	508/2,833 (18%)	-	8%	496/4,921 (10%)	2,957/22,453 (13%)	2,401/10,451 (23%)
Good neurological outcome at discharge	-	464/2,833 (16%)	18/124 (15%)	-	439/4,921 (9%)	-	-

CPR = cardiopulmonary resuscitation; EMS = emergency medical services; FRU = first-responding unit; IQR = interquartile range; ROSC = return of spontaneous circulation.

2.5.2 Trained volunteer responders in the chain of survival

In addition to professional FRUs, volunteer or citizen responders who are trained in BLS can be alerted to reach the OHCA patient. The aim of implementation of these responders is to increase the rates of bystander CPR and on-site defibrillation. In the era of smartphones, these responders can be alerted via text messages or mobile applications with mobile positioning systems, as highlighted in Table 5 (Andelius et al., 2020; Berglund et al., 2018; Caputo et al., 2017; Lee et al., 2019; Pijls, Nelemans, Rahel, & Gorgels, 2016; Ringh, Rosenqvist, et al., 2015; Stroop, Kerner, Strickmann, & Hensel, 2020; Zijlstra et al., 2014). This modern technology enables the dispatcher to orchestrate a meaningful response: if volunteer responders are available and AEDs are accessible in the vicinity of the scene, some volunteers can be directed straight to the patient while others can be guided to retrieve the nearest AED.

In contrast to the implementation of firefighters and police officers, the implementation of trained volunteer responders seems to be particularly reasonable in urban areas, Table 6. A recent Dutch study found that the optimal density of AEDs and volunteer responders in their region was as high as 2 AEDs per km^2 and >10 responders per km^2 (Stieglis et al., 2020). Currently, 4,472 AEDs are distributed in an area of 338,440 km^2 in Finland, and the criteria for 2 AEDs per km^2 is fulfilled in most city centres of the urban and suburban municipalities in Finland (Finnish Heart Association, 2021).

A randomised, controlled trial conducted in Stockholm County showed that the implementation of volunteer responders is associated with increased rates of bystander-initiated CPR (62% in the intervention group vs 48% in the control group) (Ringh, Rosenqvist, et al., 2015). In that study, approximately 70% of the OHCAs occurred at home in both groups. Moreover, observational studies suggest that the integration of volunteer responders into emergency response results in increased rates of survival to hospital discharge (Lee et al., 2019; Stroop et al., 2020). A before/after study of a volunteer responder intervention in Korea found that the proportion of patients with good neurological outcome increased from 4.5% to 8.3% between 2013–2015 and 2015–2017 (adjusted OR, 2.31 [95% CI 1.44–3.70]; NNT 27) (Lee et al., 2019). Nevertheless, a recent Cochrane review concluded that the beneficial effect of community responders on patient survival remains uncertain (Barry et al., 2019). On the contrary, a recently published systematic review and meta-analysis found that the community interventions for OHCA (e.g. implementation of professional and volunteer FRUs, public BLS courses, mandatory CPR training when acquiring a driver's license and mass media campaigns) improved both survival to hospital discharge and 30-day survival (1,158/11,812 [10%] vs 720/9,401 [8%]; OR, 1.34 [95% CI 1.14–1.57]; NNT 47) (Yu et al., 2020).

The results of these studies are at risk of becoming outdated as mobile technology develops. Fortunately, two randomised controlled trials of trained volunteers alerted via mobile application are currently in progress in the Capital region of Denmark and in two regions in Sweden (The HeartRunner trial, ClinicalTrials.gov identifier: NCT03835403; The Scandinavian AED and Mobile Bystander Activation Trial, ClinicalTrials.gov Identifier: NCT02992873). Interestingly, an observational pilot study for the trial conducted in Denmark showed that the use of mobile applications could triple PAD rates from 7% to 21% (Andelius et al., 2020).

2.5.3 Trained volunteer responders in other prehospital emergencies

Although successful implementations of first responder interventions to OHCA have been described in some urban regions in Europe, there is a lack of studies evaluating volunteer responders' role in an EMS response to other high-risk emergency patients. In addition to OHCA, trained volunteer responders can be dispatched to other prehospital emergencies, especially in sparsely populated rural areas in the Northern countries and Scotland (Roberts, Nimegeer, Farmer, & Heaney, 2014; Rørtveit & Meland, 2010). These prehospital emergencies include but are not limited to missions regarding chest pain, suspected acute myocardial infarction or unconsciousness. The clinical impact of these volunteer interventions to the emergency patient's care is uncertain.

Table 5. Characteristics of volunteer responders alerted via text messages or apps to out-of-hospital cardiac arrest.

	Zijlstra 2014	Ringh 2015	Pijls 2016	Caputo 2017	Berglund 2018	Lee 2019	Stroop 2019	Andelius 2020
Design	Prospective observational	Randomised, controlled	Prospective observational	Prospective observational	Prospective observational	Prospective observational	Retrospective observational	Prospective observational
Region	Two regions in the Netherlands	Stockholm County, Sweden	The province of Limburg, The Netherlands	The Canton of Ticino, Switzerland	Stockholm County, Sweden	Seoul, Korea	The district of Gütersloh, Germany	Copenhagen, Denmark
Population, million	1.3	2.0	1.1	0.35	2.3	10	0.37	1.8
Area, km^2	2,909	6,519	2,153	2,800	6,519	605	-	2,559
Population density, per km^2	450	310	510	130	350	17,000	-	700
Data collection	Feb 1 – Jul 31, 2010	Apr 1, 2012 – Dec 1 2013	Apr 1, 2012 – Apr 30, 2014	Jan 1, 2012 – Dec 31, 2015	Feb 11 – Aug 31, 2016	2013–2017	Oct 1, 2013 – 31 Dec, 2017	Sep 1, 2017 – Aug 31, 2018
Responders recruited	14,112	9,828	61,000	1,825	23,097	63,924	740	23,117
Number of AEDs	1,550	-	-	-	2,592	2,045	-	5,000
Alert system	TM	Application	TM	TM, application	Application	TM	Application	Application
Radius around the patient, m	1,000	500	1,000	-	240 to 1200	-	-	1,800
Radius around an AED, m	500	-	1,000	-	-	-	-	1,800

AED = automated external defibrillator; TM = text message.

Table 6. Performance of volunteer responders alerted via text messages or apps to out-of-hospital cardiac arrest.

	Zijlstra 2014	Ringh 2015	Pijls 2016	Caputo 2017	Berglund 2018	Lee 2019	Stroop 2019	Andelius 2020
EMS-treated OHCA	1,536	667	833	593	198	3,194	730	438
Volunteer alerted	893	306	422	332	685	598	-	819
≥1 Volunteer responded alert	-	199/306 (65%)	-	-	-	-	-	-
Volunteer reached scene	-	180/306 (59%)	291/422 (69%)	-	116/198 (58%)	-	342/730 (47%)	-
Volunteer first on scene	-	70/306 (23%)	-	-	51/198 (26%)	-	"One third"	184/438 (42%)
AED connected by volunteer	184/1,536 (12%)	-	27%	-	17/198 (9%)	-	-	91/438 (21%)
Comparison	-	Alert vs no alert	Volunteer attended vs not attended	TM vs application alerted volunteer	-	Before-after	Volunteer vs EMS started CPR	Volunteer vs EMS first on scene
EMS response time, median (IQR), min	-	8.3 (5.4–12.8) vs 8.2 (5.5–11.9)	-	-	-	5 (4–7) vs 6 (5–7)	7 (6–10)	7.1 (5.5–9.8) vs 5.1 (4.0–6.6)
Volunteer response time, median (IQR), min	-	-	-	5.6 (4.2–8.5) vs 3.5 (2.8–5.2)	-	-	4 (3–6)	-
Bystander CPR*	-	62% vs 48%	50% vs 24%	-	-	55% vs 60%	-	85% vs 77%
On-site defibrillation	-	-	-	-	-	-	-	21% vs 7%
ROSC	-	29% vs 29%	41% vs 31%	-	-	7% vs 13%	46% vs 43%	31% vs 30%
Survival to discharge or 30-day survival	-	11% vs 9%	27% vs 16%	17% vs 28%	-	9% vs 13%	18% vs 9%	16% vs 13%
Good neurological outcome at discharge	-	-	-	-	-	5% vs 8%	11% vs 5%	-

AED = automated external defibrillator; EMS = emergency medical services; OHCA = out-of-hospital cardiac arrest; IQR = interquartile range; ROSC = return of spontaneous circulation; TM = text message. *CPR provided by volunteer responders is considered as bystander CPR.

2.6 The chain of prevention

Early detection of the patient's physiological deterioration is vital, since this deterioration may lead to cardiac arrest and cause patient morbidity and mortality. In the in-hospital setting, the signs of impending cardiac arrest are present hours before circulatory collapse in hospitalised patients (Hillman et al., 2002; Schein et al., 1990). These signs can be used to predict IHCA and to estimate the risk of short-term mortality (i.e. mortality within first 24 hours to 30 days). Detection of impending IHCA enables a clinician to intervene in the physiological deterioration and possibly prevent an adverse outcome (Tirkkonen et al., 2020). As with the chain of survival for OHCA, a chain of prevention has been introduced and advocated for IHCA (Smith, 2010). The chain of prevention consists of five links: education, monitoring, recognition, call for help and response.

2.6.1 Early warning scores in the prehospital setting

Various EWS systems, such as the National Early Warning Score (NEWS) or the Modified Early Warning Score (MEWS), have been introduced to facilitate clinical decision-making in hospital wards and EDs and in the prehospital setting (Kivipuro et al. 2018; Nannan Panday, Minderhoud, Alam, & Nanayakkara, 2017; Smith, Prytherch, Meredith, Schmidt, & Featherstone, 2013; Williams, Tohira, Finn, Perkins, & Ho, 2016). The objective of these EWSs is to detect physiological deterioration of a patient prior to adverse outcomes (e.g. 24-hour, 48-hour and 30-day mortality, ICU admission or sepsis). These track-and-trigger systems report an aggregate weighted score based on physiological measurements of the patient's vital functions: a higher score indicates an increased risk for an adverse outcome. Some EWSs are tailored to specific patient populations, whereas others include results of laboratory studies in addition to physiological measurements (Akre et al., 2010; Olsson & Lind, 2003; Singh, McGlennan, England, & Simons, 2012).

The UK's Royal College of Physicians (RCP) introduced NEWS in 2012 and its updated version regarding chronic respiratory illness in 2017 (Royal College of Physicians, 2012, 2017). NEWS is currently one of the most widely used EWS system in the prehospital and the in-hospital settings (Nannan Panday et al., 2017). NEWS has been shown to outperform simple dichotomised medical emergency team (MET) activation criteria in hospital wards (Tirkkonen, Olkkola, Huhtala, Tenhunen & Hoppu, 2014). In a large in-hospital study comparing 34 EWSs, NEWS showed better performance than the other EWSs in relation to a composite outcome of IHCA, ICU admission and 24-hour mortality (an area under the receiver operating characteristic [AUROC] of 0.873 [95% CI 0.866–0.879]) (Smith et al., 2013). In retrospective cohorts, prehospital NEWS had good performance in predicting short-term mortality, Table 7 (Endo et al., 2020; Pirneskoski, Kuisma, Olkkola, & Nurmi, 2019; Silcock, Corfield, Gowens, & Rooney, 2015). Yet, according to systematic reviews, the predictive performance of NEWS or any other EWS in the prehospital setting remains modest, as only extreme aggregate scores (i.e. NEWS = 0 or 7) can predict that the patient is unlikely to deteriorate or that an adverse outcome is likely to occur (Patel et al., 2018; Williams et al., 2016). Furthermore, a small single centre study found a stronger association between short term adverse outcomes for NEWS at an ED than for prehospital NEWS (Abbott et al., 2018).

Table 7. Performance of the National Early Warning Score and its modifications in the prehospital setting.

	Silcock et al. 2015	Pirneskoski et al. 2019	Vihonen et al. 2020	Endo et al. 2020
Design	Retrospective	Retrospective	Retrospective	Retrospective
Region	Paisley, Scotland	Helsinki and Uusimaa, Finland	Helsinki and Uusimaa, Finland	Kawasaki city, Japan
Data collection	Oct 1–Nov 30, 2012	Aug 17, 2008–Dec 18, 2015	Aug 17, 2008–Dec 18, 2015	Apr 1, 2017–Mar 31, 2018
EWS	Standard NEWS	Standard NEWS	NEWS and BG	Standard NEWS
Initial cohort	11,052	750,964	750,964	5,640
Exclusion criteria	Age <16, pregnancy, secondary transportation	Age <18	Age <18	Age <16, OHCA, secondary transmission
Final cohort	1,684	35,800	27,141	2,847
24-h mortality, %	-	1.1	0.74	0.8
AUROC (95% CI)	0.855 (0.69–1)	0.840 (0.823–0.858)	0.851 (0.827–0.875)	0.90 (0.87–0.93)
48-h mortality, %	-	-	-	-
AUROC (95% CI)	0.871 (0.75–0.98)	-	-	-
30-day mortality, %	-	4.5	3.5	-
AUROC (95% CI)	0.740 (0.661–0.819)	0.758 (0.747–0.770)	0.756 (0.741–0.772)	-

AUROC = the area under the receiver operating characteristic; BG = blood glucose; CI = confidence interval; NEWS = the National Early Warning Score; OHCA = out-of-hospital cardiac arrest.

The performance of prehospital EWSs needs further strengthening, especially for discrimination of moderate risk patients. This could be achieved by including additional parameters to a given EWS or by statistical modelling (Linnen et al., 2019).

In-hospital studies suggest that measurements of lactate, D-dimer or some inflammatory biomarkers could be added to NEWS (Eckart et al., 2019; Jo et al., 2016; Nickel et al., 2016). In the prehospital setting, fluctuations in blood glucose (BG) level might reflect stress hyperglycaemia in non-diabetic patients (Dungan, Braithwaite, & Preiser, 2009). A recent retrospective study found that adding BG to prehospital NEWS may slightly enhance its performance as shown in Table 7 (Vihonen et al., 2020).

2.6.2 Machine learning in risk stratification

Machine learning models have been developed for various medical purposes (Rajkomar, Dean, & Kohane, 2019); for example, their purposes may range from the detection of cancer in tissue samples to the identification of multidrug-resistance pathogens in hospital wards. In the context of emergency medicine, speech recognition has been proposed to enhance dispatching process, and a machine learning model has been tested for risk stratification in EDs (Blomberg et al., 2019; Raita et al., 2019). Some studies investigating machine learning models have used inpatient vital signs and laboratory tests to predict physiological deterioration, sepsis or IHCA (Churpek et al., 2016; Giannini et al., 2019). Other recent machine learning studies have also examined predictors for survival among OHCA victims and dispatch rules for sending AED-carrying drones to suspected OHCAs (Al-Dury et al., 2020; Chu et al., 2021).

Various machine learning methods are available, ranging from simple decision trees to complex neural networks. One example of machine learning method is the random forest (RF), which is a collection of computer-generated decision trees (Ho, 1995). A single decision tree is not able to solve complicated problems, but a collection of these weak learners has been demonstrated to work well in prediction tasks involving human physiology (Lin, Hu, & Kong, 2019). In contrast to traditional EWSs, advanced machine learning models may include a vast number of parameters as predictor variables. For instance, in a large retrospective study of 560,486 patients that were presented to an ED, a total of 972 variables were used to develop a machine learning model (Hong, Haimovich, & Taylor, 2018).

Machine learning models seem to outperform traditional EWSs in predicting adverse outcomes in hospital wards, EDs or in the prehospital setting (Churpek et al., 2016; Giannini et al., 2019; Green et al., 2018; Hong et al., 2018; Raita et al., 2019;

Spangler, Hermansson, Smekal, & Blomberg, 2019). In a large multicentre, observational trial of hospitalised ward patients, different machine learning models (e.g. tree-based models, K-nearest neighbours, support vector machines and neural networks) were compared to linear and non-linear logistic regression models and MEWS (Churpek et al., 2016). The authors reported that a RF model was the most powerful machine learning method for predicting 24-hour mortality. The RF model also outperformed logistic models and MEWS (AUROC for RF, 0.936, 95% CI not reported). These findings are not directly applicable to the prehospital setting, as their RF model incorporated both physiological measurements and basic laboratory studies. Nevertheless, BG was an important predictor variable and its contribution in their RF model was similar to that of oxygen saturation and white blood cell count in terms of the Gini index. The Gini index is a metric used to quantify a model's ability to classify outcome predictors into separate classes (Raileanu & Stoffel, 2004).

3 AIMS OF THE STUDY

The aim of this thesis is to investigate the recognition of OHCA, the performance of FRUs in an EMS response and the risk stratification of emergency patients in the prehospital setting. In more detail, the specific aims of studies I–IV were:

1. To examine the association between confirmed OHCA and laypeople's spontaneous trigger words regarding physiological deterioration of a patient in the context of dispatcher-suspected or EMS-encountered cardiac arrest (I).
2. To evaluate the types of EMS missions FRUs complete and describe the general performance of professional and trained volunteer FRUs as a part of an EMS response to high-risk emergency patients (II).
3. To develop a machine learning method for predicting short-term mortality among prehospital patients (III, IV).

4 MATERIALS AND METHODS

4.1 Study design

The thesis consists of three retrospective cohort studies (I, II and III) and one prospective cohort study (IV). Studies I, II and IV were conducted in the Tampere University Hospital district and Study III was undertaken in the Helsinki and Uusimaa Hospital district. The study characteristics are summarised in Table 8.

Table 8. Summary of the studies.

	Study I	Study II	Study III	Study IV
Design	Retrospective	Retrospective	Retrospective	Prospective
EMS setting	Pirkanmaa	Pirkanmaa	Helsinki and Uusimaa	Pirkanmaa
Data collection	Jan 1 – May 31, 2017	Jan 1 – Dec 31, 2013	Aug 17, 2008 – Dec 18, 2015	June 1 – June 31, 2015
Research question	What trigger words are associated with OHCA?	How FRUs contribute to an EMS response?	Does prehospital RF outperform NEWS in predicting 24-h mortality?	Does prehospital RF outperform NEWS in predicting 30-d mortality?
Initial cohort	112 emergency calls	1,894 emergency dispatches	750,964 patients	6,202 emergency dispatches
Exclusion criteria	Trauma, unwitnessed OHCA, IHCA,	None	Age <18	Age <18, terminal care, cardiac arrest, secondary transport
Cohort in the primary analysis	80 emergency calls	1,015 patients	26,458 patients	2,853 patients

EMS = emergency medical services; FRU = first-responding unit; IHCA = in-hospital cardiac arrest; NEWS = the National Early Warning Score; OHCA = out-of-hospital cardiac arrest; RF = random forest.

4.2 EMS systems

At the time of the studies I, II and IV, the EMS system in the Tampere University Hospital district served the city of Tampere, with 220,500 inhabitants, and a surrounding rural area covering a population of 510,000 distributed across 12,600 km^2 (population density 40 inhabitants per km^2) (Statistics Finland, 2020). The EMS system area had one tertiary university hospital, one regional hospital and 18 municipal primary health care centres. During the study periods, the EMS response consisted of three tiers: (1) FRUs staffed with trained volunteers, professional firefighters and EMTs, with BLS-level ambulances, (2) ALS-level ambulances and one extended ALS-level field commander unit and (3) one physician-staffed helicopter emergency medical services (HEMS) unit.

Similarly, at the time of Study III, the Helsinki and Uusimaa Hospital district had a population of 1.6 million people across 9,600 km^2 (population density 170 inhabitants per km^2). That EMS system consisted of three tiers: (1) FRUs and BLS ambulances, (2) ALS ambulances and (3) one physician-staffed ambulance and one physician-staffed HEMS unit.

4.2.1 Emergency call handling and dispatch process

The emergency call handling system in Finland is unique. Since 2001, all emergency calls have been answered in six governmental emergency response centres. The main difference compared with most other European countries is that the same government official (i.e. emergency dispatcher) handles both call-taking and the dispatch of EMS or other resources to the incident when appropriate. In other words, emergency calls are not redirected to separate emergency response organisations; instead, any calls that require immediate action are forwarded to rescue services, health authorities, the police or social services. The length of the formal dispatcher education is 1.5 years, and the majority of emergency dispatchers do not have a medical background (Sankala, 2019).

In Finland, the emergency dispatcher should rule out the possibility of OHCA in every emergency call; thus, the dispatcher follows a strict protocol. The national call processing is protocol-based and computer-aided. During the study periods, the recognition and the dispatcher diagnosis of OHCA was based on three standardised questions: (1) Tell me exactly what happened, (2) Is she/he conscious? and (3) Is she/he breathing normally? (Nurmi et al., 2006) The emergency dispatcher did not receive any additional feedback that differed from the standard quality control.

4.2.2 Professional and trained volunteer first-responding units

The FRUs that were examined in Study II were coordinated and trained by the Pirkanmaa Fire Services. Fourteen of the professional FRUs operated from regional rescue stations and were staffed with firefighters, some of which work also as EMTs at the BLS level. These units responded to FRU dispatches within 90 seconds of the alarm. In addition to professional FRUs, approximately 400 civilians participated as first responders in the EMS system. Twenty-seven of these trained volunteer FRUs responded from home or work. By contract, these units responded to an emergency within 5 minutes of the associated dispatch. Three layperson-staffed units were available for immediate response during daytime, and during the night, these units responded within 5 minutes of a dispatch.

During the study protocol, an FRU was dispatched to an emergency by the Central Dispatch Centre when it was estimated to reach the patient 5 minutes prior to an ambulance in A-level emergencies (the most urgent, including sudden severe unconsciousness or presumed cardiac arrest) or 15 minutes prior to an ambulance in B-level emergencies (urgent mission, potential need for life support measures). In cases of witnessed cardiac arrest, high-energy fall trauma or presumed ischaemic stroke, the FRU was always dispatched, regardless of the expected time advantage over ambulance units. In cases of road traffic accidents and fires, the units were dispatched per rescue service protocol and did not perform as FRUs for the EMS response.

The professional and trained volunteer FRUs had the same treatment modalities and treatment protocol regardless of mission type. The principal means available for an FRU to help a victim of a prehospital emergency included the provision of CPR and AED-based early defibrillation, opening the airway using a supraglottic device or an oropharyngeal airway, supporting breathing with bag-mask ventilation and/or oxygen administration, wound dressing and the control of external haemorrhage, and the administration of rectal diazepam, subcutaneous glucagon, oral nitroglycerine or acetylsalicylic acid, depending on the symptoms. Several programs are available for the initial FRU training of volunteer laypersons, mostly comprised of a BLS course and an additional 30–40 hour FRU course. The basic training of a professional firefighter is 1.5 years in duration, approximately one-third of which consists of emergency care. This training is provided at two colleges in Finland.

4.3 Data collection and exclusion criteria

4.3.1 Study I

In this retrospective study, all consecutive emergency call audio recording of dispatcher-suspected OHCA or EMS-encountered OHCA in the Pirkanmaa Hospital District between January 1st, 2017 and May 31st, 2017 were collected from the EinsatzLeitSystem (ELS) database maintained by the Emergency Response Centre Agency. Cases with unwitnessed OHCA, traumatic cause for OHCA or an institutional resuscitation attempt were excluded.

Spontaneous speech was defined as something that the caller said without being prompted or asked by the dispatcher, and the caller's whole answer to a preceding question was considered as non-spontaneous speech regardless of the duration or the length of the answer. The speech was transcribed by authors EL and JK, who are professional paramedics.

4.3.2 Study II

In this retrospective study, all consecutive FRU missions in Tampere University Hospital district between January 1st, 2013 and 31st, December 2013 were collected. Missions with unreported patient evaluation were excluded. Patient and mission characteristics, including treatment modalities and treatment responses, were extracted from paper mission report forms. Reported treatment modalities were classified as resuscitation, airway management, oxygen administration, medication, spinal immobilisation or splinting and recovery position or postural treatment.

4.3.3 Study III

In this retrospective study, all consecutive EMS missions in the Helsinki and Uusimaa Hospital District between August 17th, 2008 and December 18th, 2015 were collected. Adult patients (age ≥18 years) with a known civil registration number were eligible for the study. Patients with one or more missing NEWS parameters (respiration rate, oxygen saturation, use of supplemental oxygen, temperature, systolic blood pressure, heart rate or level of consciousness) or BG or an erroneous measurement were excluded. Cut-off values for an erroneous measurement were <4 min^{-1} or >70 min^{-1} for respiration rate, <40% or >100% for oxygen saturation, <40 mmHg or >280 mmHg for systolic blood pressure and <20 min^{-1} for heart rate and <25°C or >45°C for temperature.

In the study region, oxygen saturation, use of supplemental oxygen, systolic blood pressure and heart rate were automatically recorded to an electronic patient record system (Merlot Medi), whereas respiration rate, temperature and level of consciousness had to be entered manually to the patient record system by the EMS staff.

4.3.4 Study IV

Study IV shares the prospectively collected study material with a manuscript in preparation. A utilisation, safety and performance of EMS study, which was a part of the EMS quality development program, was conducted in the Tampere Hospital District between June 1st, 2015 and June 31st, 2015. The purpose of that study was to explore characteristics and outcomes of patients in that region who encountered EMS. Therefore, the EMS personnel were mandated to complete all NEWS parameters (i.e. respiration rate, oxygen saturation [SpO2], administration of supplemental oxygen, systolic blood pressure, heart rate, level of consciousness and temperature) in all adult patients, regardless of the mission type at the scene before any intervention. During the data collection, BG was measured if it was clinically appropriate. The indications for measuring BG were (1) known type 1 or type 2 diabetes, (2) altered level of consciousness or (3) suspected acute myocardial infarction or stroke.

During the study period, the completeness of NEWS parameters in the case report forms was verified by medical students. Altogether, six medical students worked different shifts around the clock at the ED in the university hospital. The medical students rechecked the medical reports and ensured that NEWS parameters were copied from the medical records into paper case report forms. If a medical student noted a missing NEWS parameter at the ED (level of consciousness at the scene or use of supplemental oxygen during the mission), that parameter was determined by interviewing the paramedics. A second audit was made by the author JK while he transferred the paper study material to digital format.

Encountered adult patients (age ≥18 years) with a known civil registration number were eligible for the study. Exclusion criteria were cases with a missing case report form, EMS-encountered cardiac arrest or EMS-confirmed death at the scene; patients in terminal care; transportation to another hospital district; or an EMS unit from other districts. Only the first contact with the EMS personnel was included in the analysis if the same patient had multiple contacts.

4.4 Outcome measures

4.4.1 Study I

The trigger words were stratified into EMS-confirmed true cardiac arrest and EMS-confirmed non-cardiac arrest event groups according to mission report forms. After each mission, the EMS personnel filled out specific documentation that contained dispatch and transportation codes. Transportation codes for EMS-confirmed dead, or the patient had ROSC, or CPR was being performed during transportation were interpreted as true cardiac arrests.

