PhD thesis

Niels Egholm Pedersen

Early warning scores for the detection of critical illness in hospitalised patients

Academic supervisors: Doris Østergaard, Thomas Alexander Gerds, Lars S. Rasmussen, Anne Lippert

This thesis was submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen, on 1 September 2017
Academic supervisors
Doris Østergaard, MD, DMSc, MHPE, Professor
Copenhagen Academy for Medical Education and Simulation, Capital Region of Denmark,
University of Copenhagen, Herlev Ringvej 75, 2730 Herlev, Denmark

Thomas Alexander Gerds, Dr.rer.nat, Professor
Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014
Copenhagen K, Denmark

Lars S. Rasmussen, MD, PhD, Professor
Department of Anaesthesia, Centre of Head and Orthopaedics, Rigshospitalet, University of
Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark

Anne Lippert, MD, CHPE
Copenhagen Academy for Medical Education and Simulation, Capital Region of Denmark,
University of Copenhagen, Herlev Ringvej 75, 2730 Herlev, Denmark

Assessment committee
Ann Merete Møller, MD, DMSc, Professor (chair)
Department of Anaesthesia, Herlev Hospital, University of Copenhagen

Mikkel Brabrand, MD, PhD, Associate Professor
Center Sydvestjylland, University of Southern Denmark

Trond Nordseth, MD, PhD, Associate Professor
Department of Anaesthesiology and Intensive Care, St Olavs Hospital HF, Trondheim
The Medical Faculty, Norwegian University of Science and Technology

Defence
Herlev Hospital, 2017

Front page picture: The view from the Copenhagen Academy of Medical Education and
Simulation researchers’ office on 9 February 2017
Table of contents

ABBREVIATIONS ........................................................................................................................................5

ACKNOWLEDGEMENTS .............................................................................................................................6

STUDIES ......................................................................................................................................................7
Study 1 .........................................................................................................................................................7
Study 2 .........................................................................................................................................................7
Study 3 .........................................................................................................................................................7

THESIS AT A GLANCE .................................................................................................................................8

SUMMARY IN DANISH (DANSK RESUMÉ) .................................................................................................9
Baggrund .....................................................................................................................................................9
Metode .........................................................................................................................................................9
Resultater .....................................................................................................................................................10
Konklusioner ..............................................................................................................................................10
Fremtidig udvikling ....................................................................................................................................11

SUMMARY ..................................................................................................................................................12
Introduction ................................................................................................................................................12
Methods ......................................................................................................................................................12
Results .......................................................................................................................................................13
Conclusions .............................................................................................................................................13
Future development ..................................................................................................................................14

INTRODUCTION .........................................................................................................................................15
Critical illness and adverse events in hospitalised patients ........................................................................15
Rapid Response Systems (RRS) ..................................................................................................................15
Development and recommendation of EWSs ..............................................................................................16
Impact of EWS on patient outcomes ........................................................................................................17
Challenges in validation of EWSs ..............................................................................................................18
Challenges with EWSs in patients with chronic disease .........................................................................19
The EWS in the Capital Region of Denmark ..............................................................................................19
Rationale behind the studies in the thesis ................................................................................................................................................................................................. 20

AIMS ................................................................................................................................................................................................................................. 20

METHODS .................................................................................................................................................................................................................. 21

Setting ...................................................................................................................................................................................................................... 21

Study 1: Delphi consensus methodology study ..................................................................................................................................................... 21
  Selection of experts ............................................................................................................................................................................................ 22
  Delphi consensus process ............................................................................................................................................................................... 22

Study 2 and 3: Analysis of the EWS and CROS data from the Capital Region of Denmark ................................................................. 23
  Populations ........................................................................................................................................................................................................ 23
  Data collection and management ................................................................................................................................................................. 24
  End points .......................................................................................................................................................................................................... 25
  Analyses ............................................................................................................................................................................................................... 27

APPROVALS ........................................................................................................................................................................................................ 27

RESULTS .................................................................................................................................................................................................................. 28

Study 1: End points for validating early warning scores in the context of rapid response systems: a Delphi consensus study ................................................................................................................................................................................................. 28

Study 2: A critical assessment of early warning score records in 168,000 patients .................................................................................. 29

Study 3: Modifications of the National Early Warning Score in patients with chronic respiratory disease ........................................ 31

DISCUSSION ....................................................................................................................................................................................................... 32

Rationale behind Study 1 ................................................................................................................................................................................... 32

End points for validation of EWSs ................................................................................................................................................................. 33

EWS data collected during clinical practice ..................................................................................................................................................... 34

EWSs in patients with chronic respiratory disease ........................................................................................................................................... 36

Methodological considerations ................................................................................................................................................................. 37
  Study 1 ............................................................................................................................................................................................................. 37
  Study 2 ............................................................................................................................................................................................................. 38
  Study 3 ............................................................................................................................................................................................................. 39

Implication for the use of EWSs in clinical practice .................................................................................................................................................. 43

3
Abbreviations

ABCDE, airway, breathing, circulation, disability, exposure
AUC, area under the receiver-operator curve, or C-statistic
AWTTS, aggregated, weighted track and trigger system
CAMES, Copenhagen Academy for Medical Education and Simulation
COPD, chronic obstructive pulmonary disease
CREWS, Chronic Respiratory Early Warning Score
CROS, Capital Region of Denmark EWS Override System
CRN, civil registration system numbers
EHR, electronic health record
EWS, Early warning score
HDU, high dependency unit
ICU, intensive care unit
IQR, interquartile range
KISO, Clinical forms and overview. Part of the EHR. (Kliniske Inddateringsskemaer og Oversigter)
MET, medical emergency team
MEWS, modified Early Warning Score
NPV, negative predictive value
NEWS, National Early Warning Score
OPUS, the name of the EHR in the Capital Region of Denmark
PPV, positive predictive value
PROMS, patient reported outcome measures
S-NEWS, Salisbury NEWS modification
PPV, positive predictive value
RAM, random access memory
RRS, rapid response system
ViEWS, VitalPack™ Early Warning Score
Acknowledgements

First, I would like to thank my academic supervisors for their great support throughout the process of this PhD. **Doris Østergaard** for seeing possibilities in every idea, channelling those in a productive academic direction, and going to great lengths to provide the organisational support such a project needs. **Anne Lippert** for introducing me to her large network of people in the EWS / RRS research community, sharing her lengthy experience in working with EWS / RRS, for being deeply involved in every part of the work with this thesis, and providing guidance throughout the course of the project. **Thomas Alexander Gerds** for saying “I’d better help with that” when I contacted Copenhagen University’s statistics on-call service at an early stage of the process, and subsequently moving on to give great feedback on issues that would otherwise have been unresolvable for me. **Lars S. Rasmussen** for giving invaluable feedback and all-important academic advice.

Likewise, I owe great thanks to **Asger Petersen** for important contributions to the articles in this thesis, and to **Gary B. Smith** and **John Kellett** for discussing the possibilities regarding studying EWS validation and performance at an early stage of the project. The **Assessment committee** has contributed significantly to the thesis by providing highly valuable feedback.

My colleagues at CAMES have been an everlasting source of crucial inspiration, reflection, support, and laughs all through the years while working on this thesis. **Camilla**, my wife, has been a superb and ever-steady support throughout, at a level where words do not suffice. Out twins **Fritz** and **Thorkild** joined in along the way and have, besides completing this project according to the humorous note that the main constituents of a medical PhD are three articles and two children, been magnificent at diverting my attention from the thesis and giving the academic mind a break when at home. **Hanne**, my mother, stayed with us for prolonged periods through the last six months of the PhD, ensuring the timely submission of the thesis. **Family and friends** have also contributed to this, and for that I am also most thankful.
Studies

The thesis is based on the following papers, which are reproduced on the last pages of this thesis:

Study 1

Study 2

Study 3
# Thesis at a glance

<table>
<thead>
<tr>
<th>Study</th>
<th>Aim</th>
<th>Methods and data sources</th>
<th>Main findings</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>To establish current expert opinion on end points for validating EWS and other tools for identifying patients in hospital wards at risk of critical illness.</td>
<td>Delphi consensus methodology study. Expert panel consisting of 17 EWS / RRS researchers.</td>
<td>A wide range of possible approaches to end points for validating the EWS was identified. No consensus was reached on any of the suggestions suit the purpose.</td>
<td>Currently, death, cardiac arrest, and ICU admission, are the best choices for end points for validating EWSs. Some of the suggested end points could be feasible in time.</td>
</tr>
<tr>
<td>2</td>
<td>To investigate the EWS data recorded manually and stored electronically over 12 months in the Capital Region of Denmark.</td>
<td>Observational study: 2,835,331 EWS records from the year 2014 in the EHR EWS registration system in the Capital region in Denmark.</td>
<td>10% of records were incomplete. Artefacts or extreme values were found in 0.2% of the records. Digit preferences and distributional anomaly of the recorded EWS variable values were found.</td>
<td>A hospital system with an uptake area of 1.7 million inhabitants produced 2.8 million EWS records from 168,000 admissions over 12 months; 10% were incomplete. Staff practice influenced the recorded EWS variable values.</td>
</tr>
<tr>
<td>3</td>
<td>To describe the use of the CROS formalised EWS override system in order to investigate how the CROS, CREWS, and S-NEWS affected the NEWS total scores, and to compare the NEWS with the CROS, CREWS, and S-NEWS in patients with chronic respiratory disease.</td>
<td>Observational study: 305,286 complete EWS records from the year 2014 from adult in-patients with chronic respiratory disease. 48-hour mortality and ICU admission within 48 hours of a EWS record were used as end points to calculate the sensitivity, specificity, PPV, NPV, and the AUC.</td>
<td>More than 1/3 of the NEWS records had point deductions due to the CROS. The CROS, CREWS, and S-NEWS all had lower sensitivities, higher specificities, and higher PPV than the NEWS. AUC was higher for NEWS than for CROS with 48-hour mortality as the end point but higher for CROS than for NEWS with ICU-admission as the end point.</td>
<td>The CROS was used frequently by clinicians. The CROS, CREWS, and S-NEWS all reduced sensitivity for 48-hour mortality and ICU admission.</td>
</tr>
</tbody>
</table>
Summary in Danish (dansk resumé)

Baggrund

Region Hovedstadens “tidlige advarselssystem” (EWS) er baseret på det ”nationale tidlige advarselssystem” (NEWS), som anbefalet af Royal College of Physicians i Storbritannien. Det erklærerede formål er ”at sikre rettidig og ensartet vurdering og identifikation af begyndende udvikling af kritisk sygdom hos indlagte patienter”. Region Hovedstadens EWS-system (CROS) er en tilpasning af NEWS, som ikke findes andre steder end i Region Hovedstaden. Det giver mulighed for at tildele acceptable kroniske værdier til individuelle patienter. I 2014 var elektronisk registrering af de 7 manuelt målte EWS-variable (respirationsfrekvens, arteriel ilmtætning, ilttilskud, pulsfrekvens, systolisk blodtryk, bevidsthedsniveau og kropstemperatur) obligatorisk.

Det overordnede mål med denne afhandling var at undersøge Region Hovedstadens EWS. Der blev gennemført tre studier med følgende formål:

1. At bestemme den aktuelle holdning blandt eksperter til hvilke endepunkter, der er brugbare til at validere EWS og lignende værktøjer til identifikation af indlagte patienter i risiko for kritisk sygdom.
2. At analysere de manuelt indsamlede, elektronisk gemte EWS data, der blev indsamlet i løbet af 12 måneder i Region Hovedstaden.
3. At undersøge hvordan CROS blev brugt, hvordan CROS, CREWS og S-NEWS (CREWS og S-NEWS er andre udgaver af NEWS, som er foreslået til patienter med kronisk lungesygdom) påvirkede NEWS totalscorene, samt hvordan sensitivitet, specificitet, positiv prædiktiv værdi, negativ prædiktiv værdi, samt AUC (et statistisk mål) var for NEWS og de modificerede udgaver af NEWS hos patienter med kronisk lungesygdom.

Metode


**Resultater**

Ekspertpanelet foreslog 86 forskellige endepunkter til validering af EWS, omhandlende 13 overordnede temaer. Der var ikke konsensus om at nogen endepunkter var ideelle.

Korttidsdødelighed, hjertestop og indlæggelse på intensivafdeling var blandt de højest rangerende endepunkter.

Der blev foretaget 2.835.331 EWS-registreringer i 2014. Af dem var de 10 % ufuldstændige og 0,2 % indeholdt usandsynlige værdier. Vi fandt cifferpræferencer i respirationsfrekvens, iltilskud, puls og systolisk blodtryk. Der var en ophobning af pulsregistreringer lige under 91 slag i minutet, hvilket er en grænse for pointgivning i EWS.