4.4.2 Study II

The primary outcome was an improved or normalised vital function or relief of pain. A vital function was considered abnormal if systolic blood pressure <100 mm Hg, heart rate >150 or <40 beats per minute, respiration rate >30 or <10 breaths per minute, oxygen saturation ≤90%, Glasgow Coma Scale (GCS) ≤13 or an impaired level of consciousness on the AVPU scale (A = alert, V = verbal, P = pain, U = unresponsive), or hypoglycaemia (<4 mmol/L). In a subgroup analysis, clinical response was compared between professional firefighter FRUs and trained volunteer FRUs in the First Hour Quintet missions (FHQ; cardiac arrest, severe respiratory failure, chest pain, severe trauma and stroke) (Fischer et al., 2011).

4.4.3 Study III and IV

In Study III, the primary outcome was 24-hour mortality and secondary outcomes were 30-day mortality, 48-hour mortality, ICU admission and a combination of 48-hour mortality and ICU admission. The primary outcome was defined as a death within the next day after a contact with the EMS personnel.

In Study IV, the primary outcome was 30-day mortality, and secondary outcomes were 24-hour mortality, 48-hour mortality, ICU admission and a combination of 48-hour mortality and ICU admission. The patient mortality data for both Studies III and IV was obtained from the Digital and Population Data Services Agency.

4.5 Sample size

No formal sample size calculations were performed for Studies I–IV, as Study I was a hypothesis-generating pilot study, Study II was a retrospective chart review and Study III and IV were post hoc analyses. Study IV shares the same raw data with the utilisation, safety and performance of the EMS study. A sample size of approximately 1,000 patients was estimated as adequate for that study. A detailed statistical review was planned a priori and executed by the statistician Heini Huhtala.

4.6 Missing data

No imputation method was applied to handle missing data in the primary analyses of Studies I–IV (i.e. complete-case deletion was used). In a secondary analysis in Study IV, all eligible patients, regardless of the number of missing vital signs, were analysed but no imputation method was applied to calculate the NEWS score. Missing data were hypothesised to have no effect on the RF model's predictive performance but would instead improve its performance. That analysis included the last contact with the EMS if the patients had several contacts during the one-month study period.

4.7 Statistical analyses

The statistical analyses were performed using SPSS software versions 23 to 25 (SPSS Inc., Chicago, IL, USA), R (version 4.0.0) and Python (version 3.6.9), and the main open-source statistical packages used were NumPy (version 1.17.3) and sklearn (version 0.21.3). The statistician Heini Huhtala was responsible for creating a logistic regression model in Study I. She was also consulted for the statistical analyses that are reported in Studies II–IV. Author AK conducted the formal statistical analyses for Studies III and IV. He has a Master of Science (MSc) degree in information technology. Categorical variables were reported as frequencies and proportions, and numerical normally distributed variables and numerical non-normally distributed were reported as means and standard deviations (SDs) or medians and IQRs, respectively. The normality of the numerical variables was assessed using a Shapiro–Wilk test and Q–Q plots. The comparison between the groups was performed using a $\chi 2$ or a Fisher's exact test for the categorical data when appropriate and a Mann–Whitney U-test for the continuous, nonparametric data. A two-sided p-value < 0.05 was considered statistically significant in all studies.

In order to analyse the transcribed speech in Study I, different words with the same semantic meaning were manually organised into themes (Mäkelä, 1990; Krippendorf, 2004). The spontaneous trigger words were grouped into seven main themes and thirty-six subcategories, the former of which included altered level of unconsciousness, death, breathing, circulation, disability, history of present illness, and unclassified. Breathing included the following subcategories: 'is breathing', 'not breathing', 'laboured breathing', 'breathing heavily', 'breathing irregularly', 'is gasping for breath', 'a deep breath', 'is snoring', 'is wheezing' and unclassified breathing. Each emergency call could fulfil the criteria of each subcategory once. The trigger words were translated from Finnish to English (United Kingdom) by two native Finland linguists who have Master of Arts (MA) degrees in communication sciences. A univariate logistic regression was used to assess the association between the spontaneous trigger words and confirmed cardiac arrests, and the results were presented as OR with 95% CIs.

4.7.1 Development of machine learning models

RF is a supervised ensemble learner and a collection of computer-generated decision trees. It is one of the commonly used supervised classifiers that are included in the sklearn package. RF was chosen as a machine learning method for Studies III and IV as it has been demonstrated to outperform traditional regression models on predicting short-term mortality in hospital wards (Churpek et al., 2016). RF produces a prediction as a probability, and NEWS scores may also be interpreted as a probability when scaled with the maximum score value. The traditional NEWS was treated as it would also be a supervised classifier, which enabled a head-to-head comparison of the models' predictive performances.

Machine learning model development involves two phases: training and testing, Figure 1. The training phase was performed with all predictor variables since additional input features are not detrimental to the performance of RF as there are no substantial correlations. The predictor variables of the RF models are presented in Table 9. In Study III, the level of consciousness was converted from the GCS to the AVPU scale, as in previous research (Smith et al., 2013). In Study IV, level of consciousness was assessed with the GCS. In addition to the standard NEWS parameters, the BG level was included as a continuous variable in the RF models (III and IV). The training phase was a null operation for the traditional NEWS.

Table 9. Predictor description in Studies III and IV.

Predictor variable	Type of predictor	Measurement
Respiration rate	Discrete numerical	min^{-1}
Oxygen saturation	Continuous numerical	%
Any supplemental oxygen	Categorical	Air or supplemental oxygen
Systolic blood pressure	Continuous numerical	mmHg
Heart rate	Discrete numerical	min^{-1}
Consciousness	Categorical	AVPU or GCS*
Temperature	Continuous numerical	˚C
Blood glucose	Continuous numerical	mmol/l

*The AVPU scale was used in Study III and GCS (Glasgow Coma Scale) was used in Study IV.

Model evaluation was performed with ten-fold stratified cross-validation in which each fold presents a data subset to the RF algorithm and uses a different data subset to estimate predictive performance in the AUROC metric (III, IV) (Refaeilzadeh, Tang, & Liu, 2009). These generated folds and a bootstrap resampling with 10,000 sample points were later used to computationally estimate 95% CI for AUROCs and p-values for the comparison of different predictor models as the normality of the cross-validated AUROC scores was not guaranteed. Bootstrap resampling is a numerical method that approximates the true shape of the AUROC distribution and estimates the sample mean for a non-gaussian distribution (Efron, & Tibshirani, 1994). Bootstrap resampling was applied to the test data set.

The technical details regarding the configuration of the RF algorithm are as follows. The number of features randomly sampled for each split point was a square root of the input feature count. The number of trees chosen was one hundred. Nodes were expanded until all leaves are pure or until all leaves contain less than two samples. Hyperparameter optimisation was not performed since the default values in the sklearn package were considered as good initial guesses.

To assess the power of Studies III and IV, post hoc analyses were conducted with Obuchowski method for one receiver operating characteristic curve (Obuchowski, 1997). A sensitivity analysis based on the last contact in the study period was also performed for Study IV.

4.8 Ethical considerations

All studies had an observational study design, with no interventions or patient contact involved; therefore, the need for informed patient consent was waived according to Finnish legislation. The study protocols (I, II and IV) were approved by the Institutional Review Board and Ethics Committee of the Pirkanmaa Health District (R17156, November 7th, 2017 and R10111, May 5th, 2015). One study protocol (III) was approved by the Department of Emergency Medicine and Services, HUS Helsinki University Hospital (§68, November 11th, 2015). All studies followed the principles of the Helsinki Declaration.

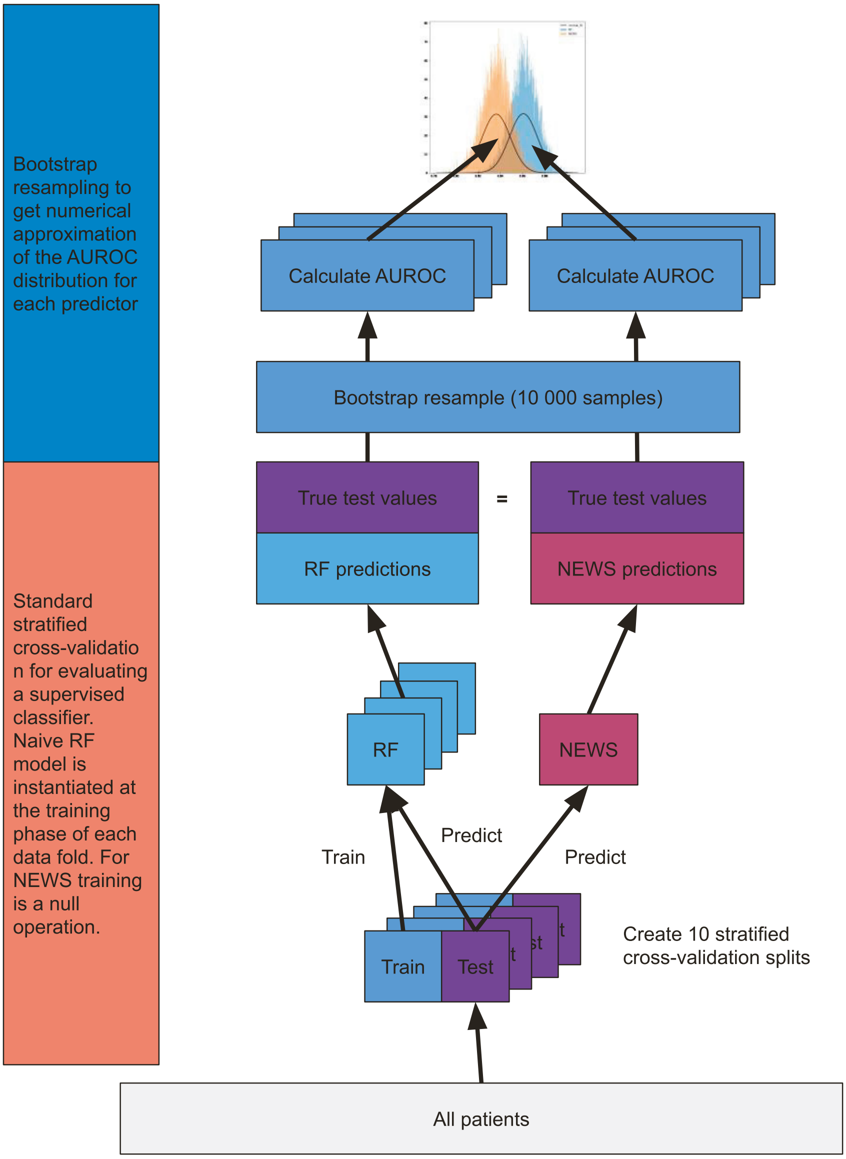

Figure 1. Flowchart of the data analysis process. AUROC = an area under the receiver operating characteristics; NEWS = the National Early Warning Score; RF = random forest. (Machine learning model predicts short-term mortality among prehospital patients: a prospective development study from Finland. Resuscitation Plus 2021;5:100089.)

5 RESULTS

5.1 Characteristics

5.1.1 Study I

In this study, 112 emergency missions regarding suspected, non-traumatic, witnessed OHCA or non-traumatic, EMS-confirmed OHCA were collected. A total of 80 emergency missions were eligible for analysis. The main reasons for exclusion were institutional CPR (n = 14) and awake patient (n = 14); four patients were excluded due to other reason (e.g. poor sound quality).

Of the 78 suspected cardiac arrests, 49 were confirmed as true cardiac arrests at the scene and 29 of the suspected cardiac arrests were regarded as non-cardiac arrest events when EMS personnel contacted the patient. In two cases, the dispatcher had not suspected later confirmed OHCA. Most cardiac arrests were suspected by the dispatcher after an ambulance was dispatched (n = 57). The median duration of spontaneous speech was similar between the groups (5 min 12 sec [IQR, 3:31–6:44] in true OHCA group vs 3 min 55 sec [IQR, 3:05–7:08] in non-OHCA group, p=0.51).

5.1.2 Study II

Figure 2 presents a flowchart of FRU missions during the study period. FRUs were dispatched on a total of 1,894 first-response missions and FRUs attended to patients during 1,622 missions (median age 67 [IQR, 52–81], 59% were male). The most common reason for dispatch was ischaemic stroke (26%). The median response time from dispatch to scene was 9 minutes, and an FRU was the first unit on scene in 860 (53%) missions. In missions in which an FRU encountered the patients prior to ambulance arrival, an FRU reached the scene in a median of 9 minutes (IQR 5–13) before ambulance. A total 1,015 patients were clinically evaluated and FRUs were involved in treatment of 793 patients. Of these patients, 223 (28%) clinical responses were noted.

During the study period, an FRU was dispatched to 1,014 FHQ mission, Table 10. There were 114 confirmed OHCAs and 83 resuscitation attempts. ROSC was obtained in 17 of these cases when CPR was attempted. Professional FRUs encountered 46 OHCA patients who were resuscitated, and volunteer FRUs were in involved in resuscitation of 37 OHCA patients. Professional FRUs had shorter response times in cardiac arrest missions as compared with trained volunteer FRUs (6 min [IQR, 5–9] vs 9 min [IQR, 7–16], respectively; p <0.001). FRUs initiated resuscitation in 42 missions at a median of 4 minutes prior to arrival of ambulance personnel (range 1-18 minutes). The number of patients in whom ROSC was obtained was similar in both groups. ROSC was achieved by an FRU alone in one patient. Regarding other prehospital emergencies, volunteer FRUs administered oxygen more liberally than professional FRUs in chest pain and stroke missions.

Table 10. Professional and volunteer first-responding units in the First Hour Quintet missions in Study II.

	Professional n = 481	Volunteer n = 533	p-value
Cardiac arrest, n (%)			
Confirmed cardiac arrest	71	43	
Resuscitation by the FRU	46 (65)	37 (86)	0.02
ROSC	8 (11)	9 (21)	0.18
Severe respiratory failure, n (%)			
Dispatches	38	76	
Airway management	0	0	
Oxygen administration	16 (42)	35 (46)	0.84
Respiratory state improved	12 (32)	25 (33)	1.00
Chest pain, n (%)			
Dispatches	78	159	
Oxygen administration	16 (21)	77 (48)	<0.001
Medication	10 (13)	24 (15)	0.70
Oxygen and medication	5 (6)	18 (11)	0.25
Chest pain relief	5 (6)	23 (14)	0.09
Shortness of breath improved	4 (5)	24 (15)	0.03
Severe trauma, n (%)			
Dispatches	58	74	
Immobilisation/splinting	13 (22)	14 (19)	0.67
Pain relief	1 (1)	1 (2)	1.00
Stroke, n (%)			
Dispatches	236	181	
Carrying/assistance	94 (34)	36 (27)	0.21
Oxygen administration	9 (4)	26 (14)	<0.001

FRU = first-responding unit; ROSC = return of spontaneous circulation.

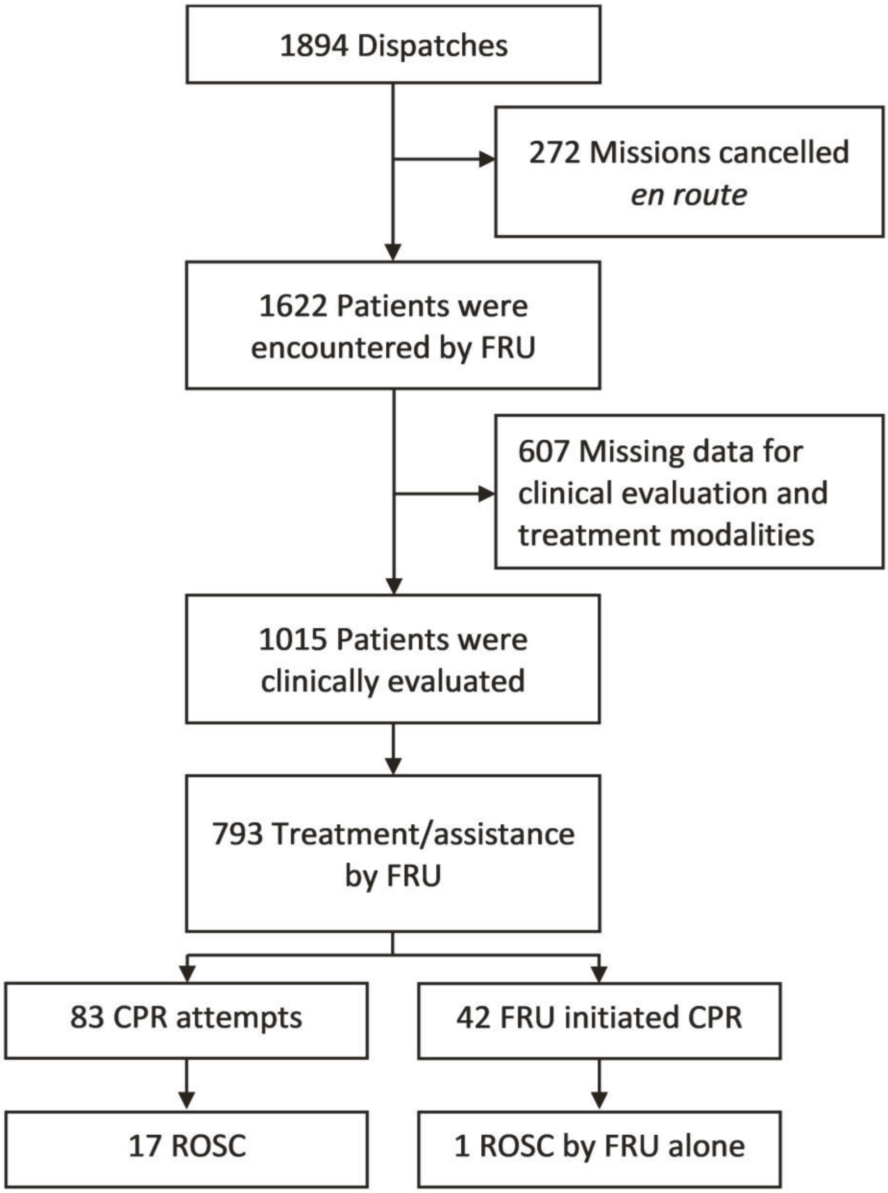

Figure 2. Flowchart of first-responding unit missions in Study II. (Modified from Professional firefighter and trained volunteer first-responding units in emergency medical service. Acta Anaesthesiologica Scandinavica 2019;63:111-116.)

5.1.3 Study III

Of the 583,937 adult emergency patients contacted by EMS in the Helsinki and Uusimaa Hospital district, 35,800 patients (6.1%) had appropriate data to calculate NEWS and 26,458 (4.5%) patients had complete NEWS data and had BG measured. Of the patients with complete vital sign data (including BG), 278 (1.0%) died within one day. Characteristics of the included patients are presented in Table 11.

Table 11. Patient characteristics in Studies III and IV.

	Study III	Study IV	
	Analysed patient n = 26,458	Eligible patient n = 3,632	Analysed patient n = 2,853
Age, mean (SD); years	66 (20)	63 (21)	66 (21)
Male sex, %	48	50	50
NEWS, median (IQR)	3 (1–6)	1 (0–3)	1 (0–3)
0, %	16	29	26
Total 1–4, %	48	58	60
3 in single parameter, %	–	19	21
Total 5–6, %	15	6.3	6.8
Total 7 or more, %	22	6.2	7.1
Respiration rate, median (IQR); min-1	16 (15–20)	16 (15–20)	16 (15–20)
Oxygen saturation, median (IQR); %	96 (93–98)	97 (95–98)	97 (95–98)
Any supplemental oxygen, %	17.2	7.6	8.2
Temperature, median (IQR); ˚C	36.8 (36.3–37.3)	36.7 (36.3–37.1)	36.7 (36.2–37.1)
Systolic blood pressure, median (IQR); mmHg	142 (123–164)	143 (127–163)	143 (127–164)
Heart rate, median (IQR); min-1	87 (73–103)	86 (73–100)	85 (72–100)
Glasgow Coma Scale > 13, %	76.6	94	94
Blood glucose, median (IQR); mmol/l	7.2 (6.0–9.1)	6.6 (5.6–8.2)	6.7 (5.7–8.2)
24-hour mortality, n (%)	278 (1.0)	16 (0.4)	13 (0.5)
48-hour mortality, n (%)	–	22 (0.6)	18 (0.6)
7-day mortality n (%)	615 (2.3)	–	–
30-day mortality, n (%)	1,115 (4.2)	114 (3.1)	97 (3.4)
ICU admission, n (%)	–	46 (1.3)	32 (1.1)
ICU admission/48-hour mortality, n (%)	–	66 (1.8)	49 (1.7)

ICU = intensive care unit; IQR = interquartile range; NEWS = the National Early Warning Score; SD = standard deviation.

5.1.4 Study IV

EMS were dispatched to 6,202 missions and a total of 3,632 individual emergency patients met the inclusion criteria. All NEWS parameters and BG were measured in 2,853 patients who were included in the complete-case analysis. The baseline characteristics for the eligible patients and for the analysed patients are presented in Table 11. The patients in both groups were similar in terms of NEWS parameters, blood glucose level, 30-day, 24-h and 48-h mortality and ICU admission. Of the analysed patients, 40% patients were transported to the ED, 19% to a general practitioner and 34% were left at the scene. As compared with the study population in Study III, the patient population in study IV had lower NEWS score, received supplemental oxygen less frequently and had higher GCS scores. Almost a quarter of the patients in Study III had NEWS scores of 7 or more, whereas only 7.1% of the patients in Study IV had a similar NEWS score, indicating a high risk. Within 30 days after a contact with EMS personnel, 114 eligible patients had died. In 97 of these patients, all NEWS parameters and BG were reported. The majority of deceased patients were admitted to the university hospital.

Missing vital signs in the eligible patients are shown in Table 12. A majority of the patients with any missing NEWS parameter had only one NEWS parameter missing (520/683 = 76%). Temperature, BG and respiration rate were the most common missing vital signs. The level of consciousness and use of supplemental oxygen were documented in all eligible patients.

Table 12. Missing data vital signs (%) in the eligible patients in Study IV.

	Missing vital sign n = 779	One missing NEWS parameter n = 520	Two missing NEWS parameters n = 67
Respiration rate	254 (33)	141 (27)	42 (63)
Oxygen saturation	127 (16)	18 (3)	17 (25)
Temperature	499 (64)	346 (67)	59 (88)
Systolic blood pressure	115 (15)	11 (2)	15 (22)
Heart rate	68 (9)	4 (1)	1 (1)
Blood glucose	403 (52)	174 (33)	46 (69)

NEWS = the National Early Warning Score.

5.2 Dispatcher-suspected cardiac arrest (I, IV)

There were 51 EMS-encountered OHCAs in Study I and 15 confirmed OHCAs in Study IV in the Tampere University Hospital District area during the study periods. Table 13 presents the sensitivities, specificities, positive predictive values and negative predictive values for the emergency dispatcher to recognise a confirmed OHCA. Only the sensitivity and positive predictive value could be calculated for Study I as the study material included no OHCA missions (true negative cases).

Table 13. Performance of emergency dispatcher in recognising out-of-hospital cardiac arrest.

	Study I	Study IV
Dispatcher suspected	78	20
OHCA at the scene	49	10
No OHCA	29	10
OHCA not suspected	2	4138
OHCA at the scene	2	5
No OHCA	-	4133
Sensitivity	96.1	66.7
Specificity	-	99.8
Positive predictive value	62.8	50.0
Negative predictive value	-	99.9

OHCA = out-of-hospital cardiac arrest.

5.3 Trigger words (I)

A total of 291 spontaneous trigger words were observed in 80 emergency calls. Figure 3 shows the distribution (%) of the spontaneous trigger words in the EMS-confirmed true cardiac arrest group and the EMS-confirmed non-cardiac arrest group. No trigger word was associated with confirmed OHCA. 'Is wheezing' and 'collapsed' were frequently used in the true cardiac arrest group ('is wheezing' 33% in true OHCA vs 17% in non-OHCA; OR, 2.40 [95% CI 0.78–7.40] and 'collapsed' 24% in true OHCA vs 7% in non-OHCA; OR, 4.15 [0.86–20.1]). 'Is snoring' was common in the non-cardiac arrest group (2% for true OHCA vs 21% for non-OHCA; OR, 0.08 [95% CI 0.009–0.67]).

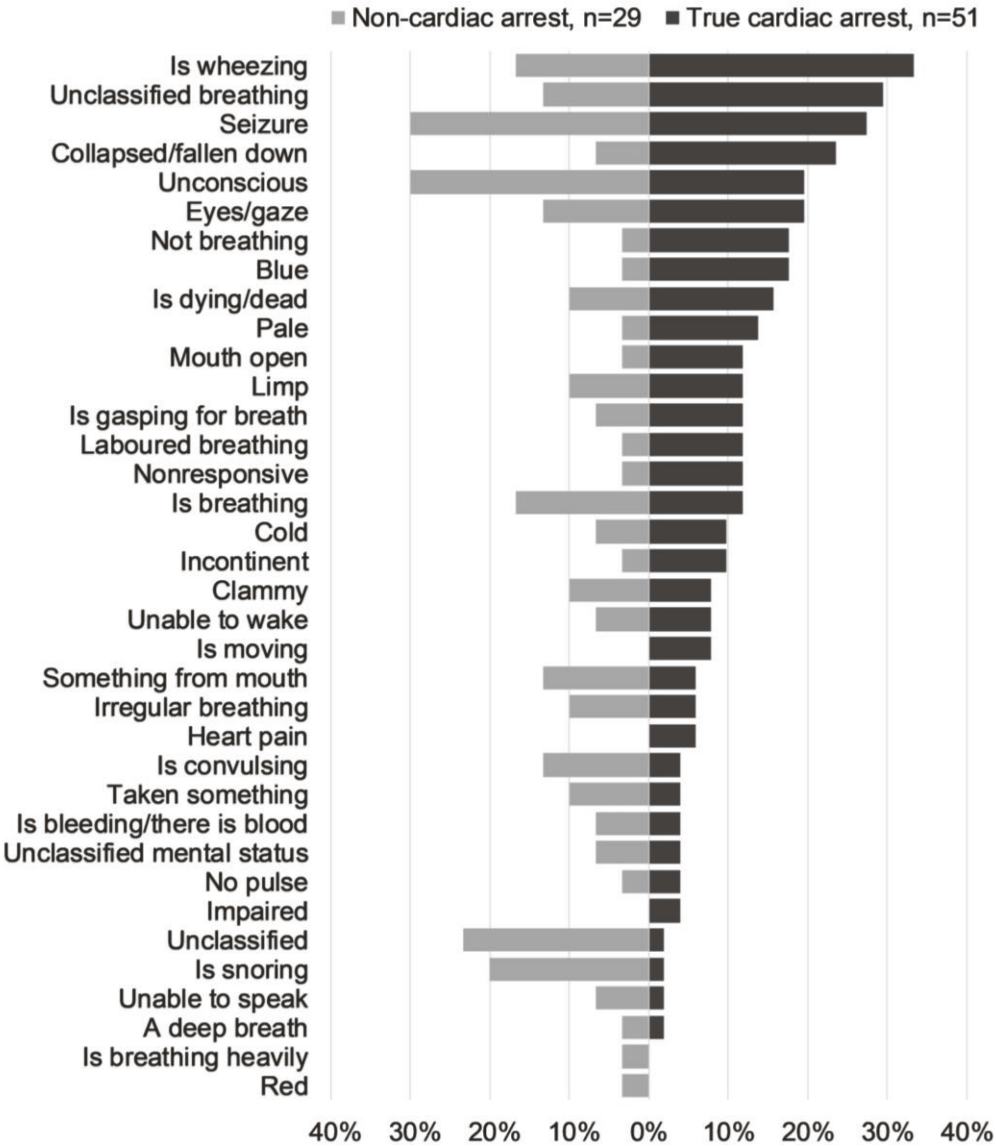

Figure 3. Distribution (%) of the spontaneous trigger words and their association with confirmed cardiac arrests. (Spontaneous trigger words associated with confirmed out-of-hospital cardiac arrest: a descriptive pilot study of emergency calls. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 2020;28:1.)

5.4 Prediction of short-term mortality (III, IV)

In Studies III and IV, physiological measurements, including NEWS parameters and BG, were used to develop the RF models. The primary analysis of Study III (n = 26,458) demonstrated that the RF model which included BG had greater AUROCs for predicting 24-hour mortality compared with the standard NEWS (AUROC for RF, 0.868 [95% CI 0.843–0.892] vs AUROC for NEWS, 0.836 [95% CI 0.810–0.860]; p <0.001), Table 14. In a secondary analysis including patients with all NEWS parameters measured (n = 35,800), the RF model without BG performed similarly to the RF model trained with NEWS parameters and BG in the primary analysis.

Table 14. Areas under the receiver operating characteristics curve with 95% confidence intervals in Study III.

	NEWS	RF 1	RF 2
Primary analysis			
BG and NEWS measured (n = 26,458)			
24-h mortality	0.836 (0.810–0.860)	0.858 (0.832–0.883)	0.868 (0.843–0.892)
Secondary analysis			
NEWS measured (n = 35,800)			
24-h mortality	0.850 (0.829–0.868)	0.873 (0.854–0.892)	-

BG = blood glucose; NEWS = the National Early Warning Score; RF 1 = random forest trained with NEWS parameters only; RF 2 = random forest trained with NEWS parameters and blood glucose.

In study IV, BG slightly improved the performance of the RF model, and the RF model showed better performance for predicting 30-day mortality compared to NEWS (AUROC for RF including BG, 0.758 [95% CI 0.705–0.807] vs AUROC for NEWS, 0.682 [95% CI 0.619–0.744]; p <0.001), Table 15. A sensitivity analysis based on the last contact showed only minor changes to the models' performances. In a secondary analysis that included the last contact of all eligible patients regardless of the number of documented vital signs, the results were essentially unchanged.

Post hoc analysis regarding the power of Studies III and IV confirmed our assumption of adequate power (Study III, significance level = 0.05; power = 0.95; AUROC = 0.56; Study IV, significance level = 0.05; power = 0.95; AUROC = 0.60). None of the 95% CIs of the AUROCs in the primary analyses (Table 14–15) crossed these AUROC values.

Table 15. Areas under the receiver operating characteristics curve with 95% confidence intervals in Study IV.

	NEWS	RF 1	RF 2
Primary analysis			
First contact, complete-case analysis (n = 2,853)			
30-d mortality	0.682 (0.619–0.744)	0.735 (0.679–0.787)	0.758 (0.705–0.807)
24-h mortality	0.890 (0.797–0.966)	0.875 (0.707–0.976)	0.940 (0.860–0.985)
48-h mortality	0.845 (0.729–0.936)	0.808 (0.629–0.957)	0.881 (0.751–0.972)
ICU admission	0.806 (0.715–0.887)	0.807 (0.714–0.890)	0.814 (0.726–0.892)
ICU admission or 48-h mortality	0.818 (0.749–0.882)	0.811 (0.739–0.877)	0.847 (0.785–0.902)
Secondary analyses			
Last contact, complete-case analysis (n = 2,853)			
30-d mortality	0.680 (0.614–0.743)	0.734 (0.672–0.791)	0.756 (0.701–0.808)
24-h mortality	0.909 (0.826–0.975)	0.916 (0.826–0.974)	0.954 (0.914–0.988)
48-h mortality	0.909 (0.827–0.975)	0.872 (0.711–0.971)	0.945 (0.895–0.986)
ICU admission	0.807 (0.717–0.890)	0.828 (0.740–0.903)	0.825 (0.743–0.899)
ICU admission or 48-h mortality	0.833 (0.763–0.896)	0.814 (0.742–0.881)	0.854 (0.791–0.907)
Last contact, missing data included (n = 3,632)			
30-d mortality	0.700 (0.642–0.753)	0.737 (0.683–0.787)	0.756 (0.705–0.804)
24-h mortality	0.886 (0.811–0.951)	0.871 (0.754–0.949)	0.923 (0.871–0.963)
48-h mortality	0.834 (0.732–0.917)	0.806 (0.659–0.925)	0.841 (0.709–0.947)
ICU admission	0.816 (0.747–0.881)	0.846 (0.779–0.904)	0.851 (0.787–0.906)
ICU admission or 48-h mortality	0.820 (0.763–0.872)	0.846 (0.792–0.894)	0.852 (0.797–0.901)

ICU = intensive care unit; NEWS = the National Early Warning Score; RF 1 = random forest trained with NEWS parameters only; RF 2 = random forest trained with NEWS parameters and glucose.

6 DISCUSSION

6.1 Summary of the main findings

Studies I and II focused on the early links in the chain of survival. In Study I, spontaneous trigger words that were associated with OHCA were examined. This hypothesis-generating pilot study found that 'is wheezing' and 'collapsed' were frequently used in the true cardiac arrest group, whereas 'is snoring' was common in the non-cardiac arrest group. No spontaneous trigger word was associated with EMS-confirmed OHCA.

In Study II, the general performance of the professional and the trained volunteer FRUs is EMS missions was described. The FRUs attended to a patient in a total of 1,622 missions. During the study period, the EMS personnel and the FRUs were dispatched to 1,014 FHQ missions and encountered 83 OHCAs in which CPR was attempted. The study showed that an FRU initiated resuscitation in half of the OHCAs at a median of 4 minutes prior to arrival of ambulance personnel, and the professional firefighter FRUs had shorter response times than the trained volunteer FRUs. ROSC was achieved in 20% of all OHCA patients.

Studies III and IV concentrated on the prediction of short-term mortality in EMS-encountered emergency patients. Study III found that a RF machine learning method including NEWS parameters outperformed the standard NEWS in predicting 24-hour mortality in adult prehospital patients, both with and without a BG variable. In Study IV, the RF models showed fair performance for predicting 30-day mortality, whereas the RF model that included BG had excellent performance in predicting 24-hour mortality.

6.2 Interpretations of the results

6.2.1 Spontaneous trigger words in OHCA (I)

Early recognition of OHCA is an essential link in the chain of survival as the dispatcher may active the EMS system and direct the caller to initiate CPR unless bystander CPR is already being performed. Dispatchers' ability to recognise OHCA was assessed in Studies I and IV. These Studies found that the sensitivity for OHCA recognition was 0.96 and 0.67, respectively. The large difference in sensitivities may be attributable to random error, as only 15 confirmed OHCAs occurred in Study IV. The ILCOR 2020 CoSTR showed that the sensitivity for a dispatcher-recognised OHCA is 0.79 (Olasveengen et al, 2020).

Could spontaneous speech be used to improve the algorithm's sensitivity for OHCA recognition without decreasing its specificity? At the beginning of an emergency call or after the dispatcher's standardised questions, a caller's spontaneous speech might give clues to the dispatcher regarding an ongoing medical emergency. According to the ERC 2021 guidelines, the dispatcher diagnosis of OHCA is based on the combination of the patient being noted as unconscious and not breathing or breathing abnormally (Olasveengen et al., 2021). Therefore, an individual trigger was postulated to possibly combine the semantic information in relation to both the level of consciousness and breathing in OHCA patients.

In our pilot study (I), two noteworthy spontaneous trigger words were identified: 'is wheezing' (Finnish: korisee) and 'is snoring' (Finnish: kuorsaa). The former does not mean obstructive wheezing but rather a death rattle, and it seems to be an idiomatic expression in the Finnish language. Both 'is wheezing' and 'is snoring' mean that the patient has difficulties maintaining the normal muscle tone of the upper respiratory tract, which, in turn, reflects a significantly altered level of consciousness. Interestingly, 'is wheezing' was the most frequently used single trigger word among confirmed OHCA victims.

Agonal breathing seems to be a pitfall in OHCA recognition since laypersons' descriptions of breathing may mislead the dispatcher. As Riou et al. suggested, the emergency dispatcher should repeat the question regarding breathing if the caller's initial answer is imprecise or vague (Riou, Ball, Williams, et al., 2018). Our material contained, altogether, nine breathing-related trigger word categories and a total of

19 unclassified breathing-related spontaneous trigger words in both groups. Additionally, 'is wheezing' was observed in one of the two later confirmed OHCAs that the dispatcher had missed. Some of the lay descriptions may possibly represent a patient's agonal breathing. Interestingly, the new ERC 2021 guidelines emphasise for the first time that slow, laboured breathing should be considered a sign of cardiac arrest (Olasveengen et al., 2021).

Spontaneous trigger words may facilitate the dispatcher to recognise OHCA. It may not be rational to add any spontaneous trigger word into the current protocol for OHCA recognition as the standardised questions should be simple and unambiguous. Conversely, the dispatcher should be vigilant and ask always about breathing when a layperson describes that an unconscious patient is wheezing.

6.2.2 Volunteer and firefighter FRUs in the Finnish EMS system (II)

The main conclusion of Study II was that the FRUs treated the patient or assisted ambulance personnel in half of the missions. The key question that arises is why this occurred in only half the cases? The answer to that question might be incomplete reporting rather than FRUs not knowing what to do. Overall, 607 (37%) of a total of 1,622 encountered patients had no documentation about their patient history, clinical evaluation or physiological measurements. This may reflect a simultaneous or nearly simultaneous arrival with ambulance personnel and that an FRU had not had time to evaluate and treat the patient. As a matter of fact, an FRU was confirmed to be the first unit on the scene in only 53% of all the cases. When an FRU was the first unit on the scene, it had a median of 9 minutes time to act before ambulance personnel arrived.

Additionally, unlike FRUs in most other European countries, the Finnish FRUs include both firefighter and trained volunteer first responders who were dispatched not only to cardiac arrest missions but also to stroke, chest pain, respiratory failure and trauma missions. Besides assisting ambulance personnel, FRUs had limited treatment options, for example, in presumed stroke missions. This fact may explain why the trained volunteer FRUs administrated oxygen more liberally than the professional FRUs in chest pain and stroke missions. After the conduction of Study II, it has been demonstrated that dispatching FRUs to stroke missions was not associated with a reduced on-scene time (Puolakka, Väyrynen, Erkkilä, & Kuisma, 2016). Furthermore, the magnitude of other treatment modalities (e.g. airway management, medication or immobilisation) was small in relation to the entire study material.

An FRU was infrequently dispatched to cardiac arrest missions since EMS-confirmed OHCAs comprised only 9% (144/1,622) of all missions. However, the median response time for the professional FRUs and for the trained volunteer FRUs were 6 and 9 minutes, respectively, which were similar to other FRUs in Europe. An FRU initiated CPR at a median of 4 minutes prior to ambulance arrival in 42 (51%) missions in which CPR was attempted. Given that approximately 400 volunteers participate as 30 layperson-staffed FRUs in the EMS system, a majority of the individual volunteer responders were not involved in resuscitation efforts during the one-year study period. This will inevitably affect the volunteers' CPR skill retention and will probably reduce the quality of CPR. As the chances for survival decrease 10% for every minute delay before initiation of CPR (Valenzuela et al., 1997), it could be speculated that the observed four-minute time advantage over ambulance units in resuscitation may have improved patient outcomes. Nevertheless, the data regarding survival to hospital discharge or 30-day survival was not obtained for Study II. The comparison between the professional and trained volunteer FRUs showed that the volunteer FRUs were more likely to be involved in CPR attempts as compared to professional FRUs, which may reflect EMT-staffed professional FRUs' stronger adherence to the national resuscitation guidelines when CPR is not attempted.

Due to a large degree of heterogeneity in the FRU teams' skill levels and the short period of time available to evaluate and treat the patient, quantifying the FRUs' performance with statistics or with hard outcomes is difficult. However, the FRUs' participation in the EMS response is warranted, especially in sparsely populated rural areas. Although an FRU may not be always capable of providing any medical treatment, the presence of a local citizen responder, in itself, may result in relief of the patient's symptoms before the ambulance staff has reached the scene. In addition, an FRU is a valuable resource for the resuscitation team when CPR is attempted in the prehospital setting. The implementation of FRUs is feasible, as both trained volunteers and professional fire departments do contribute to the first tier of the Finnish EMS response.

6.2.3 Prediction of short-term mortality in the prehospital setting (III, IV)

Cardiac arrest does not equal mortality; for instance, death may be attributed to cardiac causes as well as brain death or multiple organ dysfunction. Although the chain of survival for OHCA is a key element in this thesis, short-term mortality was chosen as the primary outcome instead of cardiac arrest in Studies III, IV for the following reasons. First, cardiac arrest encompasses a heterogenous group of aetiologies. According to updated Utstein-style reporting, the pathogenesis of cardiac arrest includes medical causes, traumatic causes, drug overdose, drowning, electrocution and asphyxia (Perkins et. al., 2015). Second, sudden cardiac arrest is less predictable event than short-term mortality since disturbances in vital functions do not necessary precede cardiac arrest. It has been shown that the commonly used EWSs perform better in predicting short-term mortality than cardiac arrest (Smith et. al., 2013). Third, a patient in unwitnessed cardiac arrest and a dead patient without secondary signs of death are clinically indistinguishable and their initial treatment is identical.

What would an ideal prehospital risk stratification tool look like? NEWS and other EWS systems were originally developed for the requirements of the in-hospital setting. If NEWS indicates a medium or a high risk in a hospital ward, an urgent clinical assessment will be performed by a rapid response team (Royal College of Physicians, 2017). The purpose of this assessment is to recognise and prevent clinical deterioration and to reduce inpatient morbidity and mortality. In the prehospital setting, however, a good risk stratification tool should also address an additional clinical dilemma. Besides detecting a physiologically deteriorating patient in a heterogenous patient cohort, this risk stratification tool could also aid a clinician to decide an appropriate transportation destination for the patient and to identify those low-risk patients that can be safety left at the scene. Similarly, identification of a high-risk emergency patient may prompt EMS personnel to contact an emergency physician earlier. Estimates of short-term mortality ranging from 24 hours to 30 days could supplement the available information on the patient history and clinical findings and may have an influence on a clinician's decision regarding emergency patient's care.

In Study III, the 24-day mortality was selected as the primary outcome measure since it was assumed to be the most suitable hard outcome for the prehospital setting. The early signs of impending physiological deterioration are present hours before cardiovascular collapse (Schein et al., 1990). Silcock et al. found that NEWS predicted 24-hour mortality and 48-hour mortality more accurately than 30-day mortality among unselected prehospital patients (AUROC for 24-hour mortality, 48-hour mortality and 30-day mortality 0.855 [95% CI 0.69–1], 0.871 [0.75–0.98] and 0.740 [0.661–0.819]; $p < 0.0001$ for each pairwise comparison) (Silcock et al., 2015). Additionally, Smith et al. reported that NEWS showed a greater ability to predict 24-hour mortality than 33 other EWSs among hospitalised patients (AUROC, 0.894 [95% CI 0.887–0.902] for NEWS (Smith et al., 2013). Similarly, Churpek et al. included 24-day mortality in their primary composite outcome in their machine learning study (Churpek et al., 2016).

By contrast, 30-day mortality was considered an appropriate primary outcome for Study IV since the patient population was less severely ill than those in Study III. In Study IV, the 30-day mortality rate was 3.4% and the 24-hour mortality rate was only 0.5% among the analysed patients. The primary analysis included those patients that were left at the scene or transported to a general practitioner, and only 40% of the analysed patients were transported to the ED. The study population had also a low clinical risk for deterioration in terms of NEWS score. These facts may explain the lower mortality rate in Study IV compared with other prehospital NEWS studies presented in Table 7.

Machine learning methods have many advantages over the traditional track-and-trigger EWS systems. First, the main strength of these EWS systems is their simplicity: an aggregate score can be easily calculated with pen and paper. However, the traditional EWS systems are based on simple linear logistic regression models, which limits their performance. The traditional regression models assess each predictor variable separately and they do not consider plausible non-linear associations or interactions between the predictor variables. Second, miscalculation of the total risk score is an inherent limitation in all EWS systems, but automation could be used to overcome human errors of this kind (Downey, Tahir, Randell, Brown, & Jayne, 2017). Third, automatic data collection from electrical medical records enables the development of advanced risk calculators which outperform the traditional risk scores (Linnen et al., 2019). Automatic data collection and monitoring of vital signs can reduce patient mortality in hospital wards; however, automatic calculation of the traditional NEWS did not reduce in-hospital mortality or ICU admission rates (Bedoya et al., 2019). Nevertheless, a large multicentre study found that an implementation of a complex risk stratification model resulted in a lower 30-

day mortality in hospitalised patients (adjusted RR, 0.84 [95% CI 0.78–0.90]) (Escobar et al., 2020). The latter model was based on electronic patient records and included vital signs, laboratory studies, severity of illness and coexisting conditions. Finally, no traditional EWS system is applicable to all patients in the prehospital setting. For instance, NEWS should not be applied to the paediatric or obstetric patient population or to patients with suspected acute myocardial infarction, gastrointestinal bleeding or spinal cord injury.

Complex machine learning models could be also developed for the purposes of the prehospital setting. Apart from Studies III and IV, one previous study has examined prehospital machine learning–based risk scores (Spangler et al., 2019). In that observational study, various machine learning methods (tree-based models including the RF method, support vector machines and neural networks) were derived from a retrospective dataset and validated in a prospectively collected dataset. The results were also compared to an analysis based on five-fold cross-validation of the retrospective dataset. A total of 30 predictor parameters (e.g. NEWS parameters, clinical signs, paramedic interventions and patient medical history) were included in that machine learning model. The authors reported that a gradient boosting technique (XGBoost) performed at least as well as other machine learning methods, and a model based on this method outperformed NEWS in predicting 48-hour mortality (AUROC, 0.89 [0.87–0.92] vs 0.85 [0.81–0.88], respectively). Similar results were observed when the prospectively tested model was compared to the cross-validated model.

In Studies III and IV, BG was hypothesised to be an informative additional parameter in the prehospital machine learning model. These studies showed that BG only slightly improved the RF models' predictive performance for all outcomes. The RF models that included BG outperformed the standard NEWS for predicting 24-hour mortality in Study III and 30-day mortality in Study IV in terms of statistical testing. Nonetheless, the clinical significance of these findings can be questioned since the 95% CIs for the AUROCs overlapped to a rather large degree, and one may argue that this only demonstrates the non-inferiority of the standard NEWS in the prehospital setting. Nevertheless, the RF model that included BG showed consistently fair performance in relation to all outcome measures in Study IV, as the lower boundaries of the 95% CIs for the AUROCs were always greater than 0.70.

Missing vital signs may reflect a medical emergency. For example, the patient may be bradycardic and therefore oxygen saturation and blood pressure could be unmeasurable. The ambulance personnel may not also have time to document the altered level of consciousness while they are concentrating on the management of the patient's bradycardia. If a traditional EWS system was used in that kind of scenario and a missing vital sign was assumed to be normal, the aggregate score would underestimate the risks for adverse outcomes. Conversely, a machine learning algorithm can use a missing vital sign as a piece of information in refining its predictions.

Missing data were speculated to improve the RF models' predictive performance. In Study IV, body temperature, respiration rate and BG were the most common missing vital signs, although the paramedics were mandated to measure all NEWS parameters. This might be a systematic flaw in documenting case report forms, as the paramedics may not have considered measuring these vital signs necessary in psychiatric patients, for example. However, a secondary analysis of all eligible patients, regardless of the number of measured vital signs, showed that the RF models' predictive performance was essentially unchanged. The RF algorithm's ability to utilise missing data would possibly have been seen in a larger patient cohort.

6.3 Methodological aspects

6.3.1 Internal validity

All Studies I–IV were single-centre observational studies, and Studies I–III had a retrospective study design. The major weaknesses of retrospective studies include, but are not limited to, selection bias, reporting bias and uncontrolled confounding factors, and these result in a high risk of systematic error. Nevertheless, structured reporting guidelines for observational studies were used while preparing these manuscripts. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline was used in Studies I–II and Transparent Reporting of a Multivariable Prediction Model for Individual Diagnosis or Prognosis (TRIPOD) guideline was used for Studies III and IV. No reporting guideline existed for prognostic machine learning studies, as the development of TRIPOD extension specific to machine learning is currently in progress (Collins & Moons, 2019).

In Study I, the diagnosis of OHCA was not based on patient records but on transportation codes in the ELS database. The authors were not blinded to the outcome when transcribing the emergency calls or categorising the trigger words. The categorisation was based only on one previously published study (Berdowski et al., 2009). In addition, an agreement between the authors EL and JK was not assessed while the authors reviewed the emergency calls, although this could have been measured by cross-abstracting a certain proportion of the calls. However, the author JT reviewed the categorisation after all the trigger words were transcribed. Finally, the exact time of trigger words and the time of OHCA suspicion in an emergency call were not considered in our analysis; however, this was not the aim of the study.

In Study II, the exact interval between patient evaluation, physiological measurements and treatment responses could not be determined, and the outcome measures may have been affected by the EMS. The FRUs are advised to describe only the assessment and treatment provided by the FRU in their documentation, whereas for EMS units fill out a separate form of documentation. In cases of the simultaneous arrival of the FRU and the EMS unit at the scene, some of the procedures performed by the EMS personnel may have also been documented on the FRU forms. Indeed, during certain missions, the FRU was always dispatched, regardless of the time benefit compared with the EMS ambulance. The arrival of the FRU and the initial treatment during these missions may well have occurred simultaneously or even after arrival of the ambulance. Nevertheless, the magnitude of these procedures (e.g. airway management, medication) is small in relation to the entire material, suggesting that the role of the FRU, in this sense, is not strong. Regarding OHCA missions, patients with an initial shockable rhythm or an initial non-shockable rhythm were not analysed separately, and the number of shocks delivered by an FRU was unknown. Whether an OHCA was witnessed or bystander CPR was provided was also not documented.

As the study material consisted of the FRUs' documentation, Study II had no control group for the FRUs, which complicates any determination of whether an FRU had an independent influence on patient outcomes or clinical responses (e.g. ROSC). As already discussed, these documentations were not filled in the cases of the simultaneous arrival of the FRU and the EMS unit or if the ambulance personnel had reached the patient first. Thus, these missions could not be used as a control group for the FRU-attended missions.

The data source in Study III was originally designed to serve the needs of prehospital care rather than to be used for research purposes and the development of a machine learning model. Due to its retrospective design, the study population in Study III was highly selected. Only 4.5% of the encountered adult patients were included in the primary analysis. This bias becomes apparent when the study populations in Study III and IV are compared. The patients in Study III had greater NEWS scores, indicating higher risk and a more severely ill patient population (Table 11). Nevertheless, the most severely ill patients may have been excluded from both Studies III and IV since the ambulance personnel may not have had time to document all studied vital signs despite being mandated to do so.

All primary analyses in Studies I–IV were complete-case analyses, and no imputation method was applied to address missing data. In Study II, no imputation method was used to handle missing data, as the study aimed to describe mission and patient characteristics. In Studies III and IV, multiple imputation was not used to calculate NEWS score in cases with missing vital signs. In a previous machine learning study, multiple imputation was applied to calculate NEWS score in cases with one or two missing vital signs (Spangler et al., 2019). Multiple imputation was not used in Study IV, as the EMS personnel were mandated to measure all vital signs, and thus, missing measurements may not occur randomly. Since the baseline characteristics were similar between the eligible patients and the analysed patients, there may not be any substantial correlation between missingness and outcomes. However, it is possible that the EMS had encountered a few critically ill patients for whom they had not managed to measure all vital signs and BG.

Random error can be controlled with an adequate sample size. Nevertheless, no sample size calculations were performed for Studies I–IV, since Study I was a hypothesis-generating pilot study, Study II was a retrospective chart review and Study III and IV were post hoc analyses. Unfortunately, Study I was underpowered to detect any association between confirmed cardiac arrests and trigger words, and the 95% CIs for ORs were wide in the logistic regression model. Evaluation of the power of Studies III and IV is more complex. In the primary analyses of Studies III and IV, the RF models had greater AUROCs than the standard NEWS according to the p-values in the statistical hypothesis tests, but the AUROC 95% CIs overlapped to a significant degree. Although the p-values and the 95% CIs seem to be conflicting, the null hypothesis can be rejected in this kind of scenario (Austin & Hux, 2002; Knezevic, 2008). Importantly, as with any binary prediction, the critical issue is not the total sample size but the numbers in the smaller group. Only 278 (1.0%) and 97 (3.4%) deaths were reported in the primary analyses of Studies III and IV.

Machine learning models have important methodological limitations which should be borne in mind when interpreting the results (Chen & Asch, 2017; Collins & Moons, 2019). First, the selection of too many predictor variables for the sample size may cause overfitting. Although the RF algorithm is thought to be resistant to overfitting and only eight predictor variables were used in Studies III and IV, the RF model's excellent performance in the secondary analysis of Study IV may, in part, be attributable to overfitting due to the small number of cases (deaths within 24 or 48 hours). Second, the RF machine learning algorithm has a significant 'black box' element, which is a metaphor for model building and its interpretation. This means that one cannot predict how the RF algorithms generate their decision trees, and the interpretation of the regression coefficients in the decision trees is extremely difficult. Third, the RF models in Studies III and IV are derived from data that date back to 2008–2015 and 2015, respectively. Clinical prehospital care may have changed in the study areas over the past years, and this may decrease the usability of these models in the present clinical practice. Fourth, the most suitable machine learning method for risk stratification in the prehospital setting is unknown. The RF method was used in Studies III and IV since it was the most powerful model among various machine learning methods in a multicentre in-hospital trial (Churpek et al., 2016). Finally, even a perfectly calibrated machine learning model may not translate into better clinical outcomes if its user is not reacting appropriately to the model's predictions. A clinician has still to decide what to do in order to prevent an adverse outcome from happening.

6.3.2 External validity

Studies I–IV were single centre studies in one hospital district in Finland; Studies I, II and IV were conducted in the Tampere University Hospital district and Study III was undertaken in the Helsinki and Uusimaa Hospital district. Study I was conducted in one dialect area in Finland. Additionally, Finnish is a small language, with roughly six million speakers worldwide (Institute for the Languages in Finland, 2021). These facts limit the generalisability of the results outside the study region. Correspondingly, Study II described trained volunteer and firefighter FRUs' performance as a part of the EMS response in a suburban and rural hospital district in Finland. In most other EMS systems, trained volunteers are not considered equal first responders compared with professional firefighter or police units. Therefore, the results may not be applicable outside of the Northern countries.