I alt 404.093 EWS-registreringer var fra patienter med kronisk lungesygdom. Af disse havde 40 % EWS pointfradrag på grund af CROS. Sensitiviteterne for CROS, CREWS og S-NEWS for 48-timers dødelighed og intensivindlæggelse var lavere end for NEWS. Specificiteterne og de positive prædiktive værdier for CROS, CREWS og S-NEWS var højere. AUC var højere for NEWS end for CROS med 48-timers mortalitet som endepunkt, men lavere for NEWS end for CROS med indlæggelse på intensivafdeling som endepunkt. CROS, CREWS og S-NEWS nedgraderede 1/2 til 1/3 af EWS totalscorerne fra et af totalscoreintervallerne ”obligatorisk lægetilstedeværelse” og ”øjeblikkelig lægetilstedeværelse og konsultation med specialist” til et lavere niveau.

**Konklusioner**

Hovedfundet i studie 1 var, at der ikke blev opnået konsensus om at nogen endepunkter var ideelle til brug ved validering af EWS. Endepunkter relateret til død, hjertestop og intensivindlæggelse blev vurderet til for nuværende at være de bedste kompromisser, mens andre forslag fremsat af ekspertpanelet kan blive relevante i fremtiden.
Menneskelige faktorer påvirker formodentlig de registrerede EWS-værdier i systemer med manuel måling og elektronisk registrering af værdier. Dette er en potentiell begrænsning i forhold til at udskifte manuelt registrerede målinger med fuldautomatiske målinger. CROS blev brugt hyppigt hos patienter med kronisk lungesygdom. Sensitiviteten for 48-timers dødelighed og intensivindlæggelse for CROS, CREWS og S-NEWS var mindre i forhold til NEWS. Specificiteterne, de positive prædiktive værdier og de negative prædiktive værdier var også påvirkede.

**Fremtidig udvikling**

Fremtidige forslag til tilpasninger og ændringer af EWS kan og bør evalueres i randomiserede, kontrollerede studier. Fortsat udvikling af EWS med de hidtil anvendte metoder og endepunkter risikerer at resultere i udviklingen af modeller, der er optimerede til at identificere uafvendeligt døende patienter. En bedre tilgang kunne være at anvende et komplekst, sammensat endepunkt, der identifierer tilstande som indebærer høj risiko for at føre til livstruende eller invaliderende sygdom, hvis de ikke behandles. Dette kunne baseres på en kombination af mange klasses af data fra elektroniske patientjournalsystemer, inklusive patient-rapporterede data, og diagnostiske klassifikationer. Alternativt kunne en forståelse af EWS som et mål for ”generel, fysiologisk afvigelse” simplificere fortolkningen af EWS betydeligt, da det er en direkte refleksion af, hvad EWS er i sig selv, uden behov for validering.
Summary

Introduction
The Capital Region of Denmark’s EWS (Early Warning Scores) is based on the NEWS recommended by the Royal College of Physicians in the United Kingdom. Its purpose is “timely and uniform evaluation and identification of incipient developing critical illness in hospitalised patients”. The CROS is a NEWS modification unique to the Capital Region of Denmark; it allows acceptable chronic values to be assigned to individual patients. By 2014, electronic registration was mandatory of the values of 7 manually collected EWS variables (respiratory rate, arterial oxygen saturation, supplemental oxygen, pulse rate, systolic blood pressure, mental state, and body temperature).
The overall aim of this thesis was to investigate the performance of the EWS in the Capital Region of Denmark. Three studies were performed with the following aims:
1. To establish current expert opinion on what end points are useful when validating EWSs and similar tools for identifying patients at risk of critically ill in hospital wards.
2. To critically assess over 12 months the EWS data recorded manually and stored electronically in the Capital Region of Denmark.
3. To investigate how the CROS was used, how the application of the CROS, CREWS, and S-NEWS affected the NEWS total scores, and how the NEWS and the NEWS modifications performed regarding sensitivity, specificity, PPV, NPV, and AUC in hospitalised patients with a diagnosis of chronic respiratory disease.

Methods
Study 1 used the Delphi consensus methodology to establish current expert opinion on end points for the validation of EWSs. Based on objective criteria, 22 experts were invited; 17 participated. Each expert’s suggestions for possible end points were collected, processed and distributed to the combined expert panel in an anonymous form. Using the Delphi process, the expert panel reviewed, rated and commented on the suggestions. The combined ratings for each suggestion were established.
Study 2 was an observational study using the EWS data collected during 2014 in the Capital Region of Denmark. The percentages were calculated of the EWS records that were incomplete, or contained extreme or implausible values. The distributions of the EWS variable values and the NEWS total scores were assessed using tables and frequency plots.
Study 3 was an observational study. Patients with a diagnosis of chronic respiratory disease were included. The use of the CROS was described. Sensitivities, specificities, PPV, NPV, and AUC for 48-hour mortality and ICU admission within 48 hours for the NEWS, CROS, CREWS, and S-NEWS were calculated. The percentages of records where the application of the CROS, CREWS, and S-NEWS affected the total score were also calculated.

Results
The expert panel suggested 86 end points for the validation of EWSs, relating to 13 overarching themes. Consensus was not reached on any of the suggested end points being ideal. Short-term mortality, cardiac arrest, and ICU admission were among the highest rated end points. There were 2,835,331 EWS records made during 2014; 10% were incomplete and 0.2% contained implausible values. We found digit preferences for respiratory rate, supplemental oxygen, pulse rate, and systolic blood pressure. There was an accumulation of pulse rate records below 91 beats per minute, which is a limit for EWS point generation. A total of 404,093 records were from patients with chronic respiratory disease; 40% of included patients had point deductions due to the CROS. The sensitivities of the CROS, CREWS, and S-NEWS for 48-hour mortality and ICU admission were lower than that of the NEWS. The specificities and the PPV of the CROS, CREWS, and S-NEWS were higher. AUC was higher for NEWS than for CROS with 48-hour mortality as the end point but lower for NEWS than for CROS with ICU-admission as the end point. The CROS, CREWS, and S-NEWS downgraded to a lower interval approximately 1/2 to 1/3 of records from one of the “mandatory doctor presence” and “immediate doctor presence and specialist consultation” total score intervals.

Conclusions
The main finding in Study 1 was that consensus was not reached on any appropriate end point for the validation of the EWS. End points related to death, cardiac arrest, and ICU admission were found currently to be the best compromises, while other suggestions made by the expert panel may be relevant in time. Human factors have probably affected the recorded EWS variable values in systems with manually collected, electronically recorded values. This is a potential limitation to the replacement of manually recorded measurements with fully automated measurements. The CROS was used frequently in patients with chronic respiratory disease. Sensitivities for 48-hour mortality and ICU admission of the CROS, CREWS, and S-NEWS were reduced compared with those of the NEWS. Specificities, PPVs, and NPVs were also affected.
It was not possible to draw firm conclusions on the performance of the EWS in the Capital Region of Denmark due to the shortcomings of the available end points for assessing EWSs.

**Future development**

Future suggested modifications and alterations of EWS can and should be evaluated in randomised, controlled studies. To continue developing EWSs using the current methodology and end points could result in the development of models optimised for identifying inevitably dying persons. A better approach could be to design a composite end point, identifying conditions, which are highly likely to lead to life-threatening of debilitating disease if they are left untreated. This could be based on a combination of many classes of data, including diagnostic classifications. Alternatively, branding EWS as a measure of “general, physiological deviation” would simplify the interpretation of EWSs greatly, as it directly reflects what EWS inherently are without any validation being required.
Introduction

Critical illness and adverse events in hospitalised patients

The in-hospital adverse events of death, cardiac arrest, and ICU admission have been found to be preceded by clinical signs of instability in the 24 hours leading up to the event in 45% to 60% of cases (1),(2). Prospectively collected abnormal vital sign observations have also been found to be associated with mortality (3). In Danish medical and surgical wards studied in 2006, it was found that 18% of patients had abnormal vital signs and that the nursing staff were not aware of these abnormalities in 43% of patients (4).

Rapid Response Systems (RRS)

A RRS is one approach to addressing the issue of staff being unaware of signs of in-hospital patients being at risk of imminent critical illness and ensuring that an appropriate response to patients at risk of critical deterioration is at hand (5),(6). In a RRS, an afferent, or “sensing”, limb monitors patients. If signs of deterioration in a patient are detected, an efferent, or “acting”, limb is activated (Figure 1). A consensus conference on the afferent limb of RRSs recommended that systematic monitoring of vital signs should be mandatory for all hospitalised patients for whom active treatment is a viable option (7). Numerous studies on the effect of the introduction

Figure 1: Schematic depiction of a rapid response system
of RRSs have been published. A controlled cluster-randomised trial found no difference in incidence of cardiac arrest, unplanned ICU admissions, or unexpected death between intervention and control hospitals (8). However, this study has been criticised for being underpowered, and for a low degree of implementation and adherence to escalation protocols (9). Several studies comparing numbers of adverse outcomes before and after the introduction of RRSs have also been published. The most recent metaanalyses reported RRSs to be associated with a reduction in hospital mortality and cardiac arrest (10),(11). As most of the studies in these metaanalyses were uncontrolled, some argue that there is insufficient evidence to support the implementation of RRS outside the context of evaluation (12). An argument to this effect is that uncontrolled studies can be seen as studies of the effect of informational and educational campaigns directed towards the detection of and reaction to in-patient deterioration just as much as studies of the effect of EWSs / RRSs. Others argue that the evidence of shortfalls in the recognition and treatment of unpredicted clinical deterioration might indicate that RRSs are beneficial, although the lack of studies demonstrating RRSs to be effective for reaching their stated aim is a concern(13). Yet others argue that RRSs work towards developing patient-safety systems and can improve end-of-life care, and dismiss the scepticism towards the result of the metaanalyses based on the pre-post studies, concluding that RRSs improve outcomes (14).

**Development and recommendation of EWSs**

EWSs have been designed as instruments for systematic monitoring of vital signs in hospitalised patients and thus fit into the afferent limb of a RRS. EWSs are aggregated weighted track and trigger systems (AWTTSs), where specific variables (respiratory rate, arterial oxygen saturation, oxygen supplementation, pulse rate, systolic blood pressure, level of consciousness, and body temperature) are routinely measured and recorded. To simplify interpretation, the measurements are aggregated to a single score (EWS), with each variable generating points depending on its deviation from the normal value.

The first published EWS was presented as a poster in 1997 (15). In 2001 the Modified Early Warning Score was the subject of the first published EWS validation study (16). Development of EWSs continued; a review published in 2007 identified and compared 33 different AWTTSs based on a palette of variables including pulse rate, respiratory rate, systolic blood pressure, body temperature, level of consciousness, arterial oxygen saturation, urine output, and age(17). In 2010, the ViEWS was developed based on expert opinion, the existing literature on the relation between antecedents to adverse clinical outcomes in hospitalised patients, and previously published EWSs. A vital signs database of manually recorded and entered,
electronically stored EWS data was used to compare the ViEWS and the 33 AWTTS identified in the review (18). In 2012 the Royal College of Physicians of the UK recommended the NEWS, which is a slight modification of the ViEWS, to be introduced nationally in the UK for 1) the assessment of acute illness, 2) the detection of clinical deterioration, and 3) the initiation of a timely and competent clinical response (19). The point scoring key for the NEWS is shown in Table 1. Along with the NEWS, an associated escalation protocol was devised (Appendix 1). This escalation protocol specifies standards for levels of monitoring and clinical responses based on the NEWS total score.

<table>
<thead>
<tr>
<th>Table 1: NEWS point scoring key</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 2 1 0 1 2 3</td>
</tr>
<tr>
<td>Respiratory rate (min-1)</td>
<td>≤ 8</td>
</tr>
<tr>
<td>Arterial oxygen saturation (%)</td>
<td>≤ 91</td>
</tr>
<tr>
<td>Supplemental oxygen (L/min)</td>
<td>0</td>
</tr>
<tr>
<td>Pulse rate (min-1)</td>
<td>≤ 40</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>≤ 90</td>
</tr>
<tr>
<td>Mental state (A: Alert, V: Verbal, P: Pain, U: Unresponsive) (20)</td>
<td>V, P, U</td>
</tr>
<tr>
<td>Body temperature (°C)</td>
<td>≤ 35.0</td>
</tr>
</tbody>
</table>

**Impact of EWS on patient outcomes**

A randomised controlled trial of an automated EWS with automated alerts, incorporating a number of variables not included in the NEWS, found no impact on the primary outcome of ICU transfer, but a modest reduction in the secondary outcome of hospital length of stay (21), (22). There was no effect on the secondary outcome of hospital mortality. Several observational studies and uncontrolled pre-post interventional studies of EWS have been made with ambiguous results with regard to health outcomes. One relatively recent review included studies reporting results concerning EWS’ predictive ability and impact on patient centred outcomes (23). It concluded that EWSs performed well for prediction of cardiac arrest and 48-hour mortality. However, the evidence base was insufficient for drawing any conclusions with regard to impact on health outcomes and resource utilisation, particularly due to the methodological flaws of the bias-prone pre-post study design and heterogeneity of the included studies. Another review included 7 pre-post studies of the impact of EWS on patient outcomes (24). This review stated inadequate methodologies and heterogeneity in the EWSs of the included studies as the main reasons for its inability to reach a general conclusion.
Challenges in validation of EWSs

Given the lack of evidence of impact on patient centred outcomes, the foundation for the use of EWSs in clinical practice relies on EWSs being able to correctly identify patients at risk of critical illness. Thus, it is desirable to obtain a thorough understanding of the relationship between EWS total scores and the risk of imminent clinical deterioration. The arguably most influential EWSs, namely the MEWS, ViEWS, and NEWS, have been validated using death, ICU/HDU admission, and cardiopulmonary arrest as the end point (16),(18),(25). Validations of EWSs using these end points contrast somewhat with EWSs’ clinical application. A key goal is to identify clinical deterioration in persons with conditions that can be corrected with an appropriate clinical response at an early state. However, EWSs are evaluated with regard to ability to find patients obtaining points identifying populations with a high fraction of unsalvageable patients, i.e. patients with conditions that cannot be stabilised even with timely, correct intervention. This is particularly the case for deaths, illustrated by a review of 1000 in-hospital deaths in the UK acute care hospitals reporting 5.2% of the deaths as having a 50% or greater chance of being preventable (26). With regard to cardiac arrests, an audit in National Health Service hospitals during a 14-day period by The United Kingdom National Confidential Enquiry into Patient Outcome and Death reported 38% of 413 cardiac arrests to have been avoidable (27).