The predictive models presented in Studies III and IV were not validated in a separate, prospectively collected dataset, which may slightly decrease their external validity. As Spangler et al. (2019) observed in their study, no substantial difference was noted in the predictive performance between the testing with the prospective dataset and the cross-validation dataset. Additionally, the risk stratification models in Studies III and IV were tailored specifically to patient populations in the study regions as the training data were collected with a narrow scope, and thus, the results may not be generalisable to other settings. In the context of this thesis and the study material, 'artificial intelligence' should be considered a sophisticated model which can give accurate answers to a simple and narrow question. This means that the RF models yielded excellent predictions for mortality among the Finnish prehospital patients by modelling complex nonlinearities visible in the vital functions prior to impending cardiovascular collapse.

6.4 Future implications

Study I was a hypothesis-generating pilot study which enables sample size calculations for future studies. Further work is required to determine whether universal descriptors of cardiac arrest exist, alone or in combination, that can help predict cardiac arrest in broader regional, national and continental contexts. Additionally, future studies should find trigger words that are associated with high-risk patients in other prehospital emergencies. For instance, patients with ischaemic stroke who are candidates for mechanical thrombectomy should ideally be recognised at the beginning of the EMS response.

Future studies examining FRUs in the EMS system should have a prospective study design, since Study II has many limitations due to its retrospective design. Most importantly, incomplete reporting of FRU missions resulted in a large degree of missing data in Study II. The FRUs' contribution to an EMS response and their effect on patient-centred outcomes could also be evaluated in a future randomised controlled trial.

6.4.1 Machine learning in the prehospital setting

Future machine learning studies on emergency calls should evaluate whether automatic speech recognition techniques could facilitate recognition of OHCA and other high-risk prehospital emergencies by the dispatcher. Trigger words and their combinations could be identified in real time by automatic speech recognition, and a probability of cardiac arrest could be calculated by an algorithm. A recently published Danish study evaluated a machine learning algorithm's ability to recognise OHCA and compared its performance to that of emergency dispatchers (Blomberg et al., 2019). The study showed that Corti AI had a higher sensitivity but a lower specificity for OHCA recognition compared with emergency dispatchers (sensitivity 0.84 vs 0.73, specificity 0.97 vs 0.99). A major limitation of the study is that the results may not be applicable outside the study region. Additionally, the machine learning algorithm has been developed to recognise agonal breathing in audio recordings that included confirmed OHCAs or normal sleep data (sensitivity 0.972, specificity 0.995), but future research is needed for its implementation in real-world conditions (Chan, Rea, Gollakota & Sunshine, 2019).

Future machine learning studies on risk stratification should consider adding more variables that are readily available in the prehospital setting to their predictive models. Indeed, other valuable information apart from vital signs could reside in patient record systems. Possible candidate variables could include paramedic's worry about the patient's condition, the patient's chief symptom, the presumed diagnosis at the scene, other simple clinical findings (e.g. breathing sounds or capillary refill) or the raw electrocardiography signal. Unfortunately, reporting of transportation codes is currently heterogeneous in our region, as no strict guideline or protocol exists (e.g. a dispatch code is used as a transportation code). These various reporting styles could be standardised if the International Classification of Primary Care (ICPC) classification were adopted. Furthermore, future studies should evaluate whether the implementation of advanced risk stratification tools in the prehospital setting affects patient outcomes.

7 SUMMARY AND CONCLUSIONS

The aims of this thesis were to examine the general performance of the professional and trained volunteer FRUs and their contribution to an EMS response, to identify laypeople's spontaneous trigger words in emergency calls that are associated with OHCA and to establish machine learning models based on readily available measurements for predicting short-term mortality in the prehospital setting. In the context of OHCA, the conclusions were as follows:

1. None of the laypeople's trigger words were associated with confirmed OHCA but 'is wheezing' (Finnish: 'korisee') was a frequently used spontaneous trigger word among the OHCA patients.
2. Professional- and layperson-staffed FRUs shortened the delay from cardiovascular collapse to the initiation of CPR in OHCA victims in half of the cases.
3. RF machine learning models that included NEWS variables and BG outperformed the standard NEWS and had a fair performance for predicting short-term mortality in the prehospital setting in two hospital districts in Finland.

REFERENCES

Abbott, T. E. F., Cron, N., Vaid, N., Ip, D., Torrance, H. D. T., & Emmanuel, J. (2018). Pre-hospital National Early Warning Score (NEWS) is associated with in-hospital mortality and critical care unit admission: A cohort study. *Annals of Medicine and Surgery*, *27*, 17-21.

Akre, M., Finkelstein, M., Erickson, M., Liu, M., Vanderbilt, L., & Billman, G. (2010). Sensitivity of the pediatric early warning score to identify patient deterioration. *Pediatrics*, *125*, e763-769.

Al-Dury, N., Ravn-Fischer, A., Hollenberg, J., Israelsson, J., Nordberg, P., Strömsöe, A., … Rawshani, A. (2020). Identifying the relative importance of predictors of survival in out of hospital cardiac arrest: a machine learning study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *28*, 60.

American Heart Association (2018). CPR through history. (Accessed 28 February 28 2021, at. https://www.heart.org/en/news/2018/05/01/cpr-through-history).

Andelius, L., Malta Hansen, C., Lippert, F. K., Karlsson, L., Torp-Pedersen, C., Kjær Ersbøll, A., … Folke, F. (2020). Smartphone activation of citizen responders to facilitate defibrillation in out-of-hospital cardiac arrest. *Journal of the American College of Cardiology*, *76*, 43-53.

Ashoor, H. M., Lillie, E., Zarin, W., Pham, B., Khan, P. A., Nincic, V., … Tricco, A. C. (2017). Effectiveness of different compression-to-ventilation methods for cardiopulmonary resuscitation: A systematic review. *Resuscitation*, *118*, 112-125.

Austin, P. C., & Hux, J. E. (2002). A brief note on overlapping confidence intervals. *Journal of Vascular Surgery*, *36*, 194-195.

Bång, A., Herlitz, J., & Martinell, S. (2003). Interaction between emergency medical dispatcher and caller in suspected out-of-hospital cardiac arrest calls with focus on agonal breathing. A review of 100 tape recordings of true cardiac arrest cases. *Resuscitation*, *56*, 25-34.

Barry, T., Doheny, M. C., Masterson, S., Conroy, N., Klimas, J., Segurado, R., … Bury, G. (2019). Community first responders for out-of-hospital cardiac arrest in adults and children. *Cochrane Database of Systematic Reviews*, *2019,* CD012764.

Bedoya, A. D., Clement, M. E., Phelan, M., Steorts, R. C., O'Brien, C., & Goldstein, B. A. (2019). Minimal impact of implemented early warning score and best practice alert for patient deterioration. *Critical Care Medicine*, *47*, 49-55.

Berdowski J., Beekhuis, F., Zwinderman, A. H., Tijssen, J. G. P., & Koster, R. W. (2009). Importance of the first link description and recognition of an out-of-hospital cardiac arrest in an emergency call, *119*, 2096-2102.

Berdowski, J., Berg, R. A., Tijssen, J. G., & Koster, R. W. (2010). Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. *Resuscitation*, *81*, 1479-1487.

Berdowski, J., Blom, M. T., Bardai, A., Tan, H. L., Tijssen, J. G. P., & Koster, R. W. (2011). Impact of onsite or dispatched automated external defibrillator use on survival after out-of-hospital cardiac arrest. *Circulation*, *124*, 2225-2232.

Berglund, E., Claesson, A., Nordberg, P., Djärv, T., Lundgren, P., Folke, F., … Ringh, M. (2018).

A smartphone application for dispatch of lay responders to out-of-hospital cardiac arrests. *Resuscitation*, *126*, 160-165.

Blom, M. T., Beesems, S. G., Homma, P. C. M., Zijlstra, J. A., Hulleman, M., Van Hoeijen, D. A., … Koster, R. W. (2014). Improved survival after out-of-hospital cardiac arrest and use of automated external defibrillators. *Circulation*, *130*, 1868-1875.

Blomberg, S. N., Folke, F., Ersbøll, A. K., Christensen, H. C., Torp-Pedersen, C., Sayre, M. R., … Lippert, F. K. (2019). Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*, *138*, 322-329.

Brooks, B., Chan, S., Lander, P., Adamson, R., Hodgetts, G. A., & Deakin, C. D. (2015). Public knowledge and confidence in the use of public access defibrillation. *Heart (British Cardiac Society)*, *101*, 967-971.

Caputo, M. L., Muschietti, S., Burkart, R., Benvenuti, C., Conte, G., Regoli, F., … Auricchio, A. (2017). Lay persons alerted by mobile application system initiate earlier cardio-pulmonary resuscitation: A comparison with SMS-based system notification. *Resuscitation*, *114*, 73-78.

Chamberlain, D. (2004). Never quite there: A tale of resuscitation medicine. *Resuscitation*, *60*, 3-11.

Chan, J., Rea, T., Gollakota, S., Sunshine J. E. (2019). Contactless cardiac arrest detection using smart devices. *NPJ Digit Med, 2*, 52.

Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine — beyond the peak of inflated expectations. *New England Journal of Medicine*, *376*, 2507-2509.

Chu, J., Leung, K. H. B., Snobelen, P., Nevils, G., Drennan I. R., Cheskes, S., & Chan, T. C. Y. (2021) Machine learning-based dispatch of drone-delivered defibrillators for out-of-hospital cardiac arrest. *Resuscitation, 162*, 120-127.

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. (2016). Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, *44*, 368-374.

Collins, G. S., & Moons, K. G. M. (2019). Reporting of artificial intelligence prediction models. *The Lancet*, *393*, 1577-1579.

Deakin, C. D. (2018). The chain of survival: Not all links are equal. *Resuscitation*, *126*, 80-82.

Downey, C. L., Tahir, W., Randell, R., Brown, J. M., & Jayne, D. G. (2017). Strengths and limitations of early warning scores: A systematic review and narrative synthesis. *International Journal of Nursing Studies*, *76*, 106-119.

Dungan, K. M., Braithwaite, S. S., & Preiser, J. C. (2009). Stress hyperglycaemia. *The Lancet*, *373*, 1798-1807.

Eckart, A., Hauser, S. I., Kutz, A., Haubitz, S., Hausfater, P., Amin, D., … Schuetz, P. (2019). Combination of the National Early Warning Score (NEWS) and inflammatory biomarkers for early risk stratification in emergency department patients: Results of a multinational, observational study. *BMJ Open*, *9*, 1-11.

Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.

Endo, T., Endo, T., Yoshida, T., Shinozaki, T., Motohashi, T., Hsu, H. C., … Fujitani, S. (2020). Efficacy of prehospital National Early Warning Score to predict outpatient disposition at an emergency department of a Japanese tertiary hospital: A retrospective study. *BMJ Open*, *10*, 1-7.

Escobar, G. J., Liu, V. X., Schuler, A., Lawson, B., Greene, J. D., & Kipnis, P. (2020). Automated identification of adults at risk for in-hospital clinical deterioration. *New England Journal of Medicine*, *383*, 1951-1960.

Finnish Heart Association. DEFI.fi, Rekisteri sydäniskureista. (Accessed 23 March 2021, at.

https://defi.fi/).

Fischer, M., Kamp, J., Garcia-Castrillo Riesgo, L., Robertson-Steel, I., Overton, J., Ziemann, A., & Krafft, T. (2011). Comparing emergency medical service systems-A project of the European Emergency Data (EED) Project. *Resuscitation*, *82*, 285-293.

Fukushima, H., Imanishi, M., Iwami, T., Seki, T., Kawai, Y., Norimoto, K., … Okuchi, K. (2015). Abnormal breathing of sudden cardiac arrest victims described by laypersons and its association with emergency medical service dispatcher-assisted cardiopulmonary resuscitation instruction. *Emergency Medicine Journal*, *32*, 314-317.

Giannini, H. M., Ginestra, J. C., Chivers, C., Draugelis, M., Hanish, A., Schweickert, W. D., … Umscheid, C. A. (2019). A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Critical Care Medicine*, *47*, 1485-1492.

Gräsner, J. T., Mancini, M. B., Avis, S., Considine, J., Perkins, G. D., Kudenchuck, P., … Olasveengen, T. M. - on behalf of the International Liaison Committee on Resuscitation Basic Life Support Task Force. (2020) CPR prior to defibrillation Consensus on Science with Treatment Recommendations. [Internet] Brussels, Belgium: International Liaison Committee on Resuscitation (ILCOR) Advanced Life Support Task Force, 2020, Jan.1. (Accessed 28 February 2021, at. https://costr.ilcor.org/document/cpr-prior-to-defibrillation-tfsr-costr).

Gräsner, J. T., Wnent, J., Herlitz, J., Perkins, G. D., Lefering, R., Tjelmeland, I., … Bossaert, L. (2020). Survival after out-of-hospital cardiac arrest in Europe - Results of the EuReCa TWO study. *Resuscitation*, *148*, 218-226.

Green, M., Lander, H., Snyder, A., Hudson, P., Churpek, M., & Edelson, D. (2018). Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. *Resuscitation*, *123*, 86-91.

Hallstrom, A. P., Ornato, J. P., Weisfeldt, M., Travers, A., Christenson, J., McBurnie M. A., … Proschan, M. (2004). Public-Access Defibrillation and survival after out-of-hospital cardiac arrest. *New England Journal of Medicine*, *35*, 637-646.

Hansen, S. M., Brøndum, S., Thomas, G., Rasmussen, S. R., Kvist, B., Christensen, A., … Hansen, P. A. (2015). Home care providers to the rescue: A novel first-responder programme. *PLoS ONE*, *10*, 1-10.

Hasselqvist-Ax, I., Riva, G., Herlitz, J., Rosenqvist, M., Hollenberg, J., Nordberg, P., … Svensson, L. (2015). Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. *New England Journal of Medicine*, *372*, 2307-2315.

Hillman, K. M., Bristow, P. J., Chey, T., Daffurn, K., Jacques, T., Norman, S. L., … Simmons, G. (2002). Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Medicine*, *28*, 1629-1634.

Hiltunen, P., Kuisma, M., Silfvast, T., Rutanen, J., Vaahersalo, J., & Kurola, J. (2012). Regional variation and outcome of out-of-hospital cardiac arrest (ohca) in Finland -- the Finnresusci study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *20*, 80.

Hiltunen, P., Silfvast, T. O., Jäntti, T. H., Kuisma, M. J., Kurola, J. O., Helena Jantti, T., … Kurola, J. O. (2015). Emergency dispatch process and patient outcome in bystander-witnessed out-of-hospital cardiac arrest with a shockable rhythm. *European Journal of Emergency Medicine* , *22*, 266-272.

Ho, T. K. (1995). Random Decision Forests. *ICDAR '95: Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278-282. IEEE Computer Society Press. doi:

10.1109/ICDAR.1995.598994
Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE*, *13*, 1-13.
Høyer, C. B., & Christensen, E. F. (2009). Fire fighters as basic life support responders: a study of successful implementation. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *17*, 16.
Hurt, R. (2005). Modern cardiopulmonary resuscitation - Not so new after all. *Journal of the Royal Society of Medicine*, *98*, 327-331.
Institute for the Languages in Finland. Suomi. (Accessed 28 February 2021, at. https://www.kotus.fi/kielitieto/kielet/suomi).
Jo, S., Yoon, J., Lee, J. B., Jin, Y., Jeong, T., & Park, B. (2016). Predictive value of the National Early Warning Score-Lactate for mortality and the need for critical care among general emergency department patients. *Journal of Critical Care*, *36*, 60-68.
Karlsson, L., Malta Hansen, C., Wissenberg, M., Møller Hansen, S., Lippert, F. K., Rajan, S., … Folke, F. (2019). Automated external defibrillator accessibility is crucial for bystander defibrillation and survival: A registry-based study. *Resuscitation*, *136*, 30-37.
Kitamura, T., Kiyohara, K., Sakai, T., Matsuyama, T., Hatakeyama, T., Shimamoto, T., … Iwami, T. (2016). Public-access defibrillation and out-of-hospital cardiac arrest in Japan. *New England Journal of Medicine*, *375*, 1649-1659.
Kivipuro, M., Tirkkonen, J., Kontula, T., Solin, J., Kalliomäki, J., Pauniaho, S. L., … Hoppu, S. (2018). National early warning score (NEWS) in a Finnish multidisciplinary emergency department and direct vs. late admission to intensive care. *Resuscitation, 128*, 164-169.
Kleinman, M. E., Perkins, G. D., Bhanji, F., Billi, J. E., Bray, J. E., Callaway, C. W., … Zideman, D. (2018). ILCOR Scientific knowledge gaps and clinical research priorities for cardiopulmonary resuscitation and emergency cardiovascular care: a consensus statement. *Circulation*, *137*, e802-e819.
Knezevic, A. (2008). Overlapping confidence intervals and statistical significance. *Stat News*. (Accessed 28 February 2021, at. https://www.cscu.cornell.edu/news/statnews/73_ci.pdf)
Kouwenhoven W. B., Jude J. R., & Knickerbocker, G. G. (1960). Closed-chest cardiac massage. *Jama*, *173*, 1064-1067.
Kragholm, K., Wissenberg, M., Mortensen, R. N., Hansen, S. M., Malta Hansen, C., Thorsteinsson, K., … Rasmussen, B. S. (2017). Bystander efforts and 1-year outcomes in out-of-hospital cardiac arrest. *New England Journal of Medicine*, *376*, 1737-1747.
Krippendorf, K. (2004) Content analysis: an introduction to its methodology. 2nd ed. *Thousand Oaks, California: Sage Publications*.
Kuisma, M., Boyd, J., Väyrynen, T., Repo, J., Nousila-Wiik, M., & Holmström, P. (2005). Emergency call processing and survival from out-of-hospital ventricular fibrillation. *Resuscitation*, *67*, 89-93.
Lee, S. Y., Shin, S. Do, Lee, Y. J., Song, K. J., Hong, K. J., Ro, Y. S., … Kong, S. Y. (2019). Text message alert system and resuscitation outcomes after out-of-hospital cardiac arrest: A before-and-after population-based study. *Resuscitation*, *138*, 198-207.
Lee, S. Y., Song, K., J., Shin, S. D., Hong, K. J., Kim, T., H. (2020). Comparison of the effects of audio-instructed and video-instructed dispatcher-assisted cardiopulmonary resuscitation on resuscitation outcomes after out-of-hospital cardiac arrest, *Resuscitation, 147*, 12-20.
Lewis, M., Stubbs, B. A., & Eisenberg, M. S. (2013). Dispatcher-assisted cardiopulmonary resuscitation: Time to identify cardiac arrest and deliver chest compression instructions. *Circulation*, *128*, 1522-1530.

Lin, K., Hu, Y., & Kong, G. (2019). Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *International Journal of Medical Informatics*, *125*, 55-61.

Linnen, D. T., Escobar, G. J., Hu, X., Scruth, E., Liu, V., & Stephens, C. (2019). Statistical modeling and aggregate-weighted scoring systems in prediction of mortality and ICU transfer: a systematic review. *Journal of Hospital Medicine*, *14*, 161-169.

Mäkelä, K. (1990) Kvalitatiivisen aineiston analyysi ja tulkinta [Analysis and Interpretation of the Qualitative Data]. *Helsinki: Gaudeamus.*

Malta Hansen, C., Kragholm, K., Pearson, D. a, Tyson, C., Monk, L., Myers, B., … Granger, C. B. (2015). Association of bystander and first-responder intervention with survival after out-of-hospital cardiac arrest in North Carolina, 2010-2013. *Jama*, *314*, 255-264.

Malta Hansen, C., Lippert, F. K., Wissenberg, M., Weeke, P., Zinckernagel, L., Ruwald, M. H., … Folke, F. (2014). Temporal trends in coverage of historical cardiac arrests using a volunteer-based network of automated external defibrillators accessible to laypersons and emergency dispatch centers. *Circulation*, *130*, 1859-1867.

Malta Hansen, C., Wissenberg, M., Weeke, P., Ruwald, M. H., Lamberts, M., Lippert, F. K., … Folke, F. (2013). Automated external defibrillators inaccessible to more than half of nearby cardiac arrests in public locations during evening, nighttime, and weekends. *Circulation*, *128*, 2224-2231.

Myat, A., Song, K. J., & Rea, T. (2018). Out-of-hospital cardiac arrest: current concepts. *The Lancet*, *391*, 970-979.

Nannan Panday, R. S., Minderhoud, T. C., Alam, N., & Nanayakkara, P. W. B. (2017). Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): A narrative review. *European Journal of Internal Medicine*, *45,* 20-31.

Nehme, Z., Andrew, E., Bernard, S., Haskins, B., & Smith, K. (2019). Trends in survival from out-of-hospital cardiac arrests defibrillated by paramedics, first responders and bystanders. *Resuscitation*, *143*, 85-91.

Nichol, G., Cobb, L. A., Yin, L., Maynard, C., Olsufka, M., Larsen, J., … Sayre, M. R. (2016). Briefer activation time is associated with better outcomes after out-of-hospital cardiac arrest. *Resuscitation*, *107*, 139-144.

Nickel, C. H., Kellett, J., Cooksley, T., Bingisser, R., Henriksen, D. P., & Brabrand, M. (2016). Combined use of the National Early Warning Score and D-dimer levels to predict 30-day and 365-day mortality in medical patients. *Resuscitation*, *106*, 49-52.

Nikolaou, N., Dainty, K. N., Couper, K., Morley, P., Tijssen, J., Vaillancourt, C., … Voorde, P. Van de. (2019). A systematic review and meta-analysis of the effect of dispatcher-assisted CPR on outcomes from sudden cardiac arrest in adults and children. *Resuscitation*, *138*, 82-105.

Nordberg, P., Hollenberg, J., Rosenqvist, M., Herlitz, J., Jonsson, M., Järnbert-Petterson, H., … Svensson, L. (2014). The implementation of a dual dispatch system in out-of-hospital cardiac arrest is associated with improved short and long term survival. *European Heart Journal: Acute Cardiovascular Care*, *3*, 293-303.

Nurmi, J., Pettilä, V., Biber, B., Kuisma, M., Komulainen, R., & Castrén, M. (2006). Effect of protocol compliance to cardiac arrest identification by emergency medical dispatchers. *Resuscitation*, *70*, 463-469.

Obuchowski, N. A. (1997). Nonparametric analysis of clustered ROC curve data. *Biometrics*, *53*, 567-578.

Olasveengen, T. M., de Caen, A. R., Mancini, M. E., Maconochie, I. K., Aickin, R., Atkins, D. L.,

… Nolan, J. P. (2017). 2017 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations: Summary. *Resuscitation*, *121*, 201-214.

Olasveengen, T. M., Mancini, M. E., Perkins, G. D., Avis, S., Brooks, S., Castrén, M. (2020). Adult Basic Life Support: International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. *Resuscitation*, *156*, A35-A79.

Olasveengen, T. M., Semeraro, F., Ristagno, G., Castrén, M., Handley, A. J., Kuzovlev, A., … Perkins, G. D. (2021). European Resuscitation Council Guidelines 2021: Basic Life Support. *Resuscitation*, *161*, 98-114.

Olsson, T., & Lind, L. (2003). Comparison of the rapid emergency medicine score and APACHE II in nonsurgical emergency department patients. *Academic Emergency Medicine*, *10*, 1040-1048.

Ong, M. E. H., Perkins, G. D., & Cariou, A. (2018). Out-of-hospital cardiac arrest: prehospital management. *The Lancet*, *391*, 980-988.

Page, R. L., Joglar, J. A., Kowal, R. C., Zagrodzky, J. D., Nelson, L. L., Ramaswamy, K., … McKenas, D. K. (2000). Use of automated external defibrillators by a U.S. airline. *New England Journal of Medicine*, *343*, 1210-1216.

Patel, R., Nugawela, M. D., Edwards, H. B., Richards, A., Le Roux, H., Pullyblank, A., & Whiting, P. (2018). Can early warning scores identify deteriorating patients in pre-hospital settings? A systematic review. *Resuscitation*, *132*, 101-111.

Perkins, G. D., Jacobs, I. G., Nadkarni, V. M., Berg, R. A., Bhanji, F., Biarent, D., … Zideman, D. A. (2015). Cardiac arrest and cardiopulmonary resuscitation outcome reports: Update of the Utstein resuscitation registry templates for out-of-hospital cardiac arrest: A statement for healthcare professionals from a task force of the international liaison committee . *Circulation*, *132*, 1286-1300.

Perkins, G. D., Neumar, R., Monsieurs, K. G., Lim, S. H., Castren, M., Nolan, J. P., … Bossaert, L. (2017). The International Liaison Committee on Resuscitation—Review of the last 25 years and vision for the future. *Resuscitation*, *121*, 104-116.

Pijls, R. W. M., Nelemans, P. J., Rahel, B. M., & Gorgels, A. P. M. (2016). A text message alert system for trained volunteers improves out-of-hospital cardiac arrest survival. *Resuscitation*, *105*, 182-187.

Pirneskoski, J., Kuisma, M., Olkkola, K. T., & Nurmi, J. (2019). Prehospital National Early Warning Score predicts early mortality. *Acta Anaesthesiologica Scandinavica*, *63*, 1-8.

Puolakka, T., Väyrynen, T., Erkkilä, E. P., Kuisma, M. (2016). Fire Engine Support and On-scene Time in Prehospital Stroke Care - A Prospective Observational Study. *Prehospital and Disaster Medicine, 31,* 278-81.

Raileanu, L. E., & Stoffel, K. (2004) Theoretical Comparison between the Gini Index and Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence 41*, 77-93.

Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., Camargo, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*, *23*, 1-13.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*, 1347-1358.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. *Encyclopedia of Database Systems*. New York: Springer, 532-538.

Ringh, M., Jonsson, M., Nordberg, P., Fredman, D., Hasselqvist-Ax, I., Håkansson, F., …

Hollenberg, J. (2015). Survival after public access defibrillation in Stockholm, Sweden - A striking success. *Resuscitation*, *91*, 1-7.
Ringh, M., Rosenqvist, M., Hollenberg, J., Jonsson, M., Fredman, D., Nordberg, P., … Svensson, L. (2015). Mobile-phone dispatch of laypersons for CPR in out-of-hospital cardiac arrest. *The New England Journal of Medicine*, *372*, 2316-2325.
Riou, M., Ball, S., Whiteside, A., Bray, J., Perkins, G. D., Smith, K., … Finn, J. (2018). 'We're going to do CPR': A linguistic study of the words used to initiate dispatcher-assisted CPR and their association with caller agreement. *Resuscitation*, *133*, 95-100.
Riou, M., Ball, S., Williams, T. A., Whiteside, A., Cameron, P., Fatovich, D. M., … Finn, J. (2018). 'She's sort of breathing': What linguistic factors determine call-taker recognition of agonal breathing in emergency calls for cardiac arrest? *Resuscitation*, *122*, 92-98.
Riou, M., Ball, S., Williams, T. A., Whiteside, A., O'Halloran, K. L., Bray, J., … Finn, J. (2017). 'Tell me exactly what's happened': When linguistic choices affect the efficiency of emergency calls for cardiac arrest. *Resuscitation*, *117*, 58-65.
Roberts, A., Nimegeer, A., Farmer, J., & Heaney D. J. (2014). The experience of community first responders in co-producing rural health care: in the liminal gap between citizen and professional. *BMC Health Services Research, 14,* 460.
Rørtveit, S., & Meland, E. (2010). First responder resuscitation teams in a rural Norwegian community: sustainability and self-reports of meaningfulness, stress and mastering. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 18,* 25.
Royal College of Physicians. (2012). National Early Warning Score (NEWS) - Standardising the assessment of acute-illness severity in the NHS. Report of a working party. *London: Royal College of Physicians.*
Royal College of Physicians. (2017). National Early Warning Score (NEWS) 2. Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. Executive summary and recommendations. *London: Royal College of Physicians.*
Saner, H., Morger, C., Eser, P., & von Planta, M. (2013). Dual dispatch early defibrillation in out-of-hospital cardiac arrest in a mixed urban-rural population. *Resuscitation*, *84*, 1197-1202.
Sankala, S. (2019). Etnografinen tutkimus hätäkeskuspäivystäjän työstä ja arjesta.
Sasson, C., Rogers, M. A. M., Dahl, J., & Kellermann, A. L. (2010). Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis. *Circ Cardiovasc Qual Outcomes*, *3*, 63-81.
Schein, R. M. H., Hazday, N., Pena, M., Ruben, B. H., & Sprung, C. L. (1990). Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest*, *98*, 1388-1392.
Semeraro, F., Greif, R., Böttiger, B. W., Burkart, R., Cimpoesu D., Georgiou M., … Monsieurs K. G. (2021). European Resuscitation Council Guidelines 2021: Systems Saving Lives. *Resuscitation*, *161*, 80-97.
Setälä, P., Hoppu, S., Virkkunen, I., Yli-Hankala, A., & Kämäräinen, A. (2017). Assessment of futility in out-of-hospital cardiac arrest. *Acta Anaesthesiologica Scandinavica*, *61*, 1334-1344.
Silcock, D. J., Corfield, A. R., Gowens, P. A., & Rooney, K. D. (2015). Validation of the National Early Warning Score in the prehospital setting. *Resuscitation*, *89*, 31-35.
Singh, S., McGlennan, A., England, A., & Simons, R. (2012). A validation study of the CEMACH recommended modified early obstetric warning system (MEOWS). *Anaesthesia*, *67*, 12-18.
Smith, G. B. (2010). In-hospital cardiac arrest: Is it time for an in-hospital "chain of prevention"? *Resuscitation*, *81*, 1209-1211.
Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E., & Featherstone, P. I. (2013). The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early

cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*, *84*, 465-470.

Spangler, D., Hermansson, T., Smekal, D., & Blomberg, H. (2019). A validation of machine learning-based risk scores in the prehospital setting. *PLoS ONE*, *14*, 1-18.

Statistics Finland. Official statistics of Finland (OSF): Population structure. (Accessed 7 January 2020, at. https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto_en.html).

Stieglis, R., Zijlstra, J. A., Riedijk, F., Smeekes, M., van der Worp, W. E., & Koster, R. W. (2020). AED and text message responders density in residential areas for rapid response in out-of-hospital cardiac arrest. *Resuscitation*, *150*, 170-177.

Stroop, R., Kerner, T., Strickmann, B., & Hensel, M. (2020). Mobile phone-based alerting of CPR-trained volunteers simultaneously with the ambulance can reduce the resuscitation-free interval and improve outcome after out-of-hospital cardiac arrest: A German, population-based cohort study. *Resuscitation*, *147*, 57-64.

Syväoja, S., Salo, A., Uusaro, A., Jäntti, H., & Kuisma, M. (2018). Witnessed out-of-hospital cardiac arrest- effects of emergency dispatch recognition. *Acta Anaesthesiologica Scandinavica*, *62*, 558-567.

Tirkkonen J., Olkkola K. T., Huhtala H., Tenhunen J., & Hoppu S. (2014). Medical emergency team activation: performance of conventional dichotomised criteria versus national early warning score. *Acta Anaesthesiol Scand, 58*, 411-419.

Tirkkonen, J., Skrifvars, M. B., Tamminen, t., Parr, M. J. A., Hillman, K., Efendijev, I., & Aneman, A. (2020). Afferent limb failure revisited - A retrospective, international, multicentre, cohort study of delayed rapid response team calls. *Resuscitation, 156*, 6-14.

Valenzuela, T. D., Roe, D. J., Cretin, S., Spaite, D. W., & Larsen, M. P. (1997). Estimating effectiveness of cardiac arrest interventions. *Circulation*, *96*, 3308-3313.

Valenzuela, T. D., Roe, D. J., Nichol, G., Clark, L. L., Spaite, D. W., & Hardman, R. G. (2000). Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. *New England Journal of Medicine*, *343*, 1206-1209.

van Alem, A. P., Vrenken, R. H., de Vos, R., Tijssen, J. G., & Koster, R. W. (2003). Use of automated external defibrillator by first responders in out of hospital cardiac arrest: prospective controlled trial. *BMJ*, *327*, 1312.

Viereck, S., Møller, T. P., Rothman, J. P., Folke, F., & Lippert, F. K. (2017). Recognition of out-of-hospital cardiac arrest during emergency calls — a systematic review of observational studies, *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 25*, 9.

Vihonen, H., Lääperi, M., Kuisma, M., Pirneskoski, J., & Nurmi, J. (2020). Glucose as an additional parameter to National Early Warning Score (NEWS) in prehospital setting enhances identification of patients at risk of death: An observational cohort study. *Emergency Medicine Journal*, *37*, 286-292.

Weisfeldt, M. L., & Becker, L. B. (2002). Resuscitation after cardiac arrest: A 3-phase time-sensitive model. *Journal of the American Medical Association, 288,* 3035-3038.

Weisfeldt, M. L., Sitlani, C. M., Ornato, J. P., Rea, T., Aufderheide, T. P., Davis, D., … Morrison, L. J. (2010). Survival after application of automatic external defibrillators before arrival of the emergency medical system. Evaluation in the Resuscitation Outcomes Consortium population of 21 million. *Journal of the American College of Cardiology*, *55*, 1713-1720.

Williams, T. A., Tohira, H., Finn, J., Perkins, G. D., & Ho, K. M. (2016). The ability of early warning scores (EWS) to detect critical illness in the prehospital setting: A systematic review. *Resuscitation*, *102*, 35-43.

Wissenberg, M., Lippert, F. K., Folke, F., Weeke, P., Hansen, C. M., Christensen, E. F., … Torp-

Pedersen, C. (2013). Association of national initiatives to improve cardiac arrest management with rates of bystander intervention and patient survival after out-of-hospital cardiac arrest. *Journal of the American Medical Association*, *310*, 1377-1384.

Yeung, J., Okamoto, D., Soar, J., & Perkins, G. D. (2011). AED training and its impact on skill acquisition, retention and performance - A systematic review of alternative training methods. *Resuscitation*, *82*, 657-664.

Yu, Y., Meng, Q., Munot, S., Nguyen, T. N., Redfern, J., & Chow, C. K. (2020) Assessment of community interventions for bystander cardiopulmonary resuscitation in out-of-hospital cardiac arrest: a systematic review and meta-analysis. *JAMA Netw Open, 3*, e209256.

Zijlstra, J. A., Koster, R. W., Blom, M. T., Lippert, F. K., Svensson, L., Herlitz, J., … Hollenberg, J. (2018). Different defibrillation strategies in survivors after out-of-hospital cardiac arrest. *Heart*, *104*, 1929-1936.

Zijlstra, J. A., Stieglis, R., Riedijk, F., Smeekes, M., van der Worp, W. E., & Koster, R. W. (2014). Local lay rescuers with AEDs, alerted by text messages, contribute to early defibrillation in a Dutch out-of-hospital cardiac arrest dispatch system. *Resuscitation*, *85*, 1444-1449.

PUBLICATIONS

PUBLICATION
I

Spontaneous trigger words associated with confirmed out-of-hospital cardiac arrest: a descriptive pilot study of emergency calls

Joonas Tamminen, Erik Lydén, Jan Kurki, Heini Huhtala, Antti Kämäräinen, Sanna Hoppu

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 2020;28:1
DOI: 10.1186/s13049-019-0696-1

Tamminen et al. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine (2020) 28:1
https://doi.org/10.1186/s13049-019-0696-1

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine

ORIGINAL RESEARCH

Open Access

Spontaneous trigger words associated with confirmed out-of-hospital cardiac arrest: a descriptive pilot study of emergency calls

Joonas Tamminen[1,2*], Erik Lydén[2], Jan Kurki[2], Heini Huhtala[3], Antti Kämäräinen[2] and Sanna Hoppu[2]

Abstract

Background: According to the International Liaison Committee on Resuscitation (ILCOR), the trigger words used by callers that are associated with cardiac arrest constitute a scientific knowledge gap. This study was designed to find hypothetical trigger words in emergency calls in order to improve the specificity of out-of-hospital cardiac arrest recognition.

Methods: In this descriptive pilot study conducted in a Finnish hospital district, linguistic contents of 80 emergency calls of dispatcher-suspected or EMS-encountered out-of-hospital cardiac arrests between January 1, 2017 and May 31, 2017 were analysed. Spontaneous trigger words used by callers were transcribed and grouped into 36 categories. The association between the spontaneous trigger words and confirmed true cardiac arrests was tested with logistic regression.

Results: Of the suspected cardiac arrests, 51 (64%) were confirmed as true cardiac arrests when ambulance personnel met the patient. A total of 291 spontaneous trigger words were analysed. 'Is not breathing' ($n = 9$ [18%] in the true cardiac arrest group vs $n = 1$ [3%] in the non-cardiac arrest group, odds ratio [OR] 6.00, 95% confidence interval [CI] 0.72–50.0), 'the patient is blue' ($n = 9$ [18%] vs $n = 1$ [3%], OR 6.00, 95% CI 0.72–50.0), 'collapsed or fallen down' ($n = 12$ [24%] vs $n = 2$ [7%], OR 4.15, 95% CI 0.86–20.1) and 'is wheezing' ($n = 17$ [33%] vs $n = 5$ [17%], OR 2.40, 95% CI 0.78–7.40) were frequently used to describe true cardiac arrest. 'Is snoring' was associated with a false suspicion of cardiac arrest ($n = 1$ [2%] vs $n = 6$ [21%], OR 0.08, 95% CI 0.009–0.67).

Conclusions: In our pilot study, no trigger word was associated with confirmed cardiac arrest. 'Is wheezing' was a frequently used spontaneous trigger word among later confirmed cardiac arrest victims.

Keywords: Out-of-hospital cardiac arrest, Bystander cardiopulmonary resuscitation, Dispatch, Emergency calls, Trigger words

Background

Survival after out-of-hospital cardiac arrest (OHCA) remains modest despite standardised dispatch protocols in emergency medical services (EMS) systems, increased community training and the introduction of post-resuscitation care [1–3]. Nevertheless, early pre-hospital interventions do have a substantial impact on the survival of OHCA victims. Bystander-initiated cardiopulmonary resuscitation (CPR) increases the chances of 30-day survival twofold and is associated with improved long-term neurological outcome [4, 5].

Early recognition of cardiac arrest is the cornerstone of the chain of survival [6–8]. The well-known clinical signs of cardiac arrest are unresponsiveness and absent or abnormal breathing [6]. However, it is unclear how these signs and symptoms, especially agonal breaths, are interpreted and described by laypeople. Besides cardiac arrest, these clinical signs and symptoms are also related to many other medical conditions, which results in significant amount of false positive suspicions of OHCA. Emergency calls could contain hypothetical trigger words

* Correspondence: joonas.i.tamminen@tuni.fi
[1]Faculty of Medicine and Health Technology, Tampere University, PO Box 2000, FI-33520 Tampere, Finland
[2]Emergency Medical Services, Tampere University Hospital, PO Box 2000, FI-33521 Tampere, Finland
Full list of author information is available at the end of the article

that current dispatch protocol may not recognise; the International Liaison Committee on Resuscitation (ILCOR) has announced that trigger words form a scientific knowledge gap [9]. These trigger words could be used to facilitate recognition of OHCA, to reduce time to dispatch EMS and to increase immediate bystander CPR rates. Importantly, they could be used to reduce the number of false positive alarms and thus to improve the specificity of recognition of cardiac arrest.

To test whether hypothetical trigger words exist and to generate more specific hypotheses, our study was designed as a descriptive pilot study. This pilot study aims to examine the association between true OHCA confirmed by ambulance personnel and laypeople's spontaneous trigger words regarding physiological deterioration of a patient in the context of emergency-dispatcher-suspected or EMS-encountered OHCA.

Methods

This descriptive pilot study was conducted in the Pirkanmaa Hospital District, Finland, which serves the city of Tampere and a surrounding rural area covering a population of 510, 000 [10]. In the study area, emergency calls are processed by trained emergency dispatchers, majority of whom are not medical professionals. The length of the formal dispatcher education is 1.5 years in Finland [11]. The national call processing is protocol-based and computer-aided. Recognition of cardiac arrest is based on three questions: (1) Tell me exactly what happened, (2) Is she/he conscious? and (3) Is she/he breathing normally? [11] During the study period, the emergency dispatcher did not receive any additional feedback that differed from the standard quality control.

Between January 1, 2017 and May 31, 2017, all audio recordings and electronic mission reports of consecutive emergency calls of dispatcher-suspected OHCA or EMS-encountered OHCA that a dispatcher had not suspected in the study area were extracted from the EinsatzLeitSystem (ELS) database maintained by the Emergency Response Centre Agency [12]. As the aim of the study was to address laypeople's interpretations of physiological deterioration of an OHCA patient, cases with unwitnessed OHCA, traumatic cause for OHCA or an institutional resuscitation attempt were excluded.

As the study was retrospective and based on registry data only, with no interventions or patient contact involved, the need for patient consent was waived. The study protocol was approved by the institutional review board of the Pirkanmaa Health District (R17156, November 7th, 2017).

Spontaneous trigger words

Spontaneous speech, defined as something that the caller said without being prompted or asked by the dispatcher, was transcribed by authors EL and JK who are professional paramedics. Caller's whole answer to a preceding question was considered as non-spontaneous speech regardless of the duration or the length of the answer. In order to analyse transcribed speech, different words with the same semantic meaning were put in a single category [13, 14]. Authors JT, EL and JK interpreted the semantic meaning of trigger words and categorised them. The basis of our categorisation was a word list introduced by Berdowski et al. [7], which included seven categories: breathing, consciousness, facial colour, death, heart problems, resuscitation, and other. In addition, the ABCDE approach was used to formulate our categorisation [15]. The ABCDE is a mnemonic for a generally accepted treatment protocol for critically ill patients. In our study, the spontaneous trigger words were grouped into seven main categories and thirty-six subcategories, the former of which included altered level of unconsciousness, death, breathing, circulation, disability, history of present illness, and unclassified. Our circulation category included facial colour and heart problems as subcategories. In cases of an ambiguous trigger word, the other two authors verified the suggested trigger word category.

Each emergency call could fulfil the criteria of each subcategory once. Subsequently, two or more trigger words were counted as a duplicate if the caller repeated the same word or if the caller used words that had a different linguistic form but had an identical semantic meaning. Ultimately, the trigger words were translated from Finnish to English (United Kingdom) by two native Finland linguists who have MA degrees in communication sciences.

Confirmation of true cardiac arrest

The trigger words were stratified into true cardiac arrest and non-cardiac arrest groups. The mission reports were used to identify true cardiac arrests, as there was no national cardiac arrest registry in Finland. After each mission, the EMS personnel filled out specific documentation that contained dispatch and transportation codes (e.g. the patient is confirmed dead, or the patient had return of spontaneous circulation, or CPR was being performed during transportation or the patient had had any other medical emergency). The true cardiac arrest events were confirmed by the EMS personnel based on these documentations. A transportation code for a non-OHCA event could be, for example, rhythm disturbance or intoxication.

Statistical methods

SPSS software version 25 (SPSS Inc., Chicago, IL, USA) was used to perform the statistical calculations. Categorical and continuous variables were reported as frequencies and proportions and as medians and interquartile ranges, respectively. The comparison between the groups was performed using a χ^2 or a two-tailed Fisher's exact test for the categorical data and a Mann–Whitney U-test for the continuous, nonparametric data. A univariate logistic regression was used to assess the association between the spontaneous trigger words and confirmed cardiac arrests, and the results were presented as odds ratio (OR) with 95% confidence interval (CI). A two-sided p-value < 0.05 was considered statistically significant.

Results

During the study period, 112 emergency calls met our inclusion criteria. A total of 32 (29%) cases were excluded because they related to an institutional resuscitation, the patient was awake or other reasons (e.g. poor sound quality), and 80 (71%) emergency calls were transcribed as presented in Fig. 1. Of the suspected cardiac arrests, 51 (64%) were confirmed as true cardiac arrests, and 29 (36%) of the suspected cardiac arrests were regarded as non-cardiac arrest events when EMS evaluated the patient.

The emergency call and mission characteristics are presented in Table 1. Most cardiac arrests were suspected after an ambulance was dispatched, and two confirmed cardiac arrests were not recognised by the dispatcher. The time of OHCA suspicion, the number of trigger words and the duration of speech intervals were similar between the groups. A total of 291 spontaneous trigger words were analysed; 93 (32%) and 41 (14%) of them concerned breathing and altered level of consciousness, respectively. The distribution of spontaneous trigger words in confirmed cardiac arrest and non-cardiac groups is presented in Fig. 2.

The results of the univariate logistic regression are shown in Table 2. The spontaneous trigger words that were more frequently used to describe true cardiac arrest were 'is not breathing' ($n = 9$ [18%] in the true cardiac arrest group vs $n = 1$ [3%] in the non-cardiac arrest group, odds ratio [OR] 6.00, 95% confidence interval [CI] 0.72–50.0), 'the patient is blue' (n = 9 [18%] vs $n = 1$ [3%], OR 6.00, 95% CI 0.72–50.0), 'collapsed or fallen down' ($n = 12$ [24%] vs $n = 2$ [7%], OR 4.15, 95% CI 0.31–20.1) and 'is wheezing' ($n = 17$ [33%] vs $n = 5$ [17%], OR 2.40, 95% CI 0.78–7.40). 'Is snoring' was associated with a false suspicion of cardiac arrest (n = 1 [2%] vs $n = 6$ [21%], OR 0.08, 95% CI 0.009–0.67).

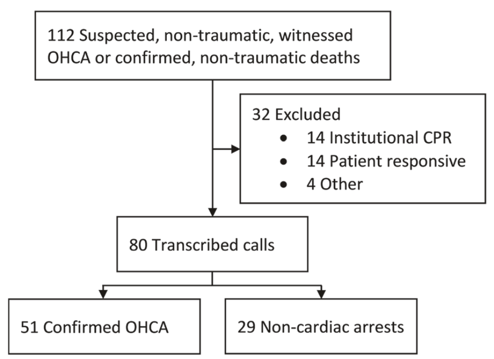

Fig. 1 Study population flowchart

Discussion

In this descriptive pilot study conducted in a Finnish hospital district, the linguistic contents of 80 emergency calls of suspected, non-traumatic, witnessed OHCAs or EMS-encountered, non-traumatic OHCAs that a dispatcher had not suspected were evaluated. The focus of the study was on spontaneous speech used by the caller since it was hypothesised to contain trigger words that the current dispatch protocol may have missed. If recognised, these trigger words could make dispatching faster and more specific. Although ILCOR notes that the trigger words associated with OHCA are a scientific knowledge gap, only one Dutch study has explored trigger words and a couple of Australian studies have examined the communication between emergency dispatchers and laypeople [7, 16, 17].

Our emergency dispatchers performed well during the five-month study period; the emergency dispatcher did not recognise two later confirmed cardiac arrests. The sensitivity for OHCA recognition was 96.2% in our material whereas a recently published systematic review concluded that the global sensitivity for OHCA recognition is 73.9% (range 14.1–96.9%) [18]. The review included three studies conducted in Finnish regions which found slightly lower sensitivities as compared with our results: 82.9, 82.3 and 79.4%, respectively [11, 19, 20]. As Viereck et al. argue, the definition of a recognised cardiac arrest is ambiguous and may result in conflicting estimates of the performance of a given EMS system.

According to the European Resuscitation Council (ERC) guidelines, recognition of OHCA is based on the combination of the patient being recognised as unconscious and apnoeic or breathing abnormally. One might argue that interpretation of the trigger words in relation to breathing is conditional on what is said about the conscious state and vice versa. However, we postulate that an individual trigger may combine the semantic information regarding both level of consciousness and breathing in the context of a medical emergency.

In our material, there were two important trigger words in the breathing category worth noting: 'is wheezing' (Finnish: *korisee*) and 'is snoring' (Finnish: *kuorsaa*). The former does not mean obstructive wheezing but rather a death rattle or choking sounds, and it seems to be an idiomatic expression in Finnish language. In addition, both trigger words mean that the patient has difficulties maintaining the normal muscle tone of the upper respiratory track, which, in turn, reflects a markedly altered level

Table 1 Emergency call and mission characteristics

	True cardiac arrest	Non-cardiac arrest	*p*-value
	n = 51	*n* = 29	
Trigger words			
Total, n (%)	194 (67)	97 (33)	
Median (IQR)	3 (2–5)	3 (2–5)	0.369
Time of OHCA suspicion, n = (%)			
Initial reason for dispatch	15 (29)	6 (21)	0.440
EMS en route	34 (67)	23 (79)	0.307
OHCA not suspected	2 (4)	0 (0)	0.532
Initial dispatch code non-specific	23 (45)	12 (41)	0.817
Duration, median (IQR); min:sec			
Emergency call	6:47 (5:12–8:35)	5:31 (3:41–8:49)	0.423
Total spontaneous speech	5:12 (3:31–6:44)	3:55 (3:05–7:08)	0.506
Initial description of situation	0:05 (0:04–0:10)	0:05 (0:04–0:09)	0.590

IQR interquartile range; *OHCA* out-of-hospital cardiac arrest; *EMS* emergency medical services

of consciousness. The latter trigger word was associated with a later confirmed non-cardiac arrest event, whereas the former was the most frequently used single trigger word in the confirmed true cardiac arrest stratum.

As discussed above, the emergency dispatcher had missed two cases, in which ambulance personnel encountered cardiac arrest. Interestingly, 'is wheezing' was the only spontaneous trigger word in the first missed case. The second case included the following trigger words: 'shallow breathing', 'I'm not sure if the patient is breathing' and 'glazed eyes'. These trigger words may

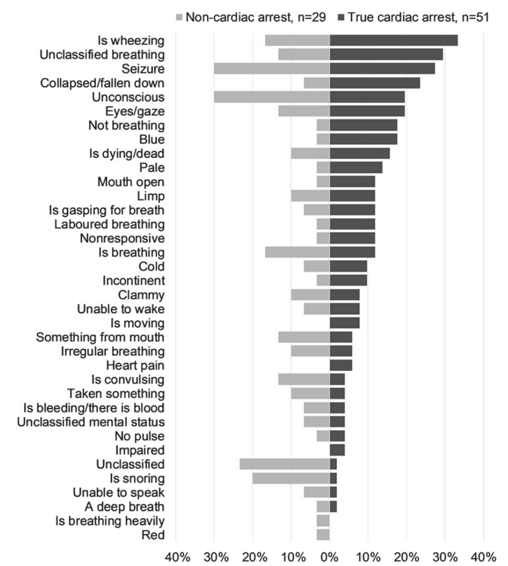

Fig. 2 The observed 291 spontaneous trigger words in 36 categories

reflect agonal breathing, which seems to be a pitfall of recognition of OHCA [21]. Indeed, subtle changes to the current algorithm may result in better sensitivity without a marked decrease in specificity. Riou et al. suggested that the emergency dispatcher should repeat the question regarding breathing pattern if the caller's initial answer is imprecise or vague [16].

In the future, trigger word combinations could be identified in real time by automatic speech recognition, and machine-learning models could calculate a probability of cardiac arrest. Corti AI, used by emergency dispatchers in Denmark, is an example of an automatic speech recognition program [22]. A recently published study evaluated this machine-learning algorithm to emergency dispatchers and showed that Corti AI seems to outperform emergency dispatchers for recognising OHCA [23].

Strengths and limitations

To the best of our knowledge, no previous study focusing on recognition of OHCA has explored spontaneous speech in emergency calls. Besides novelty, the strength of the study is the contribution of two native Finnish linguists, which increases the potential generalisability of the results beyond Finland.

This descriptive pilot study has several important limitations to consider. First, the study failed to detect any association between confirmed cardiac arrests and trigger words, and the confidence intervals for odds ratios were wide in the logistic regression model. However, this study was designed as a pilot study. A further study with a greater sample size is currently being conducted. Second, the study was underpowered to find trigger words associated with false negative cases (i.e. the dispatcher may have not suspected OHCA, even though a true

Table 2 Distribution (%) of the spontaneous trigger words and their association with confirmed cardiac arrests

	True cardiac arrest	Non-cardiac arrest	OR (95% CI)
Trigger words	$n = 51$	$n = 29$	
Dead or is dying	16	10	1.61 (0.39–6.63)
Altered level of consciousness			
Unconscious	20	31	0.54 (0.19–1.55)
Unable to wake	8	7	1.15 (0.20–6.69)
Nonresponsive	12	3	3.73 (0.43–32.7)
Impaired	4	0	NA
Unable to speak	2	7	0.27 (0.02–3.12)
Unclassified consciousness	4	7	0.55 (0.07–4.14)
Breathing			
Is breathing	12	17	0.42 (0.12–1.51)
Not breathing	18	3	6.00 (0.72–50.0)
Laboured	12	3	3.73 (0.43–32.7)
Heavily	0	3	NA
Irregularly	6	10	0.54 (0.10–2.88)
Is gasping for breath	12	7	1.80 (0.34–9.56)
A deep breath	2	3	0.56 (0.03–9.30)
Is snoring	2	21	0.08 (0.009–0.67)
Is wheezing	33	17	2.40 (0.78–7.40)
Unclassified breathing	29	14	2.60 (0.77–8.78)
Circulation			
No pulse	4	3	1.14 (0.10–13.2)
Pale	14	3	4.46 (0.52–38.2)
Red	0	1	NA
Blue	18	3	6.00 (0.72–50.0)
Cold	10	7	1.47 (0.27–8.09)
Clammy	8	10	0.74 (0.15–3.55)
Is bleeding or there is blood	4	7	0.55 (0.07–4.14)
Heart pain	6	0	NA
Disability			
Is convulsing	4	14	0.26 (0.04–1.49)
Limp	12	10	1.16 (0.27–5.01)
Incontinent	10	3	3.04 (0.34–27.4)
Eyes or gaze	20	14	1.52 (0.43–5.38)
Mouth open	12	3	3.73 (0.43–32.7)
Something from mouth	6	14	0.39 (0.08–1.88)
Is moving	8	0	NA
History of present illness			
Collapsed or fallen down	24	7	4.15 (0.86–20.1)
Seizure	27	31	0.84 (0.31–2.28)
Taken something	4	10	0.35 (0.06–2.25)
Unclassified	2	24	0.06 (0.007–0.54)

OR odds ratio; *CI* confidence interval; *NA* not applicable cardiac arrest had occurred). This was a rare event in our material, as the dispatcher had missed only two later confirmed OHCAs. Third, the authors were not blinded to the outcome when transcribing the emergency calls or categorising the trigger words. Fourth, the exact time of trigger words and the time of OHCA suspicion in an emergency call were not considered in our analysis. However, this study was not designed to address trigger words associated with prompt or late recognition of OHCA. Finally, transportation codes were used to confirm cardiac arrest. Nevertheless, it is extremely rare that EMS personnel would have used the transportation codes of OHCA for non-cardiac arrest missions and vice versa.