In-hospital cardiac arrests in 144 UK hospitals were reported to have a rate of survival to discharge of 0.18 (28). In a study including 154 in-hospital cardiac arrests in a Norwegian hospital, the rate was 0.16 (29). The rate of survival to discharge from hospital in individuals admitted to ICUs from wards was reported to be 0.47 in a study published in 1998 based on data from 15 UK acute care hospitals (30). Thus, if one of these end-points is reached, the chance of survival is very low.

The key statistic in EWS validation has traditionally been the AUC. However, it has no clinically meaningful interpretation, e.g. one cannot infer from a certain EWS total to a specific probability of reaching a certain end point. Further, serious drawbacks have been noted in its lack of reflection of model calibration, inclusion of information about cut-off values that will never be used (e.g. very high EWSs), and the low PPVs that can be associated with an excellent AUC in low-incidence settings (31),(32).

With the increasing use of electronic EWS recordings, huge amounts of EWS data are available for EWS research. However, there are no studies available of the data and quality of data in such datasets or critical appraisals of the EWS variable values gathered in such datasets.
Challenges with EWSs in patients with chronic disease

Patients suffering from chronic disease can have deviations from the normal range on various parameters, including physiological variables, without this being associated with acute illness. When these values are measured, false alerts can be generated. It is likely that false alerts affect staff compliance with EWS monitoring and escalation protocols (33). This has been an area of debate and noted as a cause of concern regarding the NEWS, especially among patients with COPD (34),(35),(36),(37),(38). In the Royal College of Physicians’ NEWS recommendations, it was noted that “the chronically disturbed physiology of some patients with COPD could influence the sensitivity of the NEWS, which should be recognised when interpreting NEWS in these patients”. However, this area warrants further work (36). The CREWS and S-NEWS are two published NEWS modifications designed for use in patients with chronic pulmonary disease (39),(40).

The EWS in the Capital Region of Denmark

Two studies from a PhD thesis titled “Effect of multi-professional education of staff on recognition and management of patients at risk” from 2009 found patients with deviating vital signs to have an increased risk of 30-day mortality. However, ward staff were not aware of these signs in more than half the patients (4),(41). The thesis also described the development of a multi-professional simulation course in the recognition and management of deteriorating hospital patients (42). It was shown that the course had no effect on staff awareness of patients at risk, 30-day mortality, or incidence of patients with abnormal vital signs (43). Thus, while it was established that there were shortcomings regarding staff awareness of deviating vital signs, the ideal corrective measure was not devised. The thesis concluded that “…routines for assessment and monitoring of patients on general wards need to be re-evaluated”. In line with this conclusion, the Capital Region of Denmark introduced a modified version of the NEWS and an associated escalation protocol as a mandatory track-and-trigger system in all its hospitals during 2012 and 2013.

Similar systems have been introduced in all health care regions in Denmark. The stated aim of the Capital Region of Denmark EWS is to ensure “timely and uniform evaluation and identification of incipient developing critical illness in hospitalised patients”, and “timely, relevant actions based on recognised deterioration of the condition of hospitalised patients” (44). The basic variables and score-triggering levels in this EWS are identical to those of the NEWS, whereas the escalation protocol is more detailed than that of the NEWS (Appendix 1). Compared with the NEWS, the largest modification is the CROS formalised EWS override system,
intended for use in patients with chronic disease. The CROS gives the option of indicating “acceptable chronic value limits” for all the NEWS variables, except temperature, on an individual level. With the CROS, when an acceptable chronic value limit has been set for a EWS variable, only EWS variable values more extreme than the acceptable chronic value limit trigger points. This system was designed based on expert opinion and has not been validated.

**Rationale behind the studies in the thesis**
Due to the lack of good quality studies clearly demonstrating a benefit to patients from EWSs in the context of a RRSs, the use of EWSs rely on the assumption that it enables clinicians to systematically identify patients at risk of critical deterioration in order to react adequately in the face of threatening deterioration. Thus, we found that there was a need to validate that the EWS introduced in the Capital Region of Denmark, particularly the CROS modification, was suitable for the identification of patients at risk. As we were able to gather a year’s worth of electronic EWS data similar to those used in other EWS validation studies, we were confident that we had a good starting point for such an effort.

We found it necessary to devote attention to one of the most the fundamental parts of any EWS validation, namely the end points, as we found that the end points used in earlier EWS studies could be problematic.
While we had access to a large amount of EWS data, a critical appraisal of EWS data gathered in everyday clinical practice was lacking. We found a descriptive study of this data to be warranted before putting the data to use.

Finally, we wished to investigate CROS and other NEWS modifications designed for use in patients with chronic respiratory disease. We wanted to describe the frequency of use of CROS and the frequency of point deductions due to CROS. We found it relevant to provide sensitivities, specificities, PPVs, and NPVs for different cut-offs of NEWS and NEWS modifications, as well as AUCs for comparison with prior and subsequent studies. Additionally, these measures are relevant to for decision makers when evaluating EWSs and action algorithms, and to clinicians for evaluating the significance of individual EWSs when calculated using relevant end points.

**Aims**

The overall aim of this thesis was to investigate the performance of the EWS in the Capital Region of Denmark. The pursuit of this overall aim resulted in 3 studies with the following aims:
1. To establish current expert opinion on what end points are useful when validating EWS and similar tools for identifying patients at risk of critically ill in hospital wards.

2. To critically assess over 12 months the EWS data recorded manually and stored electronically in the Capital Region of Denmark.

3. To investigate how the CROS was used, how the application of the CROS, CREWS, and S-NEWS affected the NEWS total scores, and how the NEWS and the NEWS modifications performed regarding sensitivity, specificity, PPV, NPV, and AUC in hospitalised patients with a diagnosis of chronic respiratory disease.

**Methods**

The thesis is made up of one study using the Delphi consensus methodology, and two descriptive studies based on EWS and CROS data from the Capital Region of Denmark.

**Setting**

The Capital Region of Denmark’s 4 acute care hospitals, 6 secondary hospitals, and 1 tertiary referral hospital provide free-for-all hospital services to a population of 1.7 million in the greater Copenhagen area. Additionally, certain specialised services are provided to the population of other regions of Denmark, Greenland, and the Faroe Islands.

During 2012 and 2013, a paper-based EWS was developed, based on the NEWS, and adding a formalised EWS override system intended for use in patients with chronic disease. During 2013, electronic recording of EWS values in the Electronic Health Record (EHR) was introduced for all adult patients, except obstetric patients, for whom a separate scoring system was used, and ICU patients. By 1 January 2014, all EWS registrations were made in the EHR. The Capital Region of Denmark’s escalation protocol differed from the NEWS escalation protocol. It had 6 total score levels with matching instructions, whereas the NEWS had 4 and used the “single parameter score” criterion differently (Appendix 1).

In conjunction with the introduction of the EWS in the Capital Region of Denmark, a EWS research group was established. I was a member of this group during work on this thesis.

**Study 1: Delphi consensus methodology study**

To collect and process current expert opinion on end points for validating the EWS, a study based on the Delphi consensus methodology was conducted (45).
Selection of experts

We used searches on “early warning score” and “rapid response system” in PubMed to identify persons who had recently published studies on the design or application of EWSs and/or RRSs. Persons who had one primary authorship or three co-authorships on these subjects within the 5 years preceding the study were invited to participate as part of the expert panel.

Delphi consensus process

A 3-round Delphi process was used (Figure 2). Round one was a brainstorm among the experts, where they were asked for suggestions for potentially useful endpoints for the validation of EWSs. This was followed by a facilitation step, where two of the authors arranged all input in themes and identical suggestions were merged. In this step, emphasis was on including all suggestions from the experts and preserving the diversity among the suggestions even if differences appeared small. Suggestions considered to have a common theme were grouped for clarity. In the second round, the experts were presented with a questionnaire including the result...
of the first round and subsequent facilitation. A random order in which the different themes appeared on the questionnaires was made for each expert, using R (46). Each expert kept his or her individual order of themes throughout the study. Experts were asked to rate all items in the questionnaire and to state all comments they might have on any item. Answers from this round were processed and a new questionnaire was prepared for each expert. This questionnaire was used in Round 3. It included the experts’ scores from the previous round, the median and IQR of all scores, and all comments made for each item on the questionnaire. In Round 3 the experts reviewed the ratings and comments from the other experts and could change their previous score if upon reflection they found it appropriate. We had predefined that Round 3 should be repeated until consensus was obtained for all items, or until its third iteration had been completed.

A numerical ranking scale was used for the questionnaires to rank the usefulness of each item from 0 to 10. Consensus was defined as ratings of each individual item on the questionnaire being within a range of 3 points on this scale.

The expert panel members were blinded to the identity of each other until the completion of the study. In the experts’ comments, which were relayed through the questionnaire, those making the comments were identified by a serial number. Thus, it was possible for the expert panel members to identify only which comments had come from the same expert. Experts’ participation in the study was voluntary and unpaid.

**Study 2 and 3: Analysis of the EWS and CROS data from the Capital Region of Denmark**

Studies 2 and 3 were retrospective cohort studies using EWS- and CROS-related data from the EHR in the Capital Region of Denmark recorded in 2014.

**Populations**

The populations included in studies 2 and 3 are depicted in Figure 3. In Study 2, all patients admitted to a hospital in the Capital Region of Denmark during 2014 who were aged 16 years or more upon admission, and who had one or more NEWS variables recorded during admission, were included.

In Study 3, patients with chronic respiratory disease, defined as patients with one of the World Health Organisation’s international classification of diseases 10 diagnoses J40-J47, J60-J67, J68.4, J70.1, J70.3, J84.1, J92.0, J96.1, J98.2, or J98.3 recorded as a primary or secondary diagnosis during any admission in 2014, were included (47).
**Figure 3: Populations included in studies 2 and 3**

<table>
<thead>
<tr>
<th>Patients with EWS records in the electronic health record in the Capital Region of Denmark during 2014</th>
<th>Excluded patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 182,403</td>
<td>Cause</td>
</tr>
<tr>
<td></td>
<td>System test patients</td>
</tr>
<tr>
<td></td>
<td>No records related to any admission</td>
</tr>
<tr>
<td></td>
<td>Patient 15 years or younger</td>
</tr>
<tr>
<td></td>
<td>Duplicate entries</td>
</tr>
<tr>
<td></td>
<td>No EWS or variable data</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patients included in study 2</th>
<th>Excluded patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 168,496</td>
<td>Cause</td>
</tr>
<tr>
<td></td>
<td>No chronic respiratory disease</td>
</tr>
<tr>
<td></td>
<td>No complete EWS record</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patients included in study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 11,266</td>
</tr>
</tbody>
</table>

**Data collection and management**

The EWS and CROS data were manually recorded and entered into the KISO module of the OPUS which is the electronic health record used by hospital staff. The data were delivered by the Capital Region of Denmark Centre for IT, Medico Technology and Telephony section Clinical Information and Data.

Information on admissions, discharges, clinical procedures, diagnoses, and time of death was retrieved from the patient administrative system by the Capital Region of Denmark Centre for Economy Data Unit. For time of death, only the date was available.

Data from the two sources were merged using the CRN (48). The CRN uniquely identifies all residents in Denmark. Non-residents admitted to a public Danish hospital have temporary numbers assigned.

Raw EWS data were received in Excel and .csv format. Upon receiving the data, each CRN was replaced with a serial number and removed from the dataset.

In the raw EWS data, one EWS record spanned multiple lines, with one line per EWS variable record and one line for each other data item associated with the record, such as a doctor’s note or nurse’s signature. Each record had an identifier, marking rows in raw data belonging to the same
record. Any point deductions from each EWS variable that had point deductions due to the CROS were also stored in a separate line and associated with the EWS record by the identifier. Raw data on admissions, discharges, clinical procedures, diagnoses, and dates of death were received in Statistical Analysis System format. Each row of data contained information on one ward admission, including dates of admissions, discharges, clinical procedures and diagnoses related to the ward admission, and dates of death, even if the death occurred after discharge. Ward admission to any ward within 24 hours of another ward admission was treated as one continuous hospital admission. This time limit was set as it reflects a clinical practice where re-admission within 24 hours is considered a continuation of the first admission.

From the raw EWS data was removed system test records, records with no timestamp, records not related to any admission, records from admissions of patients 15 years or younger, duplicate entries, records with only non-EWS information – such as a doctor’s note or nurse’s signature – and records with only artefact values. Artefact values were EWS variable values we found to be outright impossible (values of respiratory rate, arterial oxygen saturation, pulse rate, and systolic blood pressure with ciphers after the decimal point, supplemental oxygen above 30 L/min with ciphers after the decimal point, body temperature values with more than two ciphers after the decimal point, and supplemental oxygen values with more than two ciphers after the decimal point except 0.25, 0.75, 1.25, 1.75, 2.25 and 2.75 L/min).

Dates of death in patients who died and the date and time of any ICU admissions following EWS records were subsequently merged into the EWS dataset. The main components of a record in the final EWS dataset are shown in Table 2.

RStudio and the R open source statistical software package with the data.table package run on a standard laptop PC (Lenovo W540, Intel i7-4800MQ processor, 32 gigabytes of RAM) with a 64 bit version of Windows 7 were used for all data management and analyses (46),(49).

**End points**

We used 48-hour mortality and ICU admission within 48 hours of a EWS set as end points. Even though we were aware that we were lacking an optimal outcome, impairing our ability to draw strong conclusions regarding the clinical application of EWSs and EWS modifications, we found no better alternatives. We did proceed with Study 3 as we still found it relevant to calculate and report measures commonly used in similar studies, along with our interpretation of these results. As only date of death was available, a more precise definition of 48-hour mortality would be death on the same day as the EWS record was recorded or on the following day. On the death certificates, only date of death is a mandatory field; time of death is not. We had a dialogue with
the Statens Seruminstitut, which administered the death certificate registry, and unfortunately they were unable to provide complete data on time of death. Thus we had to use the abovementioned approximation of 48-hour mortality as the most reliable available measure for short term mortality.

Table 2: Main components of a record in the final EWS dataset

<table>
<thead>
<tr>
<th>Dataset variable (unit / category levels)</th>
<th>Variable type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory rate value (min⁻¹)</td>
<td>Integer</td>
</tr>
<tr>
<td>Respiratory rate points</td>
<td>Integer</td>
</tr>
<tr>
<td>Respiratory rate acceptable chronic value limit (min⁻¹)</td>
<td>Integer</td>
</tr>
<tr>
<td>Respiratory rate points after CROS point deduction</td>
<td>Integer</td>
</tr>
<tr>
<td>Arterial oxygen saturation value (%)</td>
<td>Integer</td>
</tr>
<tr>
<td>Arterial oxygen saturation points</td>
<td>Integer</td>
</tr>
<tr>
<td>Arterial oxygen saturation acceptable chronic value limit (%)</td>
<td>Integer</td>
</tr>
<tr>
<td>Arterial oxygen saturation points after CROS point deduction</td>
<td>Integer</td>
</tr>
<tr>
<td>Supplemental oxygen (L/min)</td>
<td>Double</td>
</tr>
<tr>
<td>Supplemental oxygen points</td>
<td>Integer</td>
</tr>
<tr>
<td>Supplemental oxygen acceptable chronic value limit (L/min)</td>
<td>Double</td>
</tr>
<tr>
<td>Supplemental oxygen points after CROS point deduction</td>
<td>Integer</td>
</tr>
<tr>
<td>Pulse rate value (min⁻¹)</td>
<td>Integer</td>
</tr>
<tr>
<td>Pulse rate points</td>
<td>Integer</td>
</tr>
<tr>
<td>Pulse rate acceptable chronic value limit (min⁻¹)</td>
<td>Integer</td>
</tr>
<tr>
<td>Pulse rate points after CROS point deduction</td>
<td>Integer</td>
</tr>
<tr>
<td>Systolic blood pressure value (mmHg)</td>
<td>Integer</td>
</tr>
<tr>
<td>Systolic blood pressure points</td>
<td>Integer</td>
</tr>
<tr>
<td>Systolic blood pressure acceptable chronic value limit (mmHg)</td>
<td>Integer</td>
</tr>
<tr>
<td>Systolic blood pressure points after CROS point deduction</td>
<td>Integer</td>
</tr>
<tr>
<td>Mental state (alert, verbal, pain, unresponsive)</td>
<td>Categorical</td>
</tr>
<tr>
<td>Mental state points</td>
<td>Integer</td>
</tr>
<tr>
<td>Mental state acceptable chronic value limit (alert, verbal, pain, unresponsive)</td>
<td>Categorical</td>
</tr>
<tr>
<td>Mental state points after CROS point deduction</td>
<td>Integer</td>
</tr>
<tr>
<td>Body temperature value (°C)</td>
<td>Double</td>
</tr>
<tr>
<td>Body temperature points</td>
<td>Integer</td>
</tr>
<tr>
<td>NEWS total score</td>
<td>Integer</td>
</tr>
<tr>
<td>CROS total score</td>
<td>Integer</td>
</tr>
<tr>
<td>Patient identifier, specific to this dataset (not CRN)</td>
<td>Integer</td>
</tr>
<tr>
<td>Time of EWS registration</td>
<td>Date / time</td>
</tr>
<tr>
<td>Admission time</td>
<td>Date / time</td>
</tr>
<tr>
<td>Discharge time</td>
<td>Date / time</td>
</tr>
<tr>
<td>Date of death</td>
<td>Date</td>
</tr>
<tr>
<td>ICU admission time</td>
<td>Date / time</td>
</tr>
<tr>
<td>Hospital identification code</td>
<td>String</td>
</tr>
<tr>
<td>Ward identification code</td>
<td>String</td>
</tr>
</tbody>
</table>
**Analyses**

Absolute numbers and percentages, and tables and frequency plots were used for descriptive analyses. Sensitivity, specificity, the PPV, the NPV, and the AUC were used to compare the NEWS with the CROS, CREWS, and S-NEWS. These were calculated using all complete NEWS records from all included patients.

We found sensitivities, specificities, PPVs, and NPVs for different cut-offs, and AUCs, to be relevant metrics for evaluating EWS performance. From the decision makers’ perspective, sensitivity is of particular interest as it describes how many of the patients with a certain condition are “discovered” by the instrument in question. Specificity is traditionally reported along with sensitivity, although in low-incidence environments it should be remembered that even a very high specificity can signify a rather large number of false positive records. PPVs and NPVs are relevant to both decision makers and clinicians; PPVs are particularly important in low-incidence environments. In the context of EWSs, the knowledge of PPVs of EWS total scores above different thresholds would help the clinician in interpreting individual total scores. The decision maker would gain a more informed basis for evaluating escalation protocols and strain on resources in relation to the number of patients relevantly identified. AUCs are relevant for the comparison of the discriminatory power of different models, such as NEWS and the NEWS modifications.

The percentages of records that were down- and upgraded and subsequently followed by 48-hour mortality or ICU admission were calculated. We used the terms “downgraded” and “upgraded” by the CROS, CREWS, or S-NEWS for records with a NEWS total score in one of the NEWS total score intervals of 0–5 points, 6–8 points, or 9–20 points, where application of the CROS, CREWS, or S-NEWS shifted the total score to another interval. In these three intervals, the Capital Region of Denmark EWS escalation protocol requires no mandatory doctor presence, mandatory doctor presence, or immediate doctor presence and specialist consultation, respectively. This analysis was chosen as we found it to be illustrative of the effect of the NEWS modifications on NEWS total scores.

**Approvals**

All studies were reported to the Capital Region of Denmark Ethics Committee. Ethics approval was waived. The data used in Studies 2 and 3 was information from the medical record. It was not available in the National Patient Registry or any other registry. Use of the data without obtaining informed consent from each patient required permission from the Danish Health and
Results

Study 1: End points for validating early warning scores in the context of rapid response systems: a Delphi consensus study

We identified and invited 22 persons for the expert panel used to explore current expert opinion on end points for validating EWS; 17 contributed to the study. A total of 86 suggestions for end points were made, which we arranged into 13 overarching themes. Consensus was not reached on any item. The themes containing the items with the highest median ranking were “cardiac arrest” and “death”. There were 233 comments made during the Delphi process. The least debated theme was “cardiac arrest”, with only 2 comments on the 6 items within this theme; 3 items where the general term “cardiac arrest” (median score 8) was more elaborately described (e.g. “cardiac arrest within 12, 24 and / or 48 hours of the recording of EWS”) had the higher median score of 9.

The theme “death” and the items in it received 33 comments. The item “death within 12, 24 and / or 48 hours of the recording of the EWS” (median score 9) had comments both challenging and supporting it: “death is a problematic outcome because of the large population of patients who do not want escalation of care”, and “we label an ‘unexpected death’ who have calling criteria not appropriately responded to within the 24 hours of death as ‘potentially preventable’. We believe that this is one of the more important hospital safety measures”. The theme “level of care” contained several items with a median rating of 8. In the comments, the theme was criticised on a general level for lacking generalisability between institutions due to the influence on this end point of ICU capacity and admission criteria, which differ between units.

Some experts argued in support of themes or items that had not achieved high median scores. An example is the theme “unexpected deviations from planned care”, where a comment on the general level was that “a serious adverse event is a deviation from planned care. I still rate this as important but I acknowledge that it may be hard to agree on objective measures for this.

However, for example, a patient who undergoes elective surgery has an expected trajectory. If anything deviates from this, it should be detected as early as possible and this must be the goal of any warning score...”
Study 2: A critical assessment of early warning score records in 168,000 patients

The 168,496 patients included accumulated 2,835,331 EWS records during 264,464 admissions. The EWS variables most frequently generating points were arterial oxygen saturation and pulse rate, which both generated points in close to 25% of records. One or more EWS variables were missing in 10% of records, resulting in incomplete EWS records where no total score was calculated. Temperature was the variable most frequently missing (Figure 4). Most incomplete records had recorded 5 or 6 of the 7 variables.

The percentages of incomplete records with a complete record within 12, 24, and 48 hours before or after the record were 88%, 95%, and 98%, respectively.

There were artefacts or extreme values in 0.2% of records; 0 was the most commonly recorded extreme value.

Plots of respiratory rate, oxygen supplementation, pulse rate, and systolic blood pressure showed preferences for certain values and terminal digits, and an accumulation of records just below 91 beats per minute was seen (Figure 5).

After study 2 was published, frequency plots of the recorded respiratory rates from complete NEWS sets from the 60 wards with the most NEWS records were made. The patterns of recorded respiratory rates varied between wards (Appendix 2). There was a marked variation in patterns from different wards, but no immediately obvious patterns between wards of similar specialties. The tendency towards a preference of recordings of numbers dividable by 2 in the range of 12 to 20 persisted.

---

**Figure 4: EWS variables recorded in incomplete records**

![Graph showing distribution of recorded parameters](image-url)
Figure 5: Distributions of recorded EWS variable values and NEWS total scores

- Respiratory rate
- Arterial oxygen saturation
- Supplemental oxygen
- Pulse rate
- Systolic blood pressure
- Mental state
- Body temperature
- NEWS total score
Study 3: Modifications of the National Early Warning Score in patients with chronic respiratory disease

Of the 168,496 patients with in-hospital NEWS records recorded during 2014, 11,414 patients with chronic respiratory disease were identified; 148 had no complete NEWS records and thus no NEWS total score and were excluded (Figure 3). We included 11,266 persons with chronic respiratory disease in the study; 37% of their 404,093 EWS records had point deductions due to the CROS. The median (IQR) deduction was 2 (1,3) points. Table 3 shows the percentages of records generating points for each EWS variable and the percentage of point-generating records where the points were overridden by the CROS.

When applying the CREWS, 33% of records had point deductions; when applying the S-NEWS, 20% of records had point deductions. With the S-NEWS, 15% of records had point additions. The CROS downgraded 52% from one of the total score intervals of 6–8 or 9–20 points to a lower total score interval. The CREWS downgraded 45%, and the S-NEWS 33%. The S-NEWS upgraded 5% of records with a NEWS total score of 0 to 8.