Conclusions

In conclusion, this pilot study introduces a novel method to categorise laypeople's spontaneous trigger words in emergency calls in the context of dispatcher-suspected cardiac arrest. No trigger word was associated with confirmed cardiac arrests, but 'is wheezing' was the most frequent trigger word in the confirmed cardiac arrest stratum.

Abbreviations

CI: Confidence interval; CPR: Cardiopulmonary resuscitation; ELS: EinsatzLeitSystem; EMS: Emergency medical services; ERC: European Resuscitation Council; ILCOR: International Liaison Committee on Resuscitation; OHCA: Out-of-hospital cardiac arrest; OR: Odds ratio

Acknowledgements

The authors thank linguists Satu Sihvonen, Kaisa Vihervaara and the Emergency Response Centre Agency for their contributions to the study.

Endnotes

Not applicable.

Authors' contributions

Authors EL and JK transcribed the audio recordings. JT, EL and JK interpreted the semantic meaning of trigger words and categorised them. All authors contributed substantially to the interpretation of the findings and approved the final version of this manuscript.

Funding

This study was investigator-driven, and no financial support was received.

Availability of data and materials

The data that support the findings of this study (including the complete list of Finnish trigger words with their English translations) are available from the corresponding author upon request.

Ethics approval and consent to participate

The study protocol was approved by the institutional review board of the Pirkanmaa Health District (R17156, November 7th, 2017). The study was retrospective and based on registry data only, with no interventions or patient contact involved. Thus, the need for patient consent was waived.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Faculty of Medicine and Health Technology, Tampere University, PO Box 2000, FI-33520 Tampere, Finland. [2]Emergency Medical Services, Tampere University Hospital, PO Box 2000, FI-33521 Tampere, Finland. [3]Biostatistics, Faculty of Social Sciences, Tampere University, FI-33014 Tampere, Finland.

Received: 12 August 2019 Accepted: 23 December 2019
Published online: 03 January 2020

References

1. Gräsner JT, Lefering R, Koster RW, Masterson S, Böttiger BW, Herlitz J, et al. EuReCa ONE—27 nations, ONE Europe, ONE registry: a prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. Resuscitation. 2016;105:188–95.
2. Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: systematic review of 67 prospective studies. Resuscitation. 2010;81:1479–87.
3. Kitamura T, Iwami T, Kawamura T, Nitta M, Nagao K, Nonogi H, et al. Nationwide improvements in survival from out-of-hospital cardiac arrest in Japan. Circulation. 2012;126:2834–43.
4. Hasselqvist-Ax I, Riva G, Herlitz J, Rosenqvist M, Hollenberg J, Nordberg P, et al. Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. N Engl J Med. 2015;372:2307–15.
5. Kragholm K, Wissenberg M, Mortensen RN, Hansen SM, Malta Hansen C, Thorsteinsson K, et al. Bystander efforts and 1-year outcomes in out-of-hospital cardiac arrest. N Engl J Med. 2017;376:1737–47.
6. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, et al. European resuscitation council guidelines for resuscitation 2015. Section 2Adult basic life support and automated external defibrillation. Resuscitation. 2015;95:81–99.
7. Berdowski J, Beekhuis F, Zwinderman AH, Tijssen JGP, Koster RW. Importance of the first link: description and recognition of an out-of-hospital cardiac arrest in an emergency call. 2009;119:2096–102.
8. Viereck S, Møller TP, Ersbøll AK, Bækgaard JS, Claesson A, Hollenberg J, et al. Recognising out-of-hospital cardiac arrest during emergency calls increases bystander cardiopulmonary resuscitation and survival. Resuscitation. 2017;14:141–7.
9. Olasveengen TM, de Caen AR, Mancini ME, Maconochie IK, Aickin R, Atkins DL, et al. 2017 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations summary. Resuscitation. 2017;121:201–14.
10. Official statistics of Finland (OSF): Population structure. 2017. http://www.stat.fi/til/vaerak/index_en.html. Accessed 2 July 2019.
11. Nurmi J, Pettilä V, Biber B, Kuisma M, Komulainen R, Castrén M. Effect of protocol compliance to cardiac arrest identification by emergency medical dispatchers. Resuscitation. 2006;70:463–9.
12. Pesonen J. Arvioinnin kehittäminen hätäkeskuspäivystäjien tutkinnossa. 2009. https://www.theseus.fi/bitstream/handle/10024/8102/Pesonen.Jouko.pdf?sequence=2. Accessed 2 July 2019.
13. Mäkelä, K. Kvalitatiivisen aineiston analyysi ja tulkinta [Analysis and Interpretation of the Qualitative Data]. Helsinki: Gaudeamus; 1990.
14. Krippendorf, K. Content analysis: an introduction to its methodology. 2nd ed. Thousand Oaks, California: Sage Publications; 2004.
15. The ABCE approach. 2019.https://www.resus.org.uk/resuscitation-guidelines/abcde-approach. Accessed 20 Oct 2019.
16. Riou M, Ball S, Williams TA, Whiteside A, Cameron P, Fatovich DM, et al. 'She's sort of breathing': what linguistic factors determine call-taker recognition of agonal breathing in emergency calls for cardiac arrest? Resuscitation. 2018;122:92–8.
17. Riou M, Ball S, Whiteside A, Bray J, Perkins GD, Smith K, et al. 'We're going to do CPR': a linguistic study of the words used to initiate dispatcher-assisted CPR and their association with caller agreement. Resuscitation. 2018;133:95–100.
18. Viereck S, Møller TP, Rothman JP, Folke F, Lippert FK. Recognition of out-of-hospital cardiac arrest during emergency calls - a systematic review of observational studies. Scand J Trauma Resusc Emerg Med. 2017;25:9.
19. Hiltunen PVC, TO S, Jäntti TH, Kuisma MJ, Kurola JO. Emergency dispatch process and patient outcome in bystander-witnessed out-of-hospital cardiac arrest with a shockable rhythm. Eur J Emerg Med. 2015;22:266–72.
20. Kuisma M, Boyd J, Väyrynen T, Repo J, Nousila-Wiik M, Holmström P. Emergency call processing and survival from out-of-hospital ventricular fibrillation. Resuscitation. 2005;67:89–93.

21. Fukushima H, Imanishi M, Iwami T, Seki T, Kawai Y, Norimoto K, et al. Abnormal breathing of sudden cardiac arrest victims described by laypersons and its association with emergency medical service dispatcher-assisted cardiopulmonary resuscitation instruction. Emerg Med J. 2014;1:314–7.
22. Artificial Intelligence that saves lives. 2019. https://corti.ai. Accessed 2 July 2019.
23. Blomberg SN, Folke F, Ersbøll AK, Christensen HC, Torp-Pedersen C, Sayre MR, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation. 2019;138:322–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

PUBLICATION
II

Professional firefighter and trained volunteer first-responding units in emergency medical service

Joonas Tamminen, Antti Kämäräinen, Sanna Hoppu

Acta Anaesthesiologica Scandinavica 2019;63:111–116
DOI: 10.1111/aas.13224

Received: 19 February 2018 | Revised: 5 June 2018 | Accepted: 18 June 2018

DOI: 10.1111/aas.13224

ORIGINAL ARTICLE

acta Anaesthesiologica Scandinavica

Professional firefighter and trained volunteer first-responding units in emergency medical service

Joonas I. Tamminen[1,2] | **Sanna E. Hoppu**[2] | **Antti J. J. Kämäräinen**[2]

[1]Medical School, University of Tampere, Tampere, Finland

[2]Emergency Medical Service, Tampere University Hospital, Tampere, Finland

Correspondence: Joonas I. Tamminen, Medical School, Tampere University Hospital, PO Box 2000, FI-33520 Tampere, Finland (tamminen.joonas.i@student.uta.fi).

Funding information
This study was investigator driven, and no financial support was received.

Background: Although widely dispatched to out-of-hospital cardiac arrests, the performance of prehospital first-responding units in other medical emergencies is unknown.

Methods: In this retrospective, descriptive study, the general performance of 44 first-responding units in Pirkanmaa County, Finland, were examined. A subgroup analysis compared the first-responding units made up of professional firefighters and trained volunteers.

Results: First-responding units were dispatched to patients during 1622 missions between 1 January 2013 and 31 December 2013. The median time to reach the scene was 9 minutes in any mission. Overall, first responders evaluated 1015 patients and provided treatment or assisted ambulance personnel in 793 (78%) cases. The most common treatment modalities were assistance, such as carrying (22%) and the administration of supplemental oxygen (19%). There were 83 resuscitation attempts during the time period. In 42 of these, first-responding units initiated basic life support a median of 4 minutes prior to the arrival of ambulance personnel. Return of spontaneous circulation was achieved in 20% of cases. The subgroup analysis showed that trained volunteers administered oxygen more liberally than professional firefighters in stroke and chest pain mission (stroke: professional 9/236 cases [4%] vs layperson 26/181 cases [14%], $P < 0.001$; chest pain: professional 16/78 cases [21%] vs layperson 77/159 cases [48%], $P < 0.001$).

Conclusion: First-responding units provided initial treatment or assistance to ambulance personnel in approximately half of the missions. Implementation of professional- and layperson-staffed first-responding units in emergency medical service system seems to be feasible.

KEYWORDS
emergency first responders, emergency medical services, firefighters, resuscitation, volunteers

1 | INTRODUCTION

First-responding units (FRUs) are used in many countries as a means of bringing trained help to the victims of out-of-hospital cardiac arrests (OHCAs) before ambulances.[1-7] There are variations on the concept of FRU organisation ranging from trained volunteer laypersons to professional firefighters and emergency medical technicians (EMTs) who are dispatched as one tier of emergency response.[1-7]

In addition to cardiac arrests, FRUs also respond to all other types of emergencies in Finland. Generally, FRU personnel perform initial, potentially lifesaving procedures prior to the arrival of the first ambulance. These procedures include, but are not limited to, cardiopulmonary resuscitation (CPR) with or without defibrillation; the opening of the airway with or without the use of simple airway methods, such as an oropharyngeal airway or a supraglottic device; and the control of external haemorrhage.[7]

To our knowledge, there are no previous reports on the general performance of FRUs. This study aims to evaluate the types of emergency medical service (EMS) missions FRUs complete and the procedures performed prior to ambulance arrival, as well as whether these procedures have any relief on the clinical state of the patient. Given that within the study area FRUs are staffed with both professional firefighters, EMTs, and trained volunteers depending on a given FRU's location, their contribution to emergency response was compared in terms of personnel composition.

Editorial Comment

First-responding prehospital units, besides ambulance personnel, can also utilise other rescue personnel or even trained laypersons from home. This report from Finland presents one organisation's experience with this type of prehospital response group.

2 | METHODS

The county of Pirkanmaa, Finland, (population circa 500 000) is covered by an EMS system coordinated by the Pirkanmaa Health District and consisting of 38 advanced life support (ALS)-level ambulances operated by both the Pirkanmaa Fire Services and several privately owned companies. In addition to the ALS-level ambulances, which are mainly used for immediate response, there is one extended ALS-level field commander unit, one physician-staffed helicopter emergency medical service unit and 44 FRUs. The FRUs are coordinated and trained by Pirkanmaa Fire Services. Fourteen of the FRUs operate from regional rescue stations and are staffed with professional firefighters, some of which work also as EMTs on the basic life support (BLS) level. These units respond to FRU dispatches within 90 seconds of the alarm. Twenty-seven of the FRUs are staffed with trained laypersons responding from home or work on a volunteer basis. By contract, these units respond to an emergency within 5 minutes of the associated dispatch. Three layperson-staffed units are available for immediate response during daytime, and during the night, these units will respond within 5 minutes of a dispatch. Approximately 400 civilians participate as first responders in the EMS system. Both professional and volunteer FRUs are allowed to use lights and sirens. All FRUs respond to fire and rescue dispatches, which are categorised as their primary missions in the case of simultaneous dispatches to both a rescue mission and a first-response EMS mission.

The professionally and layperson-staffed FRUs have the same treatment modalities and treatment protocol regardless of mission type. The principal means an FRU uses to help a victim of a prehospital emergency include the provision of CPR and automated external defibrillator (AED)-based early defibrillation, opening the airway using a supraglottic device or an oropharyngeal airway, supporting breathing with bag-mask ventilation and/or oxygen administration, wound dressing and the control of external haemorrhage, and the administration of rectal diazepam, subcutaneous glucagon, oral nitroglycerin or acetylsalicylic acid depending on the symptoms. There are several programs available for the initial FRU training of volunteer laypersons, mostly comprised of a BLS course and an additional 30-40 hour FRU course. The basic training of a professional firefighter is 1.5 years in duration, approximately one-third of which consists of emergency care. This training is provided via identical curricula at two colleges in Finland.

During the study protocol, an FRU was dispatched to an emergency by the Central Dispatch Centre when it was estimated to reach the patient 5 minutes prior to an ambulance in A-level emergencies (the most urgent, including sudden severe unconsciousness or presumed cardiac arrest) or 15 minutes prior to an ambulance in B-level emergencies (urgent mission, potential need for life support measures). In cases of witnessed cardiac arrest, high-energy fall trauma or presumed ischaemic stroke, the FRU was always dispatched, regardless of the expected time advantage over ambulance units. In cases of road traffic accidents, the units are dispatched per rescue service protocol and do not perform as FRUs for the EMS, thus excluding trauma cases due to motor vehicle accidents and fires from the study.

Of each mission, the FRU personnel filled out specific documentation. Based on this documentation, all FRU missions between 1 January 2013 and 31 December 2013 were analysed. The mission characteristics were analysed, focusing on the treatment provided by the FRUs and whether a clinical impact could be observed based on this treatment.

The primary endpoint was an improved or normalised vital function. A vital function was considered abnormal if systolic blood pressure <100 mm Hg, heart rate >150 or <40 beats per minute, respiration rate >30 or <10 breaths per minute, oxygen saturation ≤90%, Glasgow Coma Scale ≤13 or an impaired level of consciousness on the AVPU scale, hypoglycaemia (<4 mmol/L) or shortness of breath was recorded in the documentation. In addition, the primary endpoint also included the relief of pain. The data were further stratified into professionally and layperson-staffed unit groups to compare their contribution to emergency response, specifically focusing on the five symptoms termed the "first-hour quintet" (FHQ; cardiac arrest, severe respiratory failure, chest pain, severe trauma and stroke).[8]

Author JT manually transferred the data from the paper documentation to Microsoft Windows Excel. SPSS software version 23 (SPSS Inc, Chicago, IL, USA) was used to perform the statistical calculations. Continuous variables were reported as medians and their respective interquartile ranges and categorical variables were reported as frequencies and proportions. The comparison between the groups was performed using a Mann-Whitney *U*-test for the continuous, nonparametric data and a two-tailed Fisher's exact test for the categorical data. No systematic pattern regarding unreported patient and mission characteristics was observed, and no imputation method was applied to address missing data. Because the study was a retrospective chart review, no power calculation was performed, and the need for patient consent was waived. The study protocol was approved by the institutional review board of the Pirkanmaa Health District (R14148, 4.11.2014).

3 | RESULTS

During the 12-month study period, FRUs were dispatched on a total of 1894 medical first-response missions, yielding an incidence of 379 FRU missions per 100 000 citizens annually. Of these, the FRU mission was cancelled en route in 272 cases, and thus, FRUs attended to patients during 1622 (86%) missions (324/100 000/year). The study population is shown in Figure 1.

Patient and mission characteristics and time intervals are presented in Tables 1 and 2 respectively. The median response time from dispatch to scene was 9 minutes, and an FRU was the first unit on scene in 878 (54%) missions. An individual, professional FRU attended to a median of 44 patients per year (range 20-90), and a volunteer-staffed FRU attended to a median of 27 patients per year (range 2-66; *P* = 0.003).

Table 3 summarises the treatment characteristics and treatment responses. Overall, the FRUs evaluated 1015 out of the 1622 encountered patients and provided treatment or assisted ambulance personnel during 793 (49%) missions. CPR was attempted in 83 missions, and an FRU was the first to initiate CPR in 42 cases, which occurred at a median of 4 minutes prior to the arrival of the ambulance (range 1-18 minutes). Consequently, the return of spontaneous

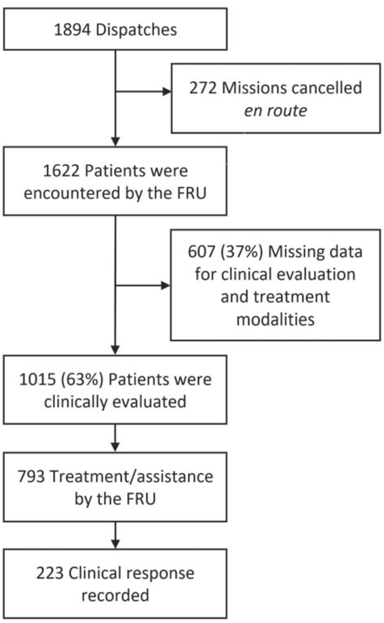

FIGURE 1 Study population. FRU, first-responding unit

TABLE 1 Mission characteristics[a]

Characteristics	Patient encounter n = 1622	%
Age, median (IQR); y	67 (52-81)	
Missing	527	32
Gender		
Male	779	48
Female	552	34
Missing	291	18
Mission type/reason for dispatch		
Ischaemic stroke	417	26
Chest pain	237	15
Trauma	132	8
Arrhythmia/collapse	130	8
Cardiac arrest	122	8
Confirmed cardiac arrest	114	7
Shortness of breath	114	7
Sudden unconsciousness	96	6
Other medical	345	21
Missing	29	2
First unit on scene		
FRU	860	53
BLS/ALS	363	22
Simultaneous arrival	82	5
Missing	96	20

IQR, interquartile range; FRU, first-responding unit; BLS/ALS, basic or advanced life support.
[a]272 (14%) of total 1894 missions were cancelled en route.

TABLE 2 Time intervals on missions when a patient was encountered

Patient encounter n = 1622	Median	IQR	Missing, %
Delay from dispatch, min			
To mobile	2	1-5	13
To scene	9	6-13	11
To patient	10	7-14	18
FRU before BLS/ALS on scene, min	9	5-13	20

IQR, interquartile range.

circulation (ROSC) was achieved in 17 (20%) of the missions during which CPR was performed.

Supplemental oxygen was administered to 309 patients, whose shortness of breath was improved in 126 cases (41%). Medication was provided to 64 patients, which resulted in the relief of chest pain in 16 of these 64 patients (25%), the correction of hypoglycaemia in three patients (5%), and the cessation of convulsions in one patient (2%). The FRU assisted ambulance personnel during 351 missions, notably by carrying the patient to the ambulance. No

TABLE 3 Treatment modalities and responses

Treatment modalities and responses	Patient treated by the FRU n = 793	%
Clinical response recorded	223	
Resuscitation	83	
ROSC	17	20
ROSC by FRU alone	1	1
Airway management excl. CPR	7	
Respiratory compromise resolved	4	57
Oxygen administration	309	
Respiratory state improved	126	41
Chest pain relief	20	6
Medication	64	
Chest pain relief	16	25
Anticonvulsive or normoglycaemic effect	4	6
Spinal immobilization/splinting	32	
Pain relief/prevention	3	9
Recovery position/postural treatment	34	
Respiratory state improved	4	12
Pain relief	4	12
Other clinical response[a]	17	

ROSC, return of spontaneous circulation; FRU, first-responding unit; CPR, cardiopulmonary resuscitation.
[a]Other clinical response includes for example, improved haemodynamic state.

clinical evaluation or treatment was recorded in 607 cases (37%) during which a patient was encountered.

The performance of the professional- and trained volunteer layperson-staffed units during FHQ missions is shown in Table 4. Statistically significant differences were observed in attempted resuscitation rates (professional 46 attempts per 71 cardiac arrests [65%] vs layperson 37 attempts per 43 cardiac arrests [86%]; $P = 0.017$) and oxygen administration rates during ischaemic stroke and chest pain missions (stroke: professional 9 per 236 cases [4%] vs layperson 26 per 181 cases [14%], $P < 0.001$; chest pain: professional 16 per 78 cases [21%] vs layperson 77 per 159 cases [48%], $P < 0.001$). Respiratory state was reported to improve more often during chest pain missions treated by laypersons as compared to professionals (4 vs 24 cases; $P = 0.031$).

TABLE 4 Comparison of professional versus trained volunteer first-responding units' performance in first hour quintet (FHQ) missions

Treatment modality/response	Professional n = 489	%	Volunteer n = 533	%	P-value
Cardiac arrest					
Confirmed cardiac arrest	71		43		
Resuscitation by the FRU	46	65	37	86	0.017
ROSC	8	11	9	21	0.182
Median time from dispatch to scene, min (IQR)	6	5-9	9	7-16	<0.001
Severe respiratory failure					
Dispatches	38		76		
Airway management	0		0		
Oxygen administration	16	42	35	46	0.842
Respiratory state improved	12	32	25	33	1.000
Chest pain					
Dispatches	78		159		
Oxygen administration	16	21	77	48	<0.001
Medication	10	13	24	15	0.697
Oxygen and medication	5	6	18	11	0.254
Chest pain relief	5	6	23	14	0.087
Shortness of breath improved	4	5	24	15	0.031
Severe trauma					
Dispatches	58		74		
Immobilisation/splinting	13	22	14	19	0.667
Pain relief/prevention	1	1	1	2	1.000
Stroke					
Dispatches	236		181		
Carrying/assistance for ambulance	94	34	36	27	0.213
Oxygen administration	9	4	26	14	<0.001

FHQ, first hour quintet; FRU, first-responding unit; ROSC, return of spontaneous circulation; IQR, interquartile range.

4 | DISCUSSION

In this study, the general performance of 44 first-responding units in the county of Pirkanmaa, Finland, was evaluated during a 1-year period. FRUs were dispatched to cardiac arrests but also to several other prehospital emergencies, such as stroke, respiratory failure, chest pain and trauma. As one would expect, stroke and chest pain were more common events than cardiac arrests in our material. This should be considered when planning training programs for FRUs and implementing FRUs in other settings. To our knowledge, there are no extensive reports concerning the general performance or clinical impact of first-responding units.

The mainstay of a first-responding unit is that it truly is the first responder or otherwise a rapid responder, especially in time-critical emergencies such as cardiac arrest. In our data, the median time needed for an FRU to reach the scene after dispatch was 9 minutes.

Compared with this study, the existing literature has described shorter response intervals ranging from 3.5 to 8 minutes, most importantly in the time-critical context of out-of-hospital cardiac arrests.[1,4-6] According to our data, this was achieved more often by professional FRUs, with the specific response times in cases of cardiac arrest being 6 and 9 minutes for the professional and layperson units respectively. A theoretical model describing the performance of EMS in Stockholm suggests that the shortest achievable interval from time of incidence to defibrillation is 6.5 minutes if the driving time to the scene is 1 minute.[9]

In cardiac arrest, every minute delay in CPR and defibrillation increases mortality, and thus, every minute saved by the use of an FRU is important.[10,11] However, in this study, the FRU was the true first responder initiating CPR in only 42 cases (51%), and ROSC was achieved in 17 of the 83 cardiac arrests cases (20%) in which CPR was attempted. A previous study of firefighter first responders showed that an FRU was first on the scene in 41% of 1961 out-of-hospital cardiac arrest missions in Stockholm, Sweden.[5] The same study reported that FRUs and EMS achieved ROSC in 29% of missions. Furthermore, in a Danish study, firefighter FRUs achieved ROSC in 7 of 29 cases (24%) when an AED was attached.[1]

Given that FRUs attended many other medical emergencies than cardiac arrest patients, we evaluated the procedures performed prior to ambulance arrival and whether these procedures had any relief to the patient. According to our data, supplemental oxygen appeared to be the most common treatment modality. This may be attributed to the fact that FRUs had no other treatment option they found reasonable, for example in stroke missions. Additionally, oxygen administration in patients presenting with dyspnoea was effective. Oxygen was supplied to 309 patients in all mission and to 51 patients with severe respiratory failure. Subsequent normalisation of oxygen saturation or relief of shortness of breath was reported in 41% and 73% of those cases respectively. Regarding the form and the clinical impact of FRU-provided treatment other than resuscitation and oxygen administration, medicinal or procedural treatment by FRUs was uncommon, occurring in approximately 8% of attended cases.

As a part of the evaluation, the contributions of 14 professional fireman- or EMT-staffed FRUs and 30 layperson-staffed FRUs to emergency response were examined in FHQ missions and no clinically significant differences were seen. However, because of missing data, the results of the comparison must be interpreted cautiously.

Among confirmed cardiac arrest missions, volunteer FRUs were more likely to be involved in CPR attempts as compared to professional FRUs. This may reflect the capability of EMT-staffed professional FRUs to critically evaluate the potential futility of a resuscitation attempt and also their stronger adherence to pre-existing guidelines when resuscitation is not attempted. Lay rescuers may also initiate CRP more frequently when no legal consequence is followed.[12] Furthermore, resuscitation attempts may result in psychological stress, especially to non-professional rescuers.[13] Both professional and volunteer first responders were able to participate in debriefing sessions during the study period.

As discussed above, oxygen administration was common. In cases of ischaemic stroke and chest pain, oxygen was administrated by layperson FRUs more often than by professionals although the units had no different treatment protocols in this regard. The potentially toxic effects of oxygen in myocardial ischaemia have been under strict evaluation during the past years, and currently, the routine administration of oxygen is not recommended unless signs of hypoxia, dyspnoea or heart failure are present.[14,15] Therefore, the more liberal administration of oxygen by laypersons and the indications for oxygen use in this study warrant further evaluation and clinical guidance.

This study has several inherent limitations because it is a retrospective chart review. First, the exact interval between patient evaluation, physiological measurements and treatment responses could not be determined. Second, the outcome measures may have been affected by EMS. The FRUs are advised to describe only the assessment and treatment provided by the FRU in their documentation, whereas for the EMS units, there is a separate form of documentation. In cases of the simultaneous arrival of the FRU and the EMS unit at the scene, some of the procedures performed by the EMS personnel may have also been documented on the FRU forms. Indeed, during certain missions, the FRU was always dispatched, regardless of the time benefit as compared with the EMS ambulance. The arrival of the FRU and the initial treatment during these missions (witnessed cardiac arrest, high-energy fall trauma and presumed ischaemic stroke) may well have occurred simultaneously or even after that of the ambulance. Nevertheless, the magnitude of these procedures (eg, airway management, medication) is small in relation to the entire material, suggesting that the role of the FRU, in this sense, is not strong. Third, a large degree of heterogeneity in terms of FRUs' skill level makes the comparison between professional versus volunteer units difficult to quantify with statistics. Finally, the paper mission forms were often incompletely filled yielding a large amount of missing data. A future prospective study is warranted to provide more complete data.