Of the records with a NEWS 0–5, 0.3% were followed by the patient dying within 48 hours of the record. Records that were downgraded from 6 or more points to 0–5 points, by any NEWS modification, were more frequently followed by 48-hour mortality or ICU admission than were the records with an unmodified NEWS total score in the same interval. The CREWS resulted in the fewest downgraded records being followed by 48-hour mortality; 1.9% of the records downgraded by the CREWS from a NEWS score of 6–8 to a total score of 0–5 were followed by 48-hour mortality; 2.2% of records downgraded from 6–8 points to 0–5 points by the CROS were followed by 48-hour mortality.

Table 3: Percentages of records where EWS points were generated, and percentages of point-generating records where the generated EWS points were overridden by the CROS

<table>
<thead>
<tr>
<th>Records generating points</th>
<th>Respiratory rate</th>
<th>Arterial oxygen saturation</th>
<th>Supplemental oxygen</th>
<th>Pulse rate</th>
<th>Systolic blood pressure</th>
<th>Mental state</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records generating points</td>
<td>24 %</td>
<td>50 %</td>
<td>45 %</td>
<td>33 %</td>
<td>20 %</td>
<td>2 %</td>
<td>8 %</td>
</tr>
<tr>
<td>Point-generating records where the points were overridden by the CROS</td>
<td>17 %</td>
<td>58 %</td>
<td>20 %</td>
<td>11 %</td>
<td>14 %</td>
<td>20 %</td>
<td>-</td>
</tr>
</tbody>
</table>
Application of the CROS, CREWS, and S-NEWS resulted in lower sensitivities and higher specificities and PPVs (Appendix 3).

Using 48-hour mortality as the end point, the AUCs (95% confidence intervals) for the NEWS, CROS, CREWS, and S-NEWS were 0.85 (0.85 – 0.86), 0.82 (0.82 – 0.83), 0.85 (0.84 – 0.85), and 0.84 (0.84 – 0.85), respectively. The AUCs (95% confidence intervals) using ICU admission as the end point were 0.79 (0.78 – 0.79), 0.81 (0.01 – 0.82), 0.81 (0.80 – 0.81), and 0.79 (0.78 – 0.80).

**Discussion**

The main objective of this thesis was to investigate the performance of an Early Warning Score (EWS) introduced in the Capital Region of Denmark. The pursuit of this objective led to three studies, which applied the Delphi consensus methodology and descriptive quantitative analyses. The main finding in Study 1 was that there was no consensus on any appropriate end point for the validation of the EWS. Items concerning death, ICU admission, and cardiac arrest were among those with the highest recorded median ratings.

The critical assessment of NEWS records accumulated in the Capital Region of Denmark during 2014 included 2,835,331 records from 264,464 admissions of 168,496 patients; 10% of records were incomplete. Digit preferences and distributional anomaly suggested that staff practice affected the recorded values.

There were 404,093 complete NEWS records from 25,978 admissions of 11,266 patients with a diagnosis of chronic respiratory disease. The CROS led to point deductions in 37% of records; 40% of patients had point deductions in one or more records. The S-NEWS, CREWS, and CROS downgraded between 1/2 to 1/3 of records from one of the “mandatory doctor presence” and “immediate doctor presence and specialist consultation” total score intervals to a lower total score interval. Sensitivities for the CROS, CREWS, and S-NEWS were lower than those for the NEWS. Specificities and PPVs were higher.

When using 48-hour mortality as the end point, the NEWS, CREWS, and S-NEWS had similar AUCs while the AUC of the CROS was lower. When using ICU admission as the end point, the AUCs for the CROS and CREWS were higher than those of the NEWS and S-NEWS.

**Rationale behind Study 1**

We found that there was a need to investigate the Capital Region of Denmark’s modification of the NEWS, i.e. how good this EWS is at identifying the correct patients, namely those at risk of
imminent critical deterioration. One of the key rationales behind EWSs, as mentioned in the EWS literature, is that unexpected in-hospital death, cardiac arrest, and ICU admission are frequently preceded by vital sign deviations (1),(2),(3),(50),(51). Accordingly, death, cardiac arrest, and unplanned ICU admission have been frequently used as end points in the EWS research. However, these end points identify a population with a high proportion of patients who cannot be saved, even with the best possible treatment. In the Royal College of Physicians’ NEWS recommendations, the stated aim is not to detect death, cardiac arrest, or ICU admission, but to detect general, clinical deterioration. Thus, it appears that there is a discrepancy between EWSs being validated for detecting death, cardiac arrest, and unplanned ICU admission and the general purpose clinical application of EWSs for detecting developing critical illness. The extrapolation of results from studies using death, cardiac arrest, and ICU admission as end points to an instrument for this different purpose requires that the physiological characteristics of the population of patients on a path to clinical deterioration, but who can be saved, are largely shared with those of the populations identified by the end points. That is, patients who are on a path towards critical illness, but who are salvageable with a timely and correct intervention, should have physiological deviations similar to those of a group of patients in which the majority of individuals are unsalvageable.

We found it remarkable that a system intended for use in practically all hospitalised patients several times per day relied on such an assumption.

By the nature of the problem identified, it was clear that we had to start by analysing the end points used to study EWSs; the end points suggested in a 2007 guideline for research on MET and RRS did not provide alternatives to the traditionally used end points (52).

**End points for validation of EWSs**

The end points favoured by the expert panel in the Delphi consensus study are in alignment with those prevalent in the EWS literature (16),(18),(25). Moreover, they are the ones used in recent studies on track-and-trigger systems other than the NEWS (53),(54),(55). Thus, the results of this study are in agreement with current practice.

The results from Study1 were not what we had hoped for as they did not provide an immediate solution to the problem of discrepancy between the end points used for validation and the clinical application of EWSs. Thus, it was necessary to accept that currently there are no good alternatives to the traditional end points. Consequently, we used these end points but realise that our results need cautious interpretation.
While not being recognised as currently useful by the expert panel, some of the 86 suggested items could be used as inspiration for the future development of end points for the validation of EWSs. The increasingly widespread use of EHRs containing all clinical data gathered during admissions and data science techniques may provide an option to define criteria reliably describing “a critically ill patient”, thus providing us with a more useful end point for the validation of EWSs.

The concrete items suggested and included in the Delphi questionnaire may not be directly applicable for this purpose. However, the items and comments in the themes of “unexpected deviations from planned care”, “physiology and EWS dynamics”, and “staff and patient satisfaction” could be useful as inspiration. Examples of unexpected deviations from planned care could be acute reoperations due to serious complications after surgery, hospital-acquired sepsis, or acute respiratory impairment in patients admitted for other reasons. The suggested item “progression to extreme physiology” could be relevant, as could PROMS.

**EWS data collected during clinical practice**

There are two components necessary to evaluate risk stratification models such as EWSs, namely the end point the model aims to identify and the input variables used to make the projection. Having studied the end points, we obtained, operationalised and analysed the input variable values.

The 10% of EWS records that were incomplete in our data are somewhat less than the 15% missing values found in a study of a paper-based EWS with 5 variables (respiratory rate, arterial oxygen saturation, pulse rate, systolic blood pressure, and body temperature) (56). It is much lower than the 35% of records where one of the variables respiratory rate, pulse rate, systolic blood pressure, or body temperature was missing in a study of records in another paper-based EWS during an outbreak of Legionnaires’ disease (57). Thus, it appears that electronic registration of the EWS encourages completeness of records compared with paper-based systems.

Our finding of artefacts or extreme values in 0.2% of records is in line with a report of 0.1% of records with entries somewhat similar to those we classify as artefacts or extreme values (58). Digit preferences or distributional anomalies similar to those found by us have not been described in other studies on EWSs. Most EWS studies do not describe the presence or absence of such features of the data used (16),(18),(25). Two studies, using data saved directly from bedside monitors without human interference, published EWS variable distributions (59),(60). Pulse rate was the only variable that showed distributional anomalies, and the anomalies
corresponded with common pacemaker settings. Accordingly, it appears most likely that the anomalies found in our EWS data are due to human factors. Preferences for certain terminal digits in manual measurements are well known (61). However, in current clinical practice in the Capital Region of Denmark, measurements of all EWS variables, other than respiratory rate and mental state, are performed by semi-automated devices. These show the result of the measurement in an electronic display. If the recorded EWS variable values were all correct readings from the semi-automated devices correctly entered into the EHR, no digit preferences or distributional anomalies should be seen. The digit preferences in staff’s recordings of respiratory rate were similar to those detected in a study of respiratory rates measured by an electronic device, ward staff, and a standardised, manual measurement (62). On a speculative level, an explanation for this could be that when staff has no concerns about a patient based on clinical judgement, they may enter an estimated value rather than spend time recording the actual respiratory rate. Another explanation could be that when measuring values just above a point-generation cut-off level, in some cases the staff round down the value. The recorded NEWS variable values could possibly in some cases be prone to recall bias due to the delay between measurement and recording of NEWS variable values. In some cases this could perhaps have caused regression towards the mean and in some cases perhaps an accentuation of an eventual nursing concern effect. Human factors effects such as those found in our data can be hypothesised to work both towards improving and weakening the ability of EWSs to predict a relevant end point. The first might be the case if the deviations are reflections of relevant, clinical considerations akin to an implicit indication of “nursing concern”. In some cases, a “standard” respiration rate of 16 could reflect “no nursing concern” rather than the patient’s actual respiration rate.

The supplementary analysis of patterns of recordings of respiratory rate in different wards is quite remarkable (Appendix 2). It shows a striking difference in the distributions of the recorded values for respiratory rate between different wards. While the presence of implausible values and distributions seemed to persist, it appeared somewhat more marked in some wards than others. This suggest that different clinical practice cultures may exist, and that human factors effects indeed appear to play a role in the recordings of respiratory rates in clinical practice.

Thus, results from validation of EWSs in a setting such as ours and the similar setting used for the main ViEWS and NEWS validation studies may not be transferrable to settings with fully automatic collection of EWS variable data. In such settings, an implicit indication of level of nursing concern or other sources of human factors effects would be lost. With devices for fully automated measurement and recording of EWS variables on general hospital wards being
actively marketed (63), the possible existence of such effects can soon become a concern in everyday clinical practice. With the patterns of recorded values varying between wards, such an effect may even be relevant to the current clinical application of EWSs.

**EWSs in patients with chronic respiratory disease**

Having analysed the two main data-based components of model evaluations, namely the input and the end points, we established the foundation for the analysis of how EWSs performed in the case of patients with chronic respiratory disease.

The CROS has been frequently applied in daily clinical practice. This supports the assumption that there is a need to address an issue of false alerts being generated by the NEWS. The percentages were high of records downgraded from the “mandatory doctor presence” and the “immediate doctor presence and specialist consultation” total score intervals with application of the CROS, CREWS, and S-NEWS. This finding must be considered in the light of staff adherence to the EWS protocol possibly being poor regarding both scoring frequency and clinical response (64).

Our finding of the CROS, CREWS, and S-NEWS all causing drops in the sensitivities for 48-hour mortality and ICU admissions compared with the NEWS indicates that the reduction in the number of alerts probably comes at a cost: we might overlook some critically ill patients. However, if the increased specificity reduces the number of false alerts, it could improve the treatment of patients at risk of critical illness. Staff resources would be available for other tasks than dealing with false alerts, and staff perception of the EWS and adherence to EWS protocols might improve. The findings of decreased sensitivity and increased specificity of the CREWS contrast with the statement that the CREWS improved specificity with unchanged sensitivity heralded by the study in which the CREWS was initially published (39). However, this study was somewhat small and differs with regard to the population studied. Further, it used 30-day mortality as the end point, so the results are not directly comparable with ours. A study including patients with acute exacerbation of COPD reported sensitivities for the CREWS and S-NEWS even lower than those found in our study, while those of the NEWS were comparable to those we found (40). Neither are results from this study directly comparable with our results, mainly due to the use of in-patient mortality as the end point, and differences in populations. We found that the AUC of the CROS and CREWS was above the AUC of the NEWS when applying ICU admission as the end point. In a comparison of the NEWS and 33 other EWSs in a population of general medical emergency patients using 24-hour mortality and ICU admission as the end points, the NEWS had the higher AUC irrespective of end point (25). Neither the CROS,
nor the CREWS were included in this study. Thus, our results show that the NEWS may not be the EWS with the best discrimination measured by the AUC using ICU admission as the end point in all populations, and further that the CROS and CREWS may have a better discrimination than the NEWS in patients with chronic respiratory disease. While the AUC, as described below, cannot be considered a reliable single measure of EWS performance, it has a central role in the EWS literature.

Methodological considerations

Study 1
To establish current expert opinion on the end points for the validation of the EWS in a structured manner, we opted for an approach based on the Delphi consensus methodology (45). The Delphi approach has the advantage of being asynchronous and geographically independent. As the experts in our study were spread across 3 continents, the ability to conduct the study without the experts needing to meet in person was paramount; this was the main reason for choosing this approach. The written communication used in the Delphi methodology may be a barrier to the free flow of ideas and views. This can be a disadvantage to the process of shared decision making (process loss). The oral communication used in the nominal group technique, which is an alternative to the Delphi methodology, does not have this weakness. Nevertheless, it necessitates the experts meeting in person. Another methodological concern in consensus methodology studies is the risk of certain views dominating not because they are logically most correct based on current knowledge, but because they are supported by strong individuals. The countermeasures against this are anonymization and carefully moderated communication processes, which are easier to apply in a Delphi process than in a nominal group process.