5 | CONCLUSION

In conclusion, first-responding units initiate treatment or assist ambulance personnel in approximately half of the cases attended. Implementation of both professional firefighters and trained volunteer as FRUs in FHQ missions seems to be feasible.

ACKNOWLEDGEMENTS

This work is dedicated to Dr Janne Virta (1969-2016).

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ORCID

Joonas I. Tamminen http://orcid.org/0000-0003-4434-0499

REFERENCES

1. Høyer CB, Christensen EF. Fire fighters as basic life support responders: a study of successful implementation. *Scand J Trauma Resusc Emerg Med*. 2009;17:16.
2. Rørtveit S, Meland E. First responder resuscitation teams in a rural Norwegian community: sustainability and self-reports of meaningfulness, stress and mastering. *Scand J Trauma Resusc Emerg Med*. 2010;18:25.
3. Roberts A, Nimegeer A, Farmer J, Heaney DJ. The experience of community first responders in co-producing rural health care: in the liminal gap between citizen and professional. *BMC Health Serv Res*. 2014;14:460.
4. Zijlstra JA, Stieglis R, Riedijk F, Smeekes M, van der Worp WE, Koster RW. Local lay rescuers with AEDs, alerted by text messages, contribute to early defibrillation in a Dutch out-of-hospital cardiac arrest dispatch system. *Resuscitation*. 2014;85:1444-1449.
5. Nordberg P, Hollenberg J, Rosenqvist M, et al. The implementation of a dual dispatch system in out-of-hospital cardiac arrest is associated with improved short and long term survival. *Eur Hear J Acute Cardiovasc Care*. 2014;3:293-303.
6. Saner H, Morger C, Eser P, von Planta M. Dual dispatch early defibrillation in out-of-hospital cardiac arrest in a mixed urban-rural population. *Resuscitation*. 2013;84:1197-1202.
7. Boland LL, Satterlee PA, Fernstrom KM, Hanson KG, Desikan P, LaCroix BK. Advanced clinical interventions performed by emergency medical responder firefighters prior to ambulance arrival. *Prehospital Emerg Care*. 2015;19:96-102.
8. Fischer M, Kamp J, Garcia-Castrillo Riesgo L, et al. Comparing emergency medical service systems – a project of the European Emergency Data (EED) Project. *Resuscitation*. 2011;82:285-293.
9. Sund B. Developing an analytical tool for evaluating EMS system design changes and their impact on cardiac arrest outcomes: combining geographic information systems with register data on survival rates. *Scand J Trauma Resusc Emerg Med*. 2013;21:8.
10. Weston CFM, Wilson RJ, Jones SD. Predicting survival from out-of-hospital cardiac arrest: a multivariate analysis. *Resuscitation*. 1997;34:27-34.
11. Pons PT, Haukoos JS, Bludworth W, Cribley T, Pons KA, Markovchick VJ. Paramedic response time: does it affect patient survival? *Acad Emerg Med*. 2005;12:594-600.
12. Mathiesen WT, Bjørshol CA, Høyland S, Braut GS, Søreide E. Exploring how lay rescuers overcome barriers to provide cardiopulmonary resuscitation: a qualitative study. *Prehosp Disaster Med*. 2017;32:27-32.
13. Mathiesen WT, Bjørshol CA, Braut GS, Søreide E. Reactions and coping strategies in lay rescuers who have provided CPR to out-of-hospital cardiac arrest victims: a qualitative study. *BMJ Open*. 2016;6:e010671.
14. Rincon F, Kang J, Maltenfort M, et al. Association between hyperoxia and mortality after stroke: a multicenter cohort study. *Crit Care Med*. 2014;42:387-396.
15. Nikolaou NI, Arntz HR, Bellou A, Beygui F, Bossaert LL, Cariou A. European resuscitation council guidelines for resuscitation 2015 section 8. initial management of acute coronary syndromes. *Resuscitation*. 2015;95:264-277.

How to cite this article: Tamminen JI, Hoppu SE, Kämäräinen AJJ. Professional firefighter and trained volunteer first-responding units in emergency medical service. *Acta Anaesthesiol Scand*. 2019;63:111–116. https://doi.org/10.1111/aas.13224

PUBLICATION
III

Random forest machine learning method outperforms prehospital National Early Warning Score for predicting one-day mortality: a retrospective study

Jussi Pirneskoski, Joonas Tamminen, Antti Kallonen, Jouni Nurmi, Markku Kuisma, Klaus T. Olkkola, Sanna Hoppu

Resuscitation Plus 2020;4:100046
DOI: 10.1016/j.resplu.2020.100046

Available online at www.sciencedirect.com

Resuscitation Plus

journal homepage: www.journals.elsevier.com/resuscitation-plus

EUROPEAN RESUSCITATION COUNCIL

Clinical paper

Random forest machine learning method outperforms prehospital National Early Warning Score for predicting one-day mortality: A retrospective study

Jussi Pirneskoski[a,*,1], *Joonas Tamminen*[b,c,1], *Antti Kallonen*[b], *Jouni Nurmi*[a], *Markku Kuisma*[a], *Klaus T. Olkkola*[d], *Sanna Hoppu*[c]

[a] *Department of Emergency Medicine and Services, University of Helsinki and HUS Helsinki University Hospital, Helsinki, Finland*
[b] *Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland*
[c] *Emergency Medical Services, Tampere University Hospital, Tampere, Finland*
[d] *Department of Anaesthesiology, Intensive Care and Pain Medicine, University of Helsinki and HUS Helsinki University Hospital, Helsinki, Finland*

Abstract

Aim of the study: The National Early Warning Score (NEWS) is a validated method for predicting clinical deterioration in hospital wards, but its performance in prehospital settings remains controversial. Modern machine learning models may outperform traditional statistical analyses for predicting short-term mortality. Thus, we aimed to compare the mortality prediction accuracy of NEWS and random forest machine learning using prehospital vital signs.

Methods: In this retrospective study, all electronic ambulance mission reports between 2008 and 2015 in a single EMS system were collected. Adult patients ($\geq$ 18 years) were included in the analysis. Random forest models with and without blood glucose were compared to the traditional NEWS for predicting one-day mortality. A ten-fold cross-validation method was applied to train and validate the random forest models.

Results: A total of 26,458 patients were included in the study of whom 278 (1.0%) died within one day of ambulance mission. The area under the receiver operating characteristic curve for one-day mortality was 0.836 (95% CI, 0.810–0.860) for NEWS, 0.858 (95% CI, 0.832–0.883) for a random forest trained with NEWS variables only and 0.868 (0.843–0.892) for a random forest trained with NEWS variables and blood glucose.

Conclusion: A random forest algorithm trained with NEWS variables was superior to traditional NEWS for predicting one-day mortality in adult prehospital patients, although the risk of selection bias must be acknowledged. The inclusion of blood glucose in the model further improved its predictive performance.

Keywords: Emergency medical services, Prehospital, Cardiac arrest prevention, Early warning score, National Early Warning Score, NEWS, Random forest, Machine learning

Introduction

The National Early Warning Score (NEWS) is a validated method for predicting deterioration in hospital wards.[1,2] It has been shown to predict short-term mortality in prehospital environments in retrospective studies,[3–7] but its role in prehospital clinical decision making remains controversial.[8]

Recent in-hospital studies have demonstrated that novel machine learning methods can surpass traditional early warning scores in predicting admission, the need for intensive care and short-term mortality at emergency departments as well as in detecting impending

* Corresponding author at: Department of Emergency Medicine and Services, University of Helsinki and HUS Helsinki University Hospital, PO Box 340, 00029 HUS, Helsinki, Finland.
E-mail address: jussi.pirneskoski@helsinki.fi (J. Pirneskoski).
[1] Equal contribution.
http://dx.doi.org/10.1016/j.resplu.2020.100046
Received 15 May 2020; Received in revised form 25 September 2020; Accepted 27 October 2020

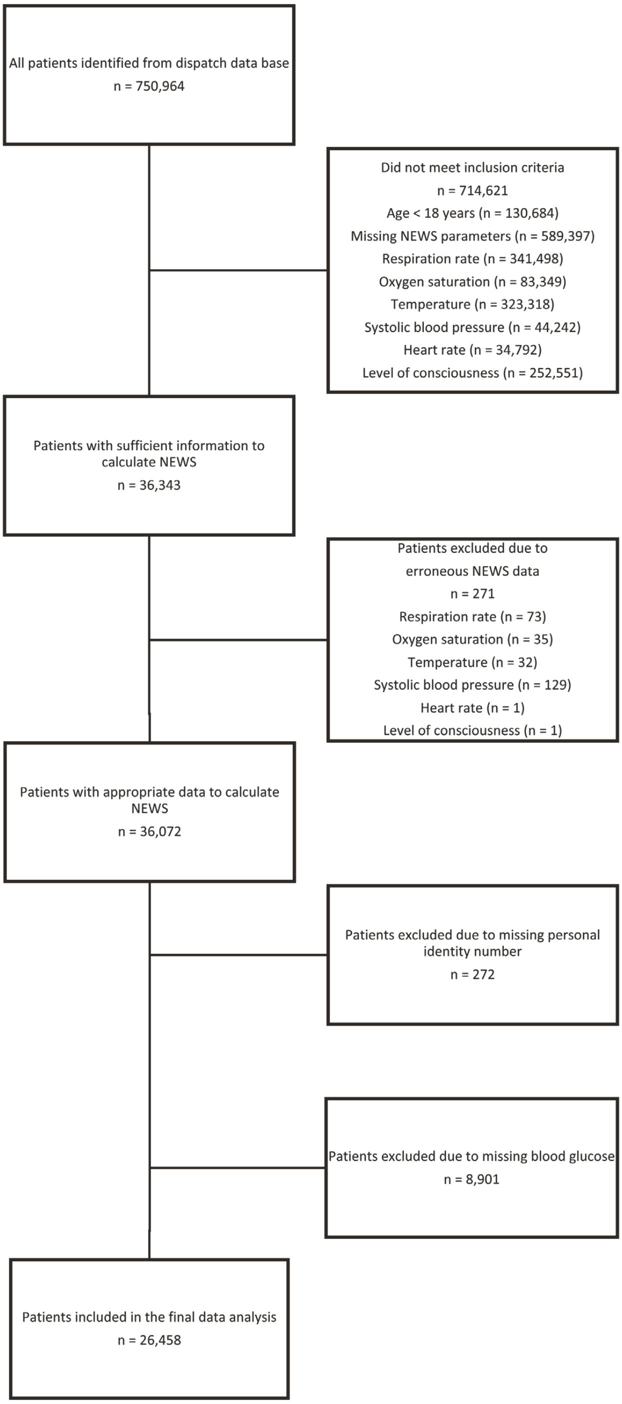

Fig. 1 – Flow chart of study cohort selection.

sepsis in wards.[9–12] However, information from prehospital environments is scarce.[13] Such machine learning methods could be trained to consider a number of the prognostically valuable variables that are recorded in prehospital electronic patient record systems. For instance, it has been suggested that adding blood glucose as a physiological parameter into the NEWS system could improve its predictive performance.[14]

In this study, we aimed to compare the predictive performance of NEWS and random forest machine learning models incorporating NEWS variables and blood glucose for one-day mortality in previously collected prehospital material.

Methods

Ethical considerations

The study protocol followed the principles of the Declaration of Helsinki and was approved by the Department of Emergency Medicine and Services, HUS Helsinki University Hospital (§68, 11.11.2015). No informed consent or ethics committee approval is required by Finnish legislation for a retrospective registry study such as this.

Study population

We collected all of the electronic ambulance mission reports in the Helsinki and Uusimaa Hospital District, Finland, made between August 17th 2008 and December 18th 2015, excluding cases without the vital signs required to calculate NEWS values and blood glucose measurement. By excluding all patients with these missing variables, we maximised the quality of the data for statistical analysis and avoided imputations in machine learning model training, while recognising the possibility of causing selection bias. In a secondary analysis, cases with appropriate data to calculate NEWS but possibly unknown blood glucose measurement were examined. Study area EMS system and dispatch process are described in detail elsewhere.[6]

Data handling and statistical analysis

The mission data had been recorded in an electronic patient record system (Merlot Medi, CGI Suomi Oy, Helsinki, Finland). The physiological variables of oxygen saturation, heart rate and blood pressure were automatically recorded from monitors whereas respiratory rate, body temperature, level of consciousness and oxygen use required manual input. The initial values for each physiological variable were used for the analysis, except for heart rate and oxygen saturation for which a mean of the first five minutes was used. One-day mortality was selected as our primary outcome since it was considered to be suitable for prehospital setting regarding clinical decision making.[3] The Digital and Population Data Services Agency.

As this was post-hoc analysis, no power calculations were performed for this specific research question. Statistical analysis was performed using Python (version 3.6.9), and the main statistical packages used were NumPy (version 1.17.3) and sklearn (version 0.21.3).

We selected the random forest as the machine learning method for this study as it has been shown to outperform traditional regression.[15] It is a supervised machine learning approach known to extract information from noisy input data and learn highly nonlinear relationships between input and target variables. Random forest models are very resistant to overfitting and can learn from imbalanced predictor class presentation.

Random forests are a collection of computer-generated decision trees. A single decision tree is not able to on complex problems, but a collection of these weak learners has been shown to work well in many prediction tasks involving human physiology.[16] In order to train a random forest, a training feature space is randomly populated with a uniform sampling of input feature thresholds at each split node in order to maximise information gain for the entire forest.[17]

Table 1 – Characteristics of the study cohort and overall adult population.

	Study cohort	All patients age > 18 years
n	26,458	620,280
Age, mean, SD (years)	65.6, 19.9	60.6, 21.4
Male sex, n, %	12,783, 48.3%	n/a
NEWS, median, IQR	3, 1–6	n/a
Respiration rate, median, IQR (min^{-1})	16, 15–20	16, 15 – 18
Blood oxygen saturation, median, IQR (%)	96, 93–98	97, 95 – 98
Use of supplemental oxygen, n, %	4,564, 17.2%	41,669, 6.7%
Body temperature, median, IQR (ºC)	36.8, 36.3–37.3	36.8, 36.4–37.3
Systolic blood pressure, median, IQR (mmHg)	142, 123–164	141, 124–160
Heart rate, median, IQR (min^{-1})	87, 73–103	86, 74–101
Level of consciousness on AVPU scale, n, %		
Alert	20,281, 76.6%	n/a
Reacts to voice	2,507, 9.5%	n/a
Reacts to pain	2,246, 8.5%	n/a
Unresponsive	1,424, 5.4%	n/a
Blood glucose, median, IQR (mmol/l)	7.2, 6.0–9.1	n/a
Primary complaint, n, %		
Trauma	1,757, 6.6%	130,538, 21.0%
Medical	24,701, 93.4%	489,742, 79.0%

IQR: interquartile range, n/a: not available.

Model evaluation was performed using ten-fold stratified cross-validation in which training is followed by testing for ten times. Each fold presents an independent data subset to the random forest algorithm and uses a different data subset to estimate predictive performance using the area under the receiver operating characteristic curve (AUROC) performance metric. These generated folds were later used to computationally estimate confidence intervals for the different predictors using bootstrap resampling with 10,000 sample points as the normality of cross-validated AUROC scores is not guaranteed. The overall performance of the model is the combination of the bootstrap samples from the ten testing folds (i.e. AUROC distributions). 95% confidence intervals (CI) were calculated for the continuous variables; all AUROC results are presented with 95% CI in parentheses. Bootstrapping method was also used to estimate p-values (null hypothesis for equal AUROCs) numerically.

Results

A total of 26,458 prehospital EMS patients were included in the study (Fig. 1). Of these patients, 278 (1.0%) died within one day. None of the deaths occurred at the scene. The demographic characteristics of included and excluded patients are presented in Table 1. Prehospital use of supplemental oxygen was more common in the study cohort patients, but otherwise the groups were similar in terms of NEWS variables.

The AUROC for one-day mortality using NEWS was 0.836 (95% CI 0.810–0.860). The corresponding AUROC values determined with the random forest models trained with NEWS variables only and with NEWS variables and blood glucose were 0.858 (0.832–0.883) and 0.868 (0.843–0.892), respectively (Fig. 2). The AUROC of the random forest models were significantly higher than that of NEWS (P=0.005 NEWS variables only and P0.001 NEWS variables and glucose). The AUROCs of the two random forests also differed significantly (P=0.032).

In a secondary analysis regarding patients with all NEWS variables measured (n=35,800), the AUROC for one-day mortality using NEWS and random forest trained with NEWS variables only were 0.850 (95% CI 0.829–0.868) and 0.873 (95% CI 0.854–0.892, P<0.001 compared with NEWS), respectively (Fig. 3).

Discussion

Principal findings

In the present study, a random forest machine learning method using NEWS variables outperformed NEWS in predicting one-day mortality

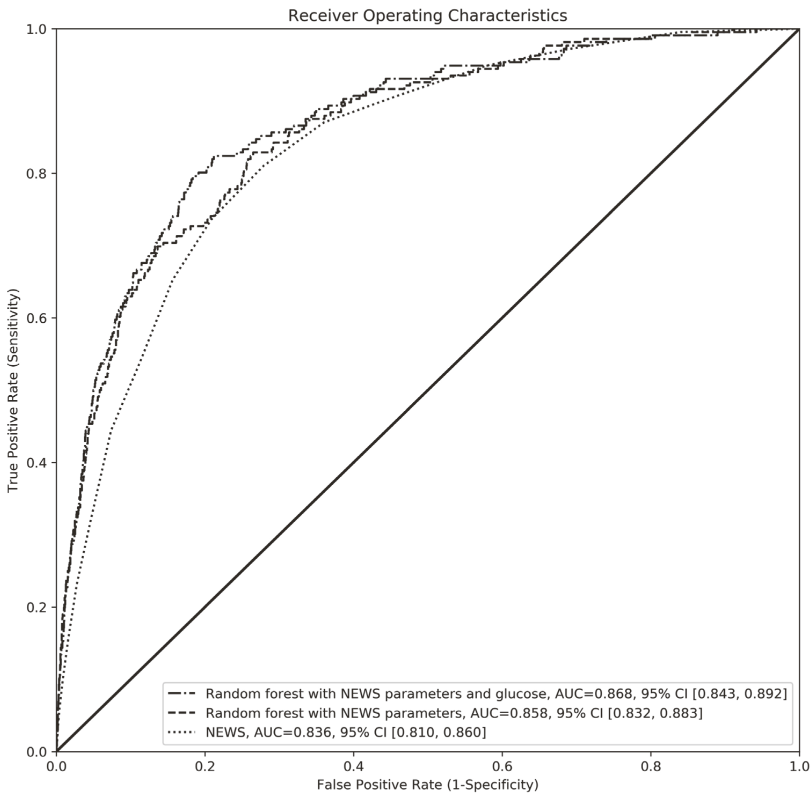

Fig. 2 – Receiving operating characteristics curves for the three models: model based on NEWS score, model based on random forest trained with NEWS variables data and model based on random forest trained with NEWS variables data and blood glucose. Random forest produces a prediction as a probability and NEWS scores may be also interpreted as a probability when scaled with the maximum score value.

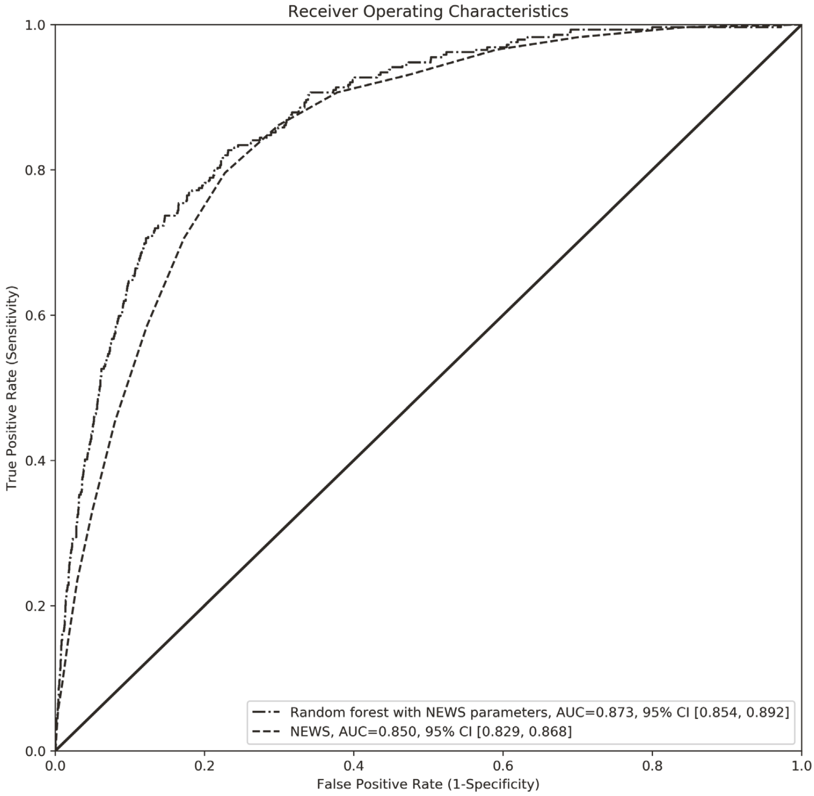

Fig. 3 – Receiving operating characteristics curves for a model based on NEWS score and a model based on random forest trained with NEWS variables.

in adult prehospital patients, both with and without a blood glucose variable.

Relation of results to other studies

Our results support the recent results of Spangler et al.[13] and confirm their finding of a machine learning method surpassing a traditional NEWS approach to prehospital risk stratification. Although a different machine learning method was used and a composite risk score of multiple outcomes was assessed, their data nevertheless further demonstrate the feasibility of using machine learning approaches to prehospital risk assessment. Our results are also in line with in-hospital emergency department studies that have compared machine learning to traditional early warning scores or triage tools and shown improved predictive performance by the machine learning techniques.[9–11]

Relevance of the study results

NEWS may not be the optimal tool for detecting impending cardiac arrest in prehospital settings since randomly selected prehospital patients may differ in terms of factors predicting mortality from in-hospital ward patients for whom NEWS was originally developed. As such, the physiological thresholds that are used in NEWS may not be valid. In a systematic review regarding the performance of prehospital NEWS, the authors concluded that only extreme aggregate scores (i.e. NEWS = 0 or 7) could reliably predict clinically relevant outcome.[8] Use of the random forest method allows for more precise physiological weighting and can model complex non-linearities in a given population. It also allows the incorporation of multiple variables including factors beyond the traditional vital signs, such as blood glucose which has been shown to improve mortality prediction in this context.[14]

On the other hand, on a secondary analysis performed on a larger cohort of patients focusing solely on NEWS parameters without glucose, the performance slightly improved, although the statistical significance of this improvement could not be tested due to the differing patient cohorts. We speculated this was likely due to the larger data set including a higher number of mortalities and therefore presenting more learning targets for random forest. This outlines the rationale of NEWS in including relevant physiological parameters to predict short-term mortality, which can still be utilized for even better predictions when analysed using novel methodology such as random forest.

The purpose of all early warning scores is to assist the detection of physiological abnormalities before they lead to cardiac arrest, and so a machine learning model could help reduce by improving the detection of patients at risk.[18] Training the model with local data would help overcome issues of generalisability of data from other populations. In that way, the model would adapt to the system-specific population and could be retrained over time with larger datasets or respond to changes in care guidelines or population demographics. In EMS that operate with electronic patient record systems, introducing automatically computed predictions for short-term mortality at the scene could help in patient-specific decision making when personnel need to consider the urgency of transport or non-conveyance. At this moment, these predictions may provide some guidance to clinicians as they are not standalone risk stratifications systems yet.

Strengths and limitations of the study

This was a retrospective study and the results are not fully generalisable because of selection bias risk from the very large exclusion rate and the fact that some mission data are collected over ten years ago. Missing values were not imputed, which is an important limitation of the study. Despite this, and the lack of power calculations, we consider the cohort of 26,489 patients including 278 mortalities sufficiently powered. Decisions to limit care such as do not resuscitate orders are not systematically entered in prehospital patient records, and it is possible that the existence of such orders could have affected the outcome of some patients.

All machine learning methods have common inherent limitations and 'artificial intelligence' should be considered as a sophisticated algorithm which can give accurate answers to a simple and narrow question. We are aware that our random forest model is a more complicated version but a more powerful version of NEWS which is tailored in our system. The most significant limitation of the random forest approach is the non-generalisability of the dataset since the exact AUROC value is likely to differ across different patient cohorts. However, we compared the same dataset using different predictive models in this study and therefore believe that the observed differences in the predictive values for one-day mortality of prehospital patients are true. Another important limitation concerning random forest is that it has a 'black box' element. As the name of the method implies, hundreds of decision trees are randomly generated into the model. Their clinical interpretation is extremely difficult although the decisions trees can be visualised (Supplemental Figure 4).

Future studies

A further prospective study is warranted to validate this new risk stratification model. Taking into account the issues of generalisability, a similar study in a different prehospital population could be considered. Given that a small minority of patients involved in EMS missions die within one day of contact, other outcomes such as the use of emergency department resources or the need for hospitalisation should also be looked at in future studies. In addition, further research into other possible variables to be considered in machine learning models is essential.

Conclusions

We have demonstrated that a random forest machine learning model was superior to NEWS in predicting one-day mortality in adult prehospital patients.

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.resplu.2020.100046.

REFERENCES

1. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. Report of a working party. London: Royal College of Physicians; 2012.
2. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation 2013;84:465–70.
3. Silcock DJ, Corfield AR, Gowens PA, Rooney KD. Validation of the National Early Warning Score in the prehospital setting. Resuscitation 2015;89:31–5.
4. Shaw J, Fothergill RT, Clark S, Moore F. Can the prehospital National Early Warning Score identify patients most at risk from subsequent deterioration? Emerg Med J 2017;34:533–7.
5. Abbott T, Cron N, Vaid N, Ip D, Torrance H, Emmanuel J. Pre-hospital National Early Warning Score (NEWS) is associated with in-hospital mortality and critical care unit admission: a cohort study. Ann Med Surg 2018;27:17–21.
6. Pirneskoski J, Kuisma M, Olkkola KT, Nurmi J. Prehospital National Early Warning Score predicts early mortality. Acta Anaesth Scand 2019;63:676–83.
7. Martín-Rodríguez F, López-Izquierdo R, del Vegas C, et al. Accuracy of National Early Warning Score 2 (NEWS2) in prehospital triage on in-hospital early mortality: a multi-center observational prospective cohort study. Prehosp Disaster Med 2019;34:1–9.
8. Patel R, Nugawela MD, Edwards HB, et al. Can early warning scores identify deteriorating patients in pre-hospital settings? A systematic review. Resuscitation 2018;132:101–11.
9. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo Jr CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019;23:64.
10. Hong W, Haimovich A, Taylor AR. Predicting hospital admission at emergency department triage using machine learning. PLoS One 2018;13:e0201016.
11. Levin S, Toerper M, Hamrock E, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index. Ann Emerg Med 2017;71:565–74.
12. Giannini HM, Ginestra JC, Chivers C, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. Crit Care Med 2019;47:1485–92.
13. Spangler D, Hermansson T, Smekal D, Blomberg H. A validation of machine learning-based risk scores in the prehospital setting. PLoS One 2019;14:e0226518.