We defined objective criteria for the inclusion of experts. However, we still found that a large fraction of the experts included were part of the network around the Society for Rapid Response Systems (65). There were, however, also researchers who were not part of this network, and it was our impression that they did indeed have a different approach to the subject. We find that the objective criteria aided in ensuring representation of competent experts from different research groups, but we would have liked to have had less overrepresentation from those in the network around the Society for Rapid Response Systems. Perhaps it had been better to select criteria identifying a wider group of experts, or to identify a number of research groups working in the area, and then select an equal number of representatives from each.
Anonymization was kept throughout the study, reducing the potential effects of professional authority and personal ties. Criteria for consensus were predefined, reducing risk of bias towards a result showing consensus.

In the moderation of the process, we put the greatest emphasis on not interfering with the experts’ suggestions and comments to avoid our own views being reflected in the study. In hindsight we realise that it might have been better if we had intervened more in the moderation process. That would have allowed us to correct those experts who misinterpreted the question asked as being concerned with how to evaluate the degree of implementation of EWSs and RRSs in clinical practice at an early stage, rather than relying on the experts correcting each other during the Delphi process. Moreover, it would probably have reduced the number of items from the 86 that were included in the questionnaire. Some items were largely overlapping, and both this and the number of items might have been difficult for the experts to oversee. Although we received no complaints about this, it may still have reduced the response rate due to the size of the task of overseeing and commenting on an 86-item questionnaire.

When the study was designed, we chose to use a well-described methodology, and hence we chose the Delphi methodology. However, it is possible that we could have gotten better results from using a different approach, for instance by arranging an online consensus conference, or using means of real-time online communication methods in another, innovative way.

**Study 2**

The most challenging aspect of this observational study was selecting what to extract and present from a huge amount of data. The most laborious aspect was the data management process before beginning to analyse the data. The successful collection of a complete set of EWS-related records from a hospital system with near-complete coverage regarding serving the entire population of the greater Copenhagen area was both a prerequisite and a major strength of this study. One key element in the work on this study and the rest of the thesis is the power of the combination of RStudio, the R statistical software package, the data.table package, which is open source, and the large online R community. Another is the ability to handle this large amount of data through the entire data management and data analysis process on an off-the-shelf laptop. Although we have a large amount of data, we know little about the way the data were collected in clinical practice. It is our impression that the intended EWS data entry method in the Capital Region of Denmark is “computers on wheels” but that stationary computers in the ward workstations are frequently used. However, while we have not verified this, if this assumption is true, it is likely that the timestamps of many EWS records are unreliable (66),(67). The effect on
the results of the studies in this thesis is probably minor. As the EHR automatically calculated the EWS total scores in complete EWS records, we would not expect errors from calculations of total scores. These errors are prevalent in systems where total scores are calculated manually (56),(68). However, a study of a system with manually entered EWS variable measurements and automatic calculations of scores similar to the system in the Capital Region of Denmark still found errors in 12% of total scores (69). Although the nature and cause of these errors were unknown, they probably arose from human factors. Similar errors may be present in our data and could contribute to the digit preferences and distributional anomaly we found. In this light, our finding of artefact or extreme values in 0.2% of records appears to play a minor role.

**Study 3**

The main strength of this study was that the large EWS dataset provided a unique opportunity to study the CROS formalised EWS override system as well as the CREWS and S-NEWS in patients with respiratory disease. That we could merge EWS data with information on end points and diagnoses through the CRN and the high validity of the diagnoses studied were further strengths (47),(70). The population included in our study differs from those in prior studies of NEWS modifications for use in patients with chronic respiratory disease. One study included patients admitted to respiratory wards also caring for general medical patients (39). Patients who had a lower target saturation according to British Thoracic Society guidelines were classified as at risk for hypercapnic respiratory failure and followed the CREWS while other patients followed the unmodified NEWS (71). The other study included acutely admitted adult patients (40). COPD-patients were defined as patients with international classification of diseases 10 diagnoses J40 through J44. In analyses of the NEWS modifications these patients all followed the modified scoring tables, without attempts to identify patients with chronically lowered hypoxaemia. In our study, we sought to evaluate CROS and the other NEWS modifications when applied in clinical practice to patients with chronic respiratory disease regardless of aetiology across an entire hospital system. Hence, we found it relevant to include patients with chronic respiratory disease in general, rather than only those with acute exacerbation of COPD. Additionally, we found it relevant to include patients who were admitted primarily for other reasons than chronic respiratory disease but who had chronic respiratory disease as a secondary diagnosis, as chronic hypoxaemia could also be of clinical significance in these patients. As a definition of chronic respiratory disease, which has been validated in a Danish population, existed, we chose to use that definition (47). The rationale behind CROS, CREWS, and S-NEWS is that arterial oxygen
saturation NEWS point trigger levels are modified in patients with chronic hypoxaemia (and, in
the case of CROS, other chronically disturbed NEWS variables in patients with chronic disease).
In patients where affection of NEWS variable values are due to acute rather than chronic disease
the NEWS scoring chart rather than the modified scoring systems should be used. Accordingly,
we found it relevant to attempt to identify patients with chronic rather than acute hypoxaemia, as
was done in the primary CREWS study (39). Chronic hypoxaemia could be defined on the
presence of an acceptable chronic arterial oxygen saturation value. That is, however, less
standardised and less useful than the definition used in the primary CREWS study, which
referred to the use of the British Thoracic Society guidelines for evaluation of risk of
hypercapnic respiratory failure. These guidelines, however, also include individual judgement of
a patient’s condition, and we found our approach to be an acceptable compromise, while of
course less than ideal.
Thus, we based our inclusion criteria on what we found to be clinically relevant, rather than try
to strictly mimic those of earlier studies. While this reduces our ability to directly compare our
results with those of earlier studies, we cannot give precise indications of the degree of impact it
has on comparability. There were other significant differences between our study and earlier
studies of NEWS modifications for use in patients with chronic respiratory disease, e.g. the
inclusion of patient from specific wards rather vs. a whole hospitals system and in-hospital or
30-day mortality vs. approximated 48-hour mortality, requiring significant reservations with
regard to the comparability of our results with those of other studies to be taken. Consequently
we found that the advantage of attempting to align the population included in our study with that
of earlier studies in the area was outweighed by the disadvantages.
We included records from both primary and repeated admissions. This was done as we found it
relevant to reflect the clinicians’ and decision makers’ perspectives. Thus, the relevant questions
were along the lines of “what is the PPV for critical illness, based on this concrete EWS total
score that I have just been presented to?” and “what is the sensitivity of EWS total scores above
a certain level for all records made in my hospital?”.
It turned out to be difficult to define relevant end points for identifying critically ill patients,
including those with reversible conditions. Thus, we could calculate metrics for evaluation in the
context of prior and subsequent studies, but regrettably found the transferability of these results
to the current application of EWSs in clinical practice so questionable that we refrained from
drawing strong clinical conclusions.
We chose to use short-term outcomes analogue to those used in other major EWS studies
(18),(25), rather than 30-day or in-patient mortality used in the other studies focusing on patients
with chronic respiratory disease (39),(40). This choice was made as we found that 30-day or inpatient mortality would be more suitable for studying NEWS and the NEWS modifications as mortality predictors in what in an EWS context could be called a medium-term perspective than as indicators of imminent risk.

Concerning the metrics applied, the AUC (also known as the C-statistic) has been widely used in pioneering EWS studies (16),(18),(25),(72). However, its use for assessing risk stratification models generally and for evaluating the EWS specifically has been criticised (31),(32). Firstly, the AUC has no intuitive, clinical interpretation. The most direct interpretation of the AUC is the probability of that a randomly selected case received the higher EWS than a randomly selected control. When 48h mortality is the end point a case is a patient who dies within 48h after the EWS measurement and a control a patient who survives the first 48h after the EWS measurement. In a clinical situation with a single patient (not a pair of patients) discrimination ability is not the only criterion. It is relevant to know how likely a high EWS record is to indicate that this patient will die within 48h. While the description of the discriminative power of a model provided by the AUC can be relevant when comparing different models, it is of little value when evaluating the practical relevance of a model. It needs to be complemented by other measures.

Similarly, from a decision-maker’s perspective, sensitivity and specificity are relevant measures when evaluating risk stratification models. From a clinician’s or patient’s perspective, positive predictive values (PPVs) and negative predictive values (NPVs) have a direct interpretation. However, when the risk of the outcome in general is very low, as it is with 48h mortality, the PPVs of a model can be very low even when it has a good specificity. This could cause the model to appear favourable when it has good sensitivity and specificity from the decision maker’s perspective, but non-favourable from the clinician’s or patient’s perspective due to many false alerts. As EWSs operates in a low-prevalence environment, the combined evaluation of all four measures is highly relevant. However, as a consequence of the surrogate end points currently used in EWS research, the values of sensitivities, specificities, PPVs, and NPVs that are calculated are extremely difficult to interpret in a relevant manner. For instance, we have no way of translating the PPV of 5.7% of a NEWS total score ≥ 6 for 48-hour mortality in patients with chronic respiratory disease into a PPV for the short-term development of critical illness. This also applies to sensitivities, specificities, and NPVs. Additionally EWS treatment algorithms encourage interventions in the case of records above certain thresholds. If these interventions are successful, they could alter the outcome. If a patient with EWS variable values correctly indicating a trajectory towards an end point (e.g. death) is “salvaged”, what would have been a correct indication of a patient at risk (a true positive) in the calculations if there had been
no intervention, instead becomes a false indication of a patient being at risk (false positive). This would work towards misleadingly low sensitivities, specificities, PPV, and NPV. Thus, while we chose to calculate these measures as we found no better alternatives, they should be interpreted with the utmost caution. Although these are significant limitations, they are common in the EWS literature.

As EWSs can be seen as models ideally predicting patients’ risk of short-term critical deterioration, it could appear relevant to also assess calibration. However, there are several reasons why this is problematic in the case of EWSs. One concern is that it seems irrelevant to assess a model for the prediction of critical illness with a calibration against mortality or ICU-admission. Another is that EWSs do not associate a given score to an estimate of risk, and hence the “observed risk”-parameter for model calibration is not available. Yet another concern is that the incidence of the available endpoints of mortality and ICU-admission have very low incidence, and the observed risks will be very close to 0 for low scores.

Nevertheless, we considered calibration plots based on NEWS total scores, although for the reasons mentioned above, we have not included these in any work which we plan to submit for publication (Appendix 4). One set of plots for NEWS records from different age groups, and one set of plots for NEWS records from admissions with different primary diagnoses were made. Predicted risks were calculated for each NEWS total score category as the relative frequency of records belonging to patients who died within 48-hours. Note that model calibration is by definition perfect when calibration is assessed in the same data in which the predicted risks are calculated. However, calibration may not necessarily be perfect in NEWS records from different groups of patients. Risk may be underestimated in records from some groups, and overestimated in others. NEWS is used as a “one size fits all”-model, i.e. NEWS is applied to patients with a wide range of medical conditions and demographic characteristics. Thus, assessing calibration in different subgroups should yield similar results. There do, however, appear to be differences between groups. For instance, a high score carries a higher risk of death in older patients than in younger patients (e.g. between patients aged 40 – 49 years and 80 – 89 years, plots 1-8). Likewise, there appears to be differences between different medical conditions. For instance, records with a given NEWS total score from patients with a primary diagnosis of acute myocardial infarction (plot 58) and ileus (plot 63) indicate a higher risk of dying than records with the same total score from patients with acute exacerbation of COPD (plot 18). Thus, relying on EWSs for decisions concerning which clinical response is suitable may trigger different responses to patients with identical risks.
Implication for the use of EWSs in clinical practice

The signs of human factors influencing the recorded EWS variable values in the Capital Region of Denmark should not be interpreted as an expression of poor clinical practice. The practice may in some aspects be imperfect, but it is probably a practice similar to that of the main EWS validation studies, and while the methodology of those validation studies has shortcomings this is the practice for which the EWS is validated. It is, however, tempting to try to introduce fully automated measurements as a time-saving and perhaps quality-improving measure. However, this would mean that the foundation of the EWS, namely the measured variable values, would have different characteristics than in the setting in which it has been validated as it would be without the human-factor influence. Clinical application of fully automated EWS variable recordings could accordingly be considered an alteration to the foundation of the EWS and monitored and studied accordingly.

The sensitivities, specificities, PPVs, and NPVs presented in this thesis should be interpreted carefully, and the usefulness of the results of this thesis regarding making recommendations on the subject of EWSs in patients with chronic respiratory disease is thus limited. The extensive use of the CROS can be seen as a reflection of clinicians finding the EWS override option highly clinically relevant.