14. Vihonen H, Lääperi M, Kuisma M, Pirneskoski J, Nurmi J. Glucose as an additional parameter to National Early Warning Score (NEWS) in prehospital setting enhances identification of patients at risk of death: an observational cohort study. Emerg Med J 2020;37:286–92.
15. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016;44: 368–74.
16. Lin K, Hu Y, Kong G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. Int J Med Inform 2019;125:55–61.
17. Breiman L. Random forests. Mach Learn 2001;45:5–32.
18. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. NEJM 2019;380:1347–58.

PUBLICATION
IV

Machine learning model predicts short-term mortality among prehospital patients: a prospective development study from Finland

Joonas Tamminen, Antti Kallonen, Sanna Hoppu, Jari Kalliomäki

Resuscitation Plus 2021;5:100089
DOI: 10.1016/j.resplu.2021.100089

Available online at www.sciencedirect.com

Resuscitation Plus

journal homepage: www.journals.elsevier.com/resuscitation-plus

EUROPEAN RESUSCITATION COUNCIL

Clincal paper

Machine learning model predicts short-term mortality among prehospital patients: A prospective development study from Finland

Joonas Tamminen[a,b,1,*], *Antti Kallonen*[a,1], *Sanna Hoppu*[b], *Jari Kalliomäki*[b,c]

[a] *Faculty of Medicine and Health Technology, Tampere University, PO Box 2000, FI-33521 Tampere, Finland*
[b] *Emergency Medical Services, Tampere University Hospital, PO Box 2000, FI-33521 Tampere, Finland*
[c] *Intensive Care Medicine, Tampere University Hospital, PO Box 2000, FI-33521 Tampere, Finland*

Abstract

Aim: To show whether adding blood glucose to the National Early Warning Score (NEWS) parameters in a machine learning model predicts 30-day mortality more precisely than the standard NEWS in a prehospital setting.

Methods: In this study, vital sign data prospectively collected from 3632 unselected prehospital patients in June 2015 were used to compare the standard NEWS to random forest models for predicting 30-day mortality. The NEWS parameters and blood glucose levels were used to develop the random forest models. Predictive performance on an unknown patient population was estimated with a ten-fold stratified cross-validation method.

Results: All NEWS parameters and blood glucose levels were reported in 2853 (79%) eligible patients. Within 30 days after contact with ambulance staff, 97 (3.4%) of the analysed patients had died. The area under the receiver operating characteristic curve for the 30-day mortality of the evaluated models was 0.682 (95% confidence interval [CI], 0.619−0.744) for the standard NEWS, 0.735 (95% CI, 0.679−0.787) for the random forest-trained NEWS parameters only and 0.758 (95% CI, 0.705−0.807) for the random forest-trained NEWS parameters and blood glucose. The models predicted secondary outcomes similarly, but adding blood glucose into the random forest model slightly improved its performance in predicting short-term mortality.

Conclusions: Among unselected prehospital patients, a machine learning model including blood glucose and NEWS parameters had a fair performance in predicting 30-day mortality.

Keywords: Machine learning, Prehospital, Risk stratification, NEWS

Introduction

Various early warning score (EWS) systems have been introduced to facilitate clinical decision-making in hospital wards; their aim is to detect an inpatient's physiological deterioration prior to adverse outcomes.[1−4] These systems report an aggregate score of physiological measurements of the patient's vital functions. A higher score indicates an increased risk of a short-term medical emergency (e.g. 24-h, 48-h and 30-day mortality, admission to an intensive care unit [ICU] or sepsis).

The signs of impending physiological deterioration and subsequent cardiac arrest can be observed hours before cardiovascular collapse,[5,6] and the Royal College of Physicians advocates the use of the National Early Warning Score (NEWS) also in the prehospital setting.[1] However, the performance of any prehospital EWS system to predict short-term mortality is modest, as only the extreme aggregate scores (i.e. NEWS = 0 or 7) predict a clinically relevant outcome.[7,8]

* Corresponding author at: Medical School, University of Tampere and Emergency Medical Services, Tampere University Hospital, PO Box 2000, FI-33521 Tampere, Finland.

E-mail address: joonas.i.tamminen@tuni.fi (J. Tamminen).

[1] Equal contribution.

http://dx.doi.org/10.1016/j.resplu.2021.100089
Received 27 September 2020; Received in revised form 18 January 2021; Accepted 20 January 2021
Available online xxx

Therefore, the standard NEWS' predictive performance should be further strengthened, especially for moderate-risk patients. Retrospective data suggest that adding blood glucose level to NEWS in the prehospital setting and some inflammatory biomarkers to NEWS in the emergency department might improve its performance.[9,10] In addition, modern machine learning methods tailored to a given patient population, such as the random forest (RF) method, seem to outperform traditional logistic regression models in predicting mortality among hospitalised ward patients.[11] RF is a modern machine learning method based on multiple randomly derived decision trees.[12]

We hypothesised that RF algorithms based on readily available physiological measurements would outperform the standard NEWS in the prehospital setting for predicting adverse outcomes. This development study compared the standard NEWS' diagnostic performance to that of RF algorithms trained with NEWS parameters and blood glucose levels for predicting 30-day mortality in unselected adult prehospital patients.

Methods

Design

This descriptive cohort study was conducted in the Tampere University Hospital (Tays) District, Finland. The city of Tampere and the surrounding rural and suburban areas cover a population of 520,000.[13] The emergency medical services (EMS) system comprises first-response units and basic level ambulances, advanced-level ambulances and a physician-staffed helicopter emergency services unit. The study area has one tertiary hospital, one regional hospital and 18 municipal primary health care centres.

The need for informed patient consent was waived, since the study design was observational, involving no interventions to standard therapy. The Tays Ethics Committee reviewed the study protocol (approval no: R10111, May 5th 2015).

Study cohort

The study cohort consisted of all consecutive adult patients (age $\geq$18 years) that the EMS personnel encountered from June 1st 2015 up to and including June 30th 2015. Cases with unknown civil registration numbers or missing case report forms, EMS-encountered cardiac arrest or EMS-confirmed death at the scene, in terminal care, transported to other hospital districts or encountered by EMS units from another district were excluded, since calculation of NEWS would be inappropriate or unfeasible in such cases.

Outcomes

The primary outcome was 30-day mortality. The patient mortality data were retrospectively extracted from the Digital and Population Data Services Agency. The secondary outcomes were 24-h and 48-h mortality, ICU admission and a composite outcome of 48-h mortality or ICU admission.

Predictors

The predictor variables of NEWS and the RF models were prospectively collected, and NEWS scores were retrospectively calculated. During the study period, the EMS was mandated to complete all NEWS parameters (i.e. respiration rate, oxygen saturation [SpO_2], administration of supplemental oxygen, systolic blood pressure, heart rate, level of consciousness and temperature) in every encountered patient regardless of the mission type at the scene before any intervention. The completeness of the NEWS parameters was verified by medical students in the emergency department of the tertiary hospital during the data collection. In the emergency department, there were altogether six medical students who worked in different shifts around the clock. The medical students audited the paper CFRs by re-checking the medical reports. A second audit was made by the author J.K while he transferred the paper CRFs to a digital format.

Contrary to the standard NEWS, the level of consciousness was assessed with the Glasgow Coma Scale, and it was entered as a categorical predictor variable into the RF models. In addition to the standard NEWS parameters, blood glucose level was included as a continuous variable in the RF models. Clinical judgement was used to ascertain whether the patient's blood glucose level was measured. The indications for measuring blood glucose were (1) known type 1 or type 2 diabetes, (2) altered level of consciousness or (3) suspected acute myocardial infarction or stroke. If the same patient had multiple contacts with the EMS personnel during the one-month study period, only the first contact was included in the analysis. Additionally, a sensitivity analysis based on the last contact in the study period was performed.

Sample size and missing data

The study material was collected for a manuscript in preparation which shares the same raw data but has a different aim and design. Since the present study was a post hoc analysis, no formal sample size calculations were performed for this research question. The development of the models was a complete-case analysis in which patients with any missing NEWS parameter or unknown blood glucose level were excluded.

Statistical analysis

All statistical analyses were done using Python language version 3.6.9 or R version 4.0.0. The main statistical packages used were NumPy version 1.17.3 and sklearn version 0.21.3 for Python. Continuous data were presented as means or medians and standard deviations or interquartile ranges, respectively, and categorical data were reported in frequencies and percentiles. The comparison between the groups was performed using a chi-squared test for the categorical data and a Mann−Whitney U-test for the continuous data.

Model development

RF was selected as a machine learning method for this study since it has outperformed logistic regression and the Modified Early Warning Score in in-hospital settings.[10] In our study, two RF models were developed: (1) an RF model derived from NEWS parameters only and (2) an RF model derived from NEWS parameters and blood glucose levels. Since additional input features are not detrimental to the RF model's performance, we decided to use all input features in the model development. The RF models were developed by applying ten-fold stratified cross-validation,[14] where each fold presents an independent subset of the data to the RF algorithm to train on and uses another subset to estimate predictive performance with an area under the receiver operating characteristics curve (AUROC) performance metric. Stratified division of the folds was used to keep the ratio of deceased patients in the training data the same as in the whole population.

Confidence intervals [CIs] for the cross-validated AUROC scores were calculated using bootstrapping with 10,000 sample points. This bootstrapped distribution of AUROC scores may exhibit non-normal distribution, so the intervals were calculated numerically using the sampled bootstrap distribution to make sure the values were representative.

Model comparisons

Performance of the different RF models and the standard NEWS was compared using the same cross-validation folds for each classifier. To make NEWS scores comparable to a supervised machine learning method, a dummy classifier was designed, which is able to output the score for a cross-validation fold. The bootstrapping method was also used to estimate p-values that were numerically calculated.

Results

EMS was dispatched to 6202 missions, and 4994 prehospital patients were contacted by ambulance personnel during the one-month study period. Of these patients, 3632 met our inclusion criteria. A total of 2853 (79%) patients had complete vital sign data and were included in the primary endpoint analysis (Fig. 1). A minority of the eligible patients with all the NEWS variables measured were excluded due to a missing blood glucose level (96/2949; [3.3%]). All missing vital signs in the eligible patients are presented in Table S1 in the Supplementary Appendix.

The study population's baseline characteristics are presented in Table 1. The 3632 patients eligible for analysis and the 2853 analysed patients were similar in terms of NEWS parameters, blood glucose level, 30-day, 24-h and 48-h mortality and ICU admission. A majority of the study population were low risk patients (i.e. had a NEWS score 0 −4). The mean age of the analysed patients was slightly higher than that of the eligible patients (66 years vs. 63 years, $p < 0.001$). Over one-third of the patients were left at the scene (34% [957] of the analysed patients and 36% [1313] of the eligible patients, p = 0.34).

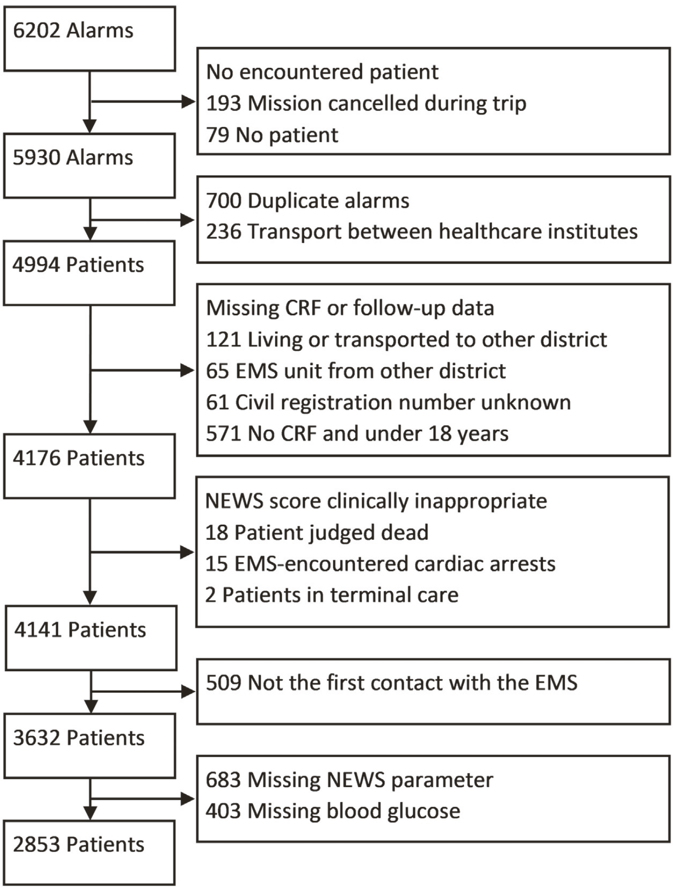

Fig. 1 – Formation of the study population. CRF = case report form; EMS = emergency medical services; NEWS = national early warning score.

Table 1 – Baseline characteristics.

N	Analysed patients 2853	Eligible patients 3632
Age, mean (SD); years	66 (21)	63 (21)
Male sex, %	50	50
NEWS score, median (IQR)	1 (0–3)	1 (0–3)
0, n (%)	735 (26)	1057 (29)
Total 1–4, n (%)	1721 (60)	2122 (58)
3 in single parameter, n (%)	607 (21)	704 (19)
Total 5–6, n (%)	195 (6.8)	228 (6.3)
Total 7 or more, n (%)	202 (7.1)	225 (6.2)
Respiration rate, median (IQR); min^{-1}	16 (15–18)	16 (15–18)
Oxygen saturation, median (IQR); %	97 (95–98)	97 (95–98)
Any supplemental oxygen, %	8.2	7.6
Temperature, median (IQR); °C	36.7 (36.2–37.1)	36.7 (36.3–37.1)
Systolic blood pressure, median (IQR); mmHg	143 (127–164)	143 (127–163)
Heart rate, median (IQR); min^{-1}	85 (72–100)	86 (73–100)
Glasgow Coma Scale >13, %	94	94
Blood glucose, median (IQR); mmol/l	6.7 (5.7–8.2)	6.6 (5.6–8.2)
Glasgow Coma Scale, median (IQR)	15 (15–15)	15 (15–15)
Transportation to, %		
Emergency department	40	38
General practitioner	19	19
Central hospital	6	5
Detoxification centre or jail	2	2
Not transported	34	36
30-day mortality, n (%)	97 (3.4)	114 (3.1)
24-h mortality, n (%)	13 (0.5)	16 (0.4)
48-h mortality, n (%)	18 (0.6)	22 (0.6)
ICU admission, n (%)	32 (1.1)	46 (1.3)
ICU admission/48-h mortality, n (%)	49 (1.7)	66 (1.8)

SD = standard deviation; NEWS = National Early Warning Score; IQR = interquartile range; ICU = intensive care unit.

Within a month after contact with the EMS personnel, 114 (3.1%) eligible patients died. Of these patients, 97 (84%) were analysed.

Fig. 2 shows the receiver operating characteristic (ROC) curves for 30-day mortality. The ROC curves for the secondary outcomes are presented in the supplementary appendix (Fig. S3). Table 2 summarises the cross-validated AUROCs with bootstrapped CIs and p-values for pairwise comparison. The RF models had greater AUROC for 30-day mortality than NEWS (NEWS 0.682 [95% CI, 0.619–0.744]; RF including NEWS parameters was 0.735 [95% CI, 0.679–0.787], $p = 0.008$ compared with NEWS; RF including NEWS parameters and blood glucose was 0.758 [95% CI, 0.705–0.807], $p < 0.001$ compared with NEWS).

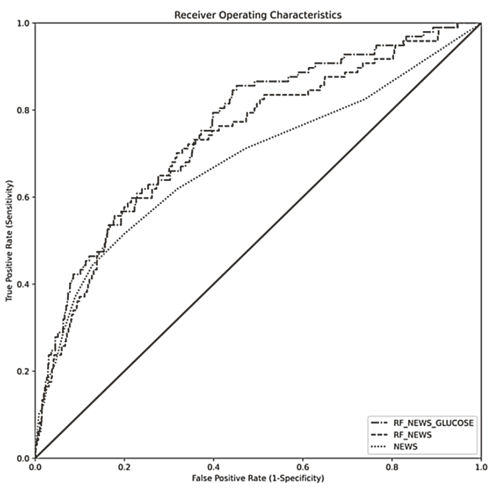

Fig. 2 – Area under the receiver operating characteristics curves for 30-day mortality.

In relation to the secondary outcomes, the AUROCs for the two RF models did not differ from the standard NEWS, but NEWS and both the RF models performed well in predicting 24-h mortality (NEWS 0.895 [95% CI, 0.816–0.961], RF including NEWS parameters was 0.899 [95% CI, 0.811–0.954] and RF including NEWS parameters and blood glucose was 0.953 [95% CI, 0.927–0.976]). The AUROCs for 48-h mortality, ICU admission and their combination were similar in all three models. A sensitivity analysis based on the last contact showed only minor changes to the models' performance (Table S4).

Discussion

Principal findings

In this prospective study, we collected NEWS parameters and blood glucose levels and developed machine learning algorithms to predict 30-day mortality among unselected adult emergency patients. We found that the RF models performed better in predicting 30-day mortality than the standard NEWS. However, the clinical significance of this finding could be questioned as the 95% CIs for the AUROCs are overlapping to a rather large degree. Regarding the secondary

Table 2 – AUROCs with 95% confidence intervals and pairwise comparisons for the cross-validated models.

	NEWS	RF 1	RF 2	p-value		
				NEWS vs RF 1	NEWS vs RF 2	RF 1 vs RF 2
30-d mortality	0.682 (0.619−0.744)	0.735 (0.679−0.787)	0.758 (0.705−0.807)	0.008	<0.001	0.074
24-h mortality	0.890 (0.797−0.966)	0.875 (0.707−0.976)	0.940 (0.860−0.985)	0.89	0.36	0.46
48-h mortality	0.845 (0.729−0.936)	0.808 (0.629−0.957)	0.881 (0.751−0.972)	0.52	0.32	0.12
ICU admission	0.806 (0.715−0.887)	0.807 (0.714−0.890)	0.814 (0.726−0.892)	0.94	0.73	0.72
ICU admission or 48-h mortality	0.818 (0.749−0.882)	0.811 (0.739−0.877)	0.847 (0.785−0.902)	0.74	0.07	0.09

AUROC = area under the receiver operating characteristics curve; NEWS = National Early Warning Score; RF 1 = random forest trained with NEWS parameters only; RF 2 = random forest trained with NEWS parameters and glucose; ICU = intensive care unit.

outcomes, including 48-h mortality and ICU admission, the standard NEWS and the RF model that included blood glucose levels performed equally. That RF model showed excellent performance in predicting 24-h mortality.

Relation of results to other studies

Machine learning models have been developed for various medical purposes.[15] In emergency medicine, speech recognition has been proposed to enhance dispatch, and a machine learning model has been tested for risk stratification at emergency departments.[16,17] Some in-hospital studies have used patients' vital signs and laboratory tests to predict physiological deterioration, sepsis or in-hospital cardiac arrest.[11,18] A recent prehospital study used an RF model to evaluate predictors of 30-day survival in patients with out-of-hospital cardiac arrest.[19]

To the best our knowledge, only one study has also used readily available information in a prehospital setting to train a machine learning model as a risk stratification tool.[20] Spangler et al. found that a different machine learning method to ours (XGBoost) trained with ambulance record data (i.e. vital signs, patient demographics and mission characteristics) was superior to the traditional NEWS in predicting 48-h mortality (AUROC for NEWS, 0.85 [95% CI, 0.83−0.87] vs. AUROC for XGBoost, 0.89 [95% CI, 0.87−0.91]). Contrary to this study, their study population was more selective, as they excluded patients left at the scene or transported to the non-emergency department. In our material, a third of the patients were left at the scene and a quarter of the patients had a NEWS score of 0, which may indicate that our patient population was less severely ill.

Disturbances in glucose homeostasis might precede impending physiological deterioration or be its consequence in diabetic and non-diabetic emergency patient populations.[9,21] Vihonen et al. found that the standard NEWS and their NEWSgluc logistic regression model had similar AUROCs for 30-day and 24-h mortality, but severe hypoglycaemia was noted to be an important prehospital predictor for death at 30 days (blood glucose 3.0 mmol/L or less; unadjusted odds ratio, 2.06 [95% CI, 1.28−3.19]). However, their study had a notable selection bias, as only 4% of the included patients were analysed. In our study, we observed that measuring blood glucose when clinically appropriate slightly improved the RF model's ability to predict 24-h, 48-h and 30-day mortality. This may indicate that an elevated blood glucose level and stress hyperglycaemia should be suspected at a low threshold among moderate-risk emergency patients but not be measured routinely.

Clinical implications

Machine learning algorithms could be utilised more extensively in the prehospital setting, as digital reporting is becoming more common in ambulances. Currently, some NEWS parameters (except for respiratory rate, level of consciousness and body temperature) are already automatically sent to an electronic emergency patient record system in most hospital districts in Finland.[22] Within the next two years, all Finnish EMS systems will have a uniform electronic patient record system. These electronic data could be entered simultaneously during patient evaluation to a machine learning algorithm, which would calculate estimates of short-term (e.g. 24-h, 48-h and 30-day) mortality. These estimates could facilitate EMS staff's recognition of high-risk patients who might otherwise be left at the scene or transported to inappropriate destinations. Future machine learning studies should utilise all available data that are documented in electronic patient record systems, as based on this study, the traditional NEWS parameter and blood glucose combined have a limited potential to predict 30-day mortality.

Strengths and limitations

Our study has limitations attributable to its observational design. First, blood glucose measurements were based on clinical judgement. This introduces selection bias, since these patients were more likely to be higher-risk patients. Nevertheless, only 3.6% of the patients who were otherwise eligible for analysis had an unknown blood glucose level. Second, RF was selected as a machine learning method, although it is unknown which machine learning method is the most suitable for our research question. Third, bootstrapped CIs for AUROC and p-values seem to be conflicting: p-values suggested a statistical significance whereas CIs partly overlapped. Nevertheless, the null hypothesis can be rejected at the $\alpha = 0.05$ level in this kind of scenario.[23] Finally, our RF model's performance should be externally validated in another prospectively collected dataset.

The most noteworthy strengths of the study are related to its design. During the prospective data collection, the ambulance personnel were mandated to measure the standard NEWS parameters in all adult patients encountered, regardless of the type of mission. Additionally, the study population included patients left at the scene or transported to a general practitioner, which further strengthens model's ability to detect moderate-risk patients among all prehospital patients.

Conclusion

An RF algorithm combining traditional NEWS parameters and blood glucose levels showed a fair performance in predicting 30-day mortality among unselected prehospital patients. Blood glucose improved the RF model's predictive power slightly.

Conflicts of interest

None.

CRediT authorship contribution statement

Joonas Tamminen: Conceptualization, Methodology, Formal analysis, Writing - original draft. **Antti Kallonen:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization. **Sanna Hoppu:** Conceptualization, Investigation, Resources, Writing - review & editing, Supervision, Project administration. **Jari Kalliomäki:** Conceptualization, Investigation, Resources, Data curation, Writing - review & editing, Supervision, Project administration.

Acknowledgements

This study was financially supported by the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (Grant 9V006).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.resplu.2021.100089.

REFERENCES

1. Royal College of Physicians. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS: Updated report of a working party: Executive summary and recommendations. London: Royal College of Physicians; 2017.
2. Subbe CP. Validation of a modified early warning score in medical admissions. QJM 2001;94:521–6.
3. Singh S, McGlennan A, England A, Simons R. A validation study of the CEMACH recommended modified early obstetric warning system (MEOWS). Anaesthesia 2012;67:12–8.
4. Akre M, Finkelstein M, Erickson M, Liu M, Vanderbilt L, Billman G. Sensitivity of the pediatric early warning score to identify patient deterioration. Pediatrics 2010;125:e763.
5. Schein RMH, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. Chest 1990;98:1388–92.
6. Hillman KM, Bristow PJ, Chey T, et al. Duration of life-threatening antecedents prior to intensive care admission. Intensive Care Med 2002;28:1629–34.
7. Silcock DJ, Corfield AR, Gowens PA, Rooney KD. Validation of the National Early Warning Score in the prehospital setting. Resuscitation 2015;89:31–5.
8. Patel R, Nugawela MD, Edwards HB, et al. Can early warning scores identify deteriorating patients in pre-hospital settings? A systematic review. Resuscitation 2018;132:101–11.
9. Vihonen H, Lääperi M, Kuisma M, Pirneskoski J, Nurmi J. Glucose as an additional parameter to National Early Warning Score (NEWS) in prehospital setting enhances identification of patients at risk of death: an observational cohort study. Emerg Med J 2020;286–92.
10. Eckart A, Hauser SI, Kutz A, et al. Combination of the National Early Warning Score (NEWS) and inflammatory biomarkers for early risk stratification in emergency department patients: results of a multinational, observational study. BMJ Open 2019;9:1–11.
11. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med. 2016;44:368–74.
12. Ho TK. Random decision forests. Proc 3rd Int Conf Doc Anal Recognit 1995 Aug 14–16; Montreal: IEEE Xplore; 1995. p. 278–282.
13. Statistics Finland. Official statistics of Finland (OSF): Population structure. (Accessed 7 January 2020, at. . https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto_en.html).
14. Refaeilzadeh P, Tang L, Liu H. Cross-validation. Encyclopedia of database systems. New York: Springer; 2009. p. 532–8.
15. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347–58.
16. Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 2019;138:322–9.
17. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019;23:1–13.
18. Giannini HM, Ginestra JC, Chivers C, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. Crit Care Med 2019;47:1485–92.
19. Al-Dury N, Ravn-Fischer A, Hollenberg J, et al. Identifying the relative importance of predictors of survival in out of hospital cardiac arrest: a machine learning study. Scand J Trauma Resusc Emerg Med 2020;28:60.
20. Spangler D, Hermansson T, Smekal D, Blomberg H. A validation of machine learning-based risk scores in the prehospital setting. PLoS One 2019;14:1–18.
21. Dungan KM, Braithwaite SS, Preiser JC. Stress hyperglycaemia. Lancet 2009;373:1798–807.
22. Pirneskoski J, Kuisma M, Olkkola KT, Nurmi J. Prehospital National Early Warning Score predicts early mortality. Acta Anaesthesiol Scand 2019;63:676–83.
23. Knezevic A. Overlapping confidence intervals and statistical significance. StatNews 2008. . (Accessed 22 June 2020, at https://cscu.cornell.edu/news/statnews/stnews73.pdf).