The shortcomings of the methodology available to EWS research uncovered in this thesis are perhaps the most striking overall finding, and they should be considered in relation to the use of EWSs in clinical practice. While these shortcomings rendered us unable to draw any strong conclusions on the performance of the EWS override system unique to the Capital Region of Denmark, they also apply to the main NEWS validation studies.

In clinical practice, the lack of a meaningful clinical interpretation of the EWS total scores (with or without the CROS) necessitates staff relying on the EWS escalation protocol when considering the clinical implication of any given EWS total score. However, without reliable estimates of parameters such as sensitivity, specificity, PPVs, and NPVs, or results of studies directly assessing the clinical impact of a given EWS implementation, the composition and evaluation of an escalation protocol becomes extremely challenging. The escalation protocols, with or without the CROS, thus must rely on judgements made based on incomplete evidence. This is a severe limitation that should be taken into account regarding use of EWSs.

Patients with treatment limitations

The issue of patients with treatment limitations is important, both when studying responses to elevated EWS total scores in the context of RRSs, and when considering end points for the
validation of EWSs. This is illustrated by death with and without limits on treatment being one of the most discussed subjects during the Delphi process in study one. One argument for including deaths in patients with limits on treatment is that a limit on treatment does not mean “it’s ok to crash”; hence these patients should be included in model validation studies as the model should be evaluated also on its ability to identify these patients. The argument that patients dying with limits on treatment “…are patients who are allowed to experience the physiologic winding-down known as death without a bunch of people trying to reverse the process…” can be seen as working both for and against the inclusion of these patients in EWS validation studies: If the purpose of EWSs is to identify patients who are about to die, i.e. EWS as a mortality prediction system, these patients should be included. If the purpose of EWSs is to identify patients with reversible conditions, deaths among patients who are on an irreversible end-of-life pathway should not be included in validations of EWSs using death as the end point. In the work on this thesis, we did not have the choice between the two approaches, as there was no practically possible way of reliably identifying patients with treatment limitations. However, treatment limitations have possibly had an influence on our results, which became apparent in the comparison of NEWS and CROS by AUCs in study 3. While NEWS does not include information concerning treatment limitations, CROS total scores may in some cases reflect patient status with regard to treatment limitations. It appears probable that in some cases of patients being on an end of life pathway with treatment limitations, clinicians may have applied acceptable chronic values, resulting in the CROS total score being lower than the NEWS total score. Thus, some patients with high NEWS total scores but limits to treatment deeming them not eligible for ICU admission could have had lower CROS total scores. This could contribute to the unexpected find of the AUC of CROS calculated using ICU admission as the outcome being higher than the AUC calculated using 48 hour mortality as the outcome. The mechanism involved could be that patients with treatment limitations and elevated NEWS total scores would frequently become “false positives” in validations using ICU admission as the outcome, as they would not be admitted to ICU. In validations using ICU admission, however, the lower CROS total score would cause them to more frequently become “true negatives”. Thus limits on treatment appear to influence results of EWS validation studies, and should therefore be considered in these.

**Perspectives on the end point problem**

A key concern of ours regarding the end points used in EWS validation studies is whether an ability of EWSs to identify patients at imminent risk of critical illness who can be saved with
timely intervention can be inferred from results concerning EWSs’ abilities to identify patients admitted to ICU, having cardiac arrests, or dying. As mentioned above, our initial concern was that the traditionally used end points identify a group of patients with a very high proportion of patients on an inevitable end-of-life pathway. A recent study of 358 patients with cardiac arrest found that 30-day survival was much lower for patients with a high NEWS less than 12 hours before the arrest than it was for patients with a low EWS; 47% of patients with a total NEWS score of 0-4 were alive 30 days after the cardiac arrest, compared to 20% of patients with a total NEWS of 5 or 6 or one single score of 3 points, and 10% of patients with a total NEWS score of 7 or more (73). One interpretation of these results is that elevated NEWSs indicate a particularly high risk, which could perhaps been reduced with timely intervention, in patients who experience cardiac arrests. Another interpretation could be that NEWS has a tendency to identify patients with a threatening cardiac arrest who cannot be saved, for instance due to low physiological reserves. Studies investigating the proportions of preventable cardiac arrests in patients who experience cardiac arrests with high versus low EWSs are warranted in order to elaborate on this. Another study including 245 persons with gram negative sepsis found MEWS to be a good predictor of mortality (74). MEWS on the day of onset of sepsis was significantly higher in patients who died within 28 days compared to those who survived (5.6 ± 2.6 vs. 3.4 ± 2.2, p=0.0001). Short term survival was also poorer in patients with a more markedly elevated MEWS. If this was due to delayed detection and treatment of the sepsis, it would support the idea that EWSs can save patients by detecting deterioration and getting treatment started earlier. However, this appears not to have been the case as MEWS on the day before the diagnosis of sepsis was unable to discriminate between those who survived and those who died (AUC 0.53 (0.42 – 0.64)). Additionally, the authors report that treatment was actually started earlier in the group with the highest MEWSs. A response with declining MEWS after the initiation of treatment was found to be a marker of a better prognosis. This could suggest that MEWS in this case shows a tendency towards being an indicator of low physiological reserves and poor prognosis rather than an indicator of level of risk of imminent critical illness in salvageable patients.

A recent study including patients with suspected infection compared different scores for use in non-ICU patients at the time of suspected infection (75). It used in-hospital mortality and ICU admission as outcomes, which with a median time to one of these outcomes of 14 hours (IQR 6 to 66 hours) were comparable to the end points traditionally used in EWS validation studies. It found that NEWS was more accurate for predicting in-hospital mortality and ICU admission than the quick Sepsis-related Organ Failure Assessment (qSOFA). Sensitivities and PPVs of NEWS
were excellent with a PPV of 23.6% at a sensitivity of 79.9% at NEWS ≥ 8. However, this could equally well be an indication of NEWS being good at detecting inevitably dying patients or hospital staff reacting to elevated NEWS total scores by admitting patients to intensive care as it could be an indication of NEWS being suitable for detecting risk of progression of a suspected infection to imminent, critical illness.

The calibration plots, which we made using 48-hour mortality as the end point (Appendix 4), highlight another paradox: If this end point were considered to be proportional to the risk of critical illness, the interpretation would be that a given NEWS total score can signify markedly different risks in different patient groups, making a one-size-fits-all action algorithm inappropriate. Thus, assuming that 48-hour mortality is a good proxy for critical illness in the evaluation of EWSs would lead to a conclusion of EWSs appearing unsuitable as one-size-fits-all instruments for the detection of imminent critical illness in hospitalised patients.

**Conclusions**

An expert panel was not able to achieve consensus on which end points are suitable for the validation of EWSs. End points related to death, cardiac arrest, and ICU admission were found currently to be the best compromises, while other suggestions made by the expert panel can possibly become relevant in time with the dissemination of advanced EHR and data science techniques.

Over 12 months, 168,000 patients had 2.8 million NEWS records made during admission to hospital in the Capital Region of Denmark; 10% of records were incomplete. Digit preferences and an accumulation of recorded values just below a cut-off for EWS point generation indicated that human factors are probably affecting the recorded EWS variable values in systems with manually collected, electronically recorded values.

The CROS was frequently used in patients with chronic respiratory disease. The CROS, CREWS, and S-NEWS all had reduced sensitivity compared with the NEWS for 48-hour mortality and ICU admission. Sensitivity, specificity, PPVs, and NPVs for 48-hour mortality and ICU admission were also affected. NEWS had a higher AUC than CROS with 48-hour mortality as the outcome, while CROS had a higher AUC than NEWS with ICU-admission as the outcome.

The combined result of the thesis regarding the investigation of the performance of the EWS in the Capital Region of Denmark was ambiguous. While it was found that currently there are no better alternatives to the methodology traditionally used in research on EWS performance, this
methodology was found to have significant shortcomings. Thus, it has not been possible to draw any strong conclusions on how the Capital Region of Denmark’s EWS performed regarding its stated aim of identifying developing critical illness in hospitalised patients.

**Future development**

With regard to the use of EWSs in the context of RRSs, the main issue is the lack of controlled studies investigating effects of introducing these systems on patient centred outcomes. In a Danish context this appears impossible to address as EWSs and RRSs are already in place across the healthcare system. Even if this had not been the case, vital signs would still have been measured, and staff should ideally always respond to deteriorating patients. Thus, in a controlled study with EWS / RRS in the intervention arm and usual practice in the control arm, the control group should be subjected to an educational effort towards the detection and management of at risk of similar proportions as the educational effort in the intervention group. Otherwise the study could risk becoming just as much a study of the effect of an information and education campaign directed towards the detection of and reaction to in-patient deterioration as a study of the effect of EWSs / RRSs.

Future suggested modifications and alterations of EWSs and RRSs can and should be evaluated by such randomised, controlled studies. These could use hard outcomes such as mortality, but could also focus on clinical responses to deteriorating patients. Such an evaluation of two different paediatric EWSs is underway in the Central Region of Denmark(76). The EWS research group in the Capital Region of Denmark is also planning a cluster randomised study of the currently used EWS with the CROS modification and a new EWS modification without the CROS, but with the option for staff to let their level of concern for the patient have substantial influence on the EWS total score.

While such studies are highly warranted in order to ensure that further EWS and RRS developments are based on high-level evidence, observational studies of EWSs and EWS modifications are also needed as a part of the future development of EWSs. EWSs should be optimised for the detection of developing critical illness in hospitalised patients with the highest possible sensitivities and PPVs. This will ensure that detection of deterioration is as reliable as possible and that the strain on resources and risk of alarm fatigue as low as possible.

The digit preferences found in recorded NEWS variable values and the variations in distributions of recorded values of respiratory rate between wards suggest that human factors indeed do play a role in the process of manually measuring and recording respiratory rates. This in itself is an
interesting finding, which could be relevant to explore further. It appears a relevant question if EWS total scores calculated using fully automated recordings of especially respiratory rate can be regarded as identical to scores calculated using manual measurements. If the manual measurements are influenced by human factors, EWS total scores collected by fully automated systems could be systematically different from EWS total scores collected using manual measurements. One hypothesis regarding human factors in play is that there in some cases could be an element of reflection of nursing concern, i.e. a respiratory rate of for instance 16 in some cases indicating “no nursing concern”. It could be relevant to investigate whether EWSs in general and fully automated EWSs where no indirect influence by staff concern appears possible in particular would benefit from adding an explicit “nursing concern” EWS variable. In RRSs with single parameter calling criteria as the afferent limb, clinical concern is a common trigger for MET calls, and it identifies a different population than objective calling criteria (77). In the NEWS recommendations as well as the Capital Region of Denmark action algorithms it is stated that clinical concern should always trigger a response, also in cases of low NEWS total scores. Thus clinical concern is actually already an “auxiliary afferent limb” in NEWS, although it is not an explicit NEWS variable, and it appears possible that including this parameter could improve EWSs, particularly EWSs based on fully automated EWS variable recordings.

Three possible pathways for future development of EWSs for the detection of in-patient deterioration can be outlined:

**Continue development using current methodology**

The “straightforward” path of EWS development would be to use the current methodology regarding end points and statistical methods to continue EWS perfection. Work along this path is already underway (54),(55). This would probably lead to models being developed on the basis of large datasets with a multitude of parameters from modern EHRs and a further increase in reported AUCs. However, as described above, this could result in the development of models optimised for identifying inevitably dying persons for whom nothing can be done beyond comfort care. While it is relevant to be able to identify these patients in order to ensure the best possible end-of-life care, EWSs are implemented in order to detect developing critical illness in patients who can benefit from timely interventions at an early stage, so further deterioration can be avoided. Thus, this path does not appear favourable.

**Establish better end points**

Based on the findings in this thesis, the culprit in the case of our inability to apply any analytical methods allowing us to draw clinically relevant conclusions with regard to EWS and the
detection of developing critical illness is the lack of suitable end points. Establishing better end points appears necessary in order to ensure that current and future EWSs reliably identify patients on a path towards imminent critical illness.

The problem faced here is akin to that of the International Classification of Diseases system lacking diagnostic codes related to human and system factors errors, making it impossible to assess key patient safety factors (78). The advent of diagnosis codes representing cases of unexpected deterioration would, if reliably reported, be exactly the end point that is missing. Such codes would mean that for instance cases of patients who were believed to be stable arriving at the operation ward with arterial oxygen saturation below 80 % due to respiratory failure could be recorded as episodes of unexpected critical illness. These could be used in EWS validation studies, and would mark a huge step forward to patient safety research. However, the introduction of such codes would require a large, focused effort by the political and administrative levels of the health care system, which is perhaps unrealistic.

Another option would be to obtain similar information from existing sources. Registries that could come to play a role are in existence, but they will need to be significantly improved with regard to coverage and / or structure to become useful. The Danish Patient Compensation Association database stores information about complications due to treatment. Adverse events are recorded only for patients filing claims and thus adverse events are probably massively underreported (79). The Danish Patient Safety Database has information that could be relevant to EWS research, but it does not store CRNs. Thus, its data cannot be linked to other data sources and consequently it is of little use to EWS research.

In the short term, EHRs and other existing sources of data are perhaps the most probable sources of alternative end points for use in EWS research. One approach could be to define a set of criteria for the conditions which EWSs are to identify. Such criteria could be based on a combination of many classes of EHR data and the current diagnostic classification. Some of the suggestions from the experts in study 1 could also be relevant, for instance based on items and comments in the themes of “unexpected deviations from planned care”, “physiology and EWS dynamics”, or “staff and patient satisfaction”.

The aim should be to design a highly composite end point, identifying conditions which are highly likely to lead to life-threatening or debilitating disease if left untreated. Such an end point could then be used to validate and compare current and future EWSs, whether they are simple and empirically derived such as NEWS or complex and developed using data science technologies.
PROMS, such as patients’ or relatives’ reports of unexpected and severe deterioration, could have a place as part of a composite outcome measure. Future EWSs could perhaps also benefit from using information retrieved directly from patients or as predictive variables.

**Adapt the use of EWS to what it actually is**

If the idea of better end points for EWS research is abandoned, and the methodological and interpretational problems with EWSs are deemed too severe, an alternative approach could be to adapt the way EWSs are applied in clinical practice. Inherently, EWSs are expressions of the combined deviation of a number of primary physiological variables from reference values. To “brand” EWSs as a measure of “general, physiological deviation” would simplify the interpretation of EWSs greatly, and as it directly reflects what EWSs are, no validation would be required. The interpretation would be even more straightforward if the score was distributionally derived, e.g. as in the centile based score assigned 3, 2, and 1 points to EWS variable records outside the 1%, 5%, and 10% quantiles, respectively (59).

The branding of the EWS equivalent in Region Zealand as the Basic Observation Score (BOS) could be a suitable approach. It indicates a score reflecting the status of basic physiological parameters, which should be understood and interpreted based on knowledge of the significance of these parameters. This could appear more appropriate than a branding indicating an ability to give early warnings, but where the clinicians do not know exactly what it warns against, or what level of risk a certain score represents.

**Conflicts of interest**

Niels Egholm Pedersen does not report any conflicts of interest. Neither do the academic advisors or co-authors of the papers included in the thesis.

**Source of funding**

The PhD was funded by Centre for Health and Centre for Human Resources, CAMES, the Capital Region of Denmark.
# Appendix 1: Escalation protocols

## The NEWS escalation protocol

<table>
<thead>
<tr>
<th>NEWS SCORE</th>
<th>FREQUENCY OF MONITORING</th>
<th>CLINICAL RESPONSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Minimum 12 hourly</td>
<td>- Continue routine NEWS monitoring with every set of observations.</td>
</tr>
<tr>
<td>Total: 1-4</td>
<td>Minimum 4-6 hourly</td>
<td>- Inform registered nurse who must assess the patient; - Registered nurse to decide whether increased frequency of monitoring and/or escalation of clinical care is required.</td>
</tr>
<tr>
<td>Total: 5 or more or 3 in one variable</td>
<td>Increased frequency to a minimum of 1 hourly</td>
<td>- Registered nurse to urgently inform the medical team caring for the patient; - Urgent assessment by a clinician with core competencies to assess acutely ill patients; - Clinical care in an environment with monitoring facilities.</td>
</tr>
<tr>
<td>Total: 7 or more</td>
<td>Continuous monitoring of vital signs</td>
<td>- Registered nurse to immediately inform the medical team caring for the patient – this should be at least at Specialist Registrar level; - Emergency assessment by a clinical team with critical care competencies, which also includes a practitioner/s with advanced airway skills; - Consider transfer of clinical care to a level 2 or 3 care facility, i.e. higher dependency or ITU.</td>
</tr>
</tbody>
</table>
The capitol Region of Denmark EWS escalation protocol effective in 2014 is shown below in English. Translated from Danish by the author. Original version follows.

<table>
<thead>
<tr>
<th>EWS score</th>
<th>Minimum observation interval</th>
<th>Escalation protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>Every 12 hours</td>
<td>Continue scoring every 12 hours. The frequency of scoring can be increased by doctor’s prescription.</td>
</tr>
<tr>
<td>2</td>
<td>Every 6 hours</td>
<td>Nursing staff optimise ABCDE * see pocket reference card. In the case of a single score of 2, the on-call doctor is notified. The nurse responsible for the patient’s care is notified before a doctor is called. The frequency of scoring can be increased.</td>
</tr>
<tr>
<td>3-5</td>
<td>Every 4 hours</td>
<td>The nurse optimises ABCDE * see pocket reference card. Doctor is notified. The doctor makes a plan and the frequency of scoring can be increased.</td>
</tr>
<tr>
<td>6</td>
<td>Every 4 hours</td>
<td>The nurse optimises ABCDE * see pocket reference card. Call for on-call doctor. The doctor attends the patient and makes a plan. The frequency of scoring can be increased.</td>
</tr>
<tr>
<td>7-8</td>
<td>Every 1 hours</td>
<td>The nurse optimises ABCDE * see pocket reference card. Call for on-call doctor. The doctor attends the patient within 30 minutes and makes a plan. The frequency of scoring can be increased. Think MET or assistance from anaesthesiologist.</td>
</tr>
<tr>
<td>9 or above</td>
<td>Every 30 minutes</td>
<td>The nurse optimises ABCDE * see pocket reference card. Call for on-call doctor. The doctor attends the patient within 15 minutes. The on-call doctor confers with a consultant doctor or MET / assistance for anaesthesiologist. The doctor makes a plan and the frequency of scoring can be increased.</td>
</tr>
</tbody>
</table>
The Capital Region of Denmark escalation protocol in Danish

The capital Region of Denmark EWS escalation protocol effective in 2014 is shown below in Danish. Available in English above.

<table>
<thead>
<tr>
<th>EWS score</th>
<th>Minimums observations-interval</th>
<th>Handlingsalgoritme</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>Hver 12. time</td>
<td>Fortsæt scoring hver 12. time Scoringshynnigheden kan øges efter lægeoordination</td>
</tr>
<tr>
<td>2</td>
<td>Hver 6.time</td>
<td>Plejepersonale ABCDE optimiserer * se lommekort Ved enkeltscore på 2 orienteres vagthavende læge. Sygeplejersinde med plejeansvar for patienten orienteres inden tilkald af læge. Scoringshynnigheden kan øges</td>
</tr>
<tr>
<td>3-5</td>
<td>Hver 4. time</td>
<td>Sygeplejersken ABCDE optimiserer * se lommekort Lægen underrettes. Lægen lægger en plan og scoringshynnigheden kan øges</td>
</tr>
<tr>
<td>6</td>
<td>Hver 4. time</td>
<td>Sygeplejersken ABCDE optimiserer * se lommekort Tilkald vagthavende læge Lægen tilser patienten og lægger en plan Scoringshynnigheden kan øges</td>
</tr>
<tr>
<td>7-8</td>
<td>Hver 1. time</td>
<td>Sygeplejersken ABCDE optimiserer * se lommekort Tilkald vagthavende læge Lægen tilser patienten indenfor 30 min. og lægger en plan Scoringshynnigheden kan øges Tænk MAT eller anæstesiologisk assistance</td>
</tr>
<tr>
<td>9 eller derover</td>
<td>Hver 30. minut</td>
<td>Sygeplejersken ABCDE optimiserer * se lommekort Patienten tilses af vagthavende læge indenfor 15 minutter Vagthavende læge konfererer med speciallæge eller MAT/anæstesiologisk assistance Lægen lægger en plan og scoringshynnigheden kan øges</td>
</tr>
</tbody>
</table>

Appendix 2: Recorded respiration rates by ward

On the next pages, frequency plots of the recorded respiratory rates from complete EWS sets from the 60 wards with the most EWS records can be found.
Ward specialty: Cardiology
N = 27,249

Ward specialty: Nefrology
N = 26,433

Ward specialty: Cardiology
N = 26,386

Ward specialty: Respiratory medicine
N = 26,090

Ward specialty: Surgery
N = 25,005

Ward specialty: Oncology
N = 23,997

Ward specialty: Endocrinology
N = 23,760

Ward specialty: Geriatrics
N = 23,460
Ward specialty: Medical
N = 15,645

Ward specialty: Medical admissions
N = 15,348

Ward specialty: Cardiology
N = 15,057

Ward specialty: Neurology
N = 14,926
Appendix 3: Sensitivities, specificities, PPV, and NPV

Sensitivities, specificities, PPVs, and NPVs for 48-hour mortality and ICU admission when applying the NEWS, CROS, CREWS, and S-NEWS in patients with chronic respiratory disease.

### 48-hour mortality as outcome

<table>
<thead>
<tr>
<th>Score</th>
<th>Measure</th>
<th>NEWS</th>
<th>CROS</th>
<th>CREWS</th>
<th>S-NEWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 1</td>
<td>N</td>
<td>337,949</td>
<td>29,3649</td>
<td>316,990</td>
<td>319,148</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>84 %</td>
<td>73 %</td>
<td>78 %</td>
<td>79 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>99.0 %</td>
<td>95.8 %</td>
<td>98.4 %</td>
<td>98.4 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>16.5 %</td>
<td>27.6 %</td>
<td>21.8 %</td>
<td>21.2 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>1.2 %</td>
<td>1.4 %</td>
<td>1.3 %</td>
<td>1.3 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.9 %</td>
<td>99.8 %</td>
<td>99.9 %</td>
<td>99.9 %</td>
</tr>
<tr>
<td>≥ 2</td>
<td>N</td>
<td>282,826</td>
<td>231,745</td>
<td>260,052</td>
<td>262,288</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>70 %</td>
<td>57 %</td>
<td>64 %</td>
<td>65 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>97.8 %</td>
<td>93.1 %</td>
<td>96.9 %</td>
<td>97.1 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>30.3 %</td>
<td>43.0 %</td>
<td>36.0 %</td>
<td>35.4 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>1.5 %</td>
<td>1.7 %</td>
<td>1.6 %</td>
<td>1.6 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.9 %</td>
<td>99.8 %</td>
<td>99.9 %</td>
<td>99.9 %</td>
</tr>
<tr>
<td>≥ 3</td>
<td>N</td>
<td>217,714</td>
<td>153,641</td>
<td>178,279</td>
<td>204,639</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>54 %</td>
<td>38 %</td>
<td>44 %</td>
<td>51 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>94.5 %</td>
<td>84.2 %</td>
<td>90.8 %</td>
<td>93.3 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>46.5 %</td>
<td>62.5 %</td>
<td>56.4 %</td>
<td>49.8 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>1.8 %</td>
<td>2.3 %</td>
<td>2.1 %</td>
<td>1.9 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.9 %</td>
<td>99.7 %</td>
<td>99.9 %</td>
<td>99.9 %</td>
</tr>
<tr>
<td>≥ 4</td>
<td>N</td>
<td>160,622</td>
<td>102,180</td>
<td>121,091</td>
<td>161,940</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>40 %</td>
<td>25 %</td>
<td>30 %</td>
<td>40 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>89.6 %</td>
<td>75.5 %</td>
<td>83.0 %</td>
<td>88.2 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>60.8 %</td>
<td>75.2 %</td>
<td>70.6 %</td>
<td>60.4 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>2.3 %</td>
<td>3.1 %</td>
<td>2.9 %</td>
<td>2.3 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.9 %</td>
<td>99.7 %</td>
<td>99.9 %</td>
<td>99.8 %</td>
</tr>
<tr>
<td>≥ 5</td>
<td>N</td>
<td>114,880</td>
<td>68,357</td>
<td>80,410</td>
<td>108,180</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>28 %</td>
<td>17 %</td>
<td>20 %</td>
<td>27 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>82.6 %</td>
<td>66.4 %</td>
<td>74.0 %</td>
<td>79.6 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>72.1 %</td>
<td>83.6 %</td>
<td>80.7 %</td>
<td>73.8 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>3.0 %</td>
<td>4.1 %</td>
<td>3.9 %</td>
<td>3.1 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.7 %</td>
<td>99.6 %</td>
<td>99.7 %</td>
<td>99.7 %</td>
</tr>
<tr>
<td>≥ 6</td>
<td>N</td>
<td>75,915</td>
<td>42,048</td>
<td>49,062</td>
<td>70,893</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>19 %</td>
<td>10 %</td>
<td>12 %</td>
<td>18 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>73.1 %</td>
<td>53.4 %</td>
<td>60.7 %</td>
<td>68.3 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>81.8 %</td>
<td>90.0 %</td>
<td>88.4 %</td>
<td>83.0 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>4.0 %</td>
<td>5.3 %</td>
<td>5.2 %</td>
<td>4.0 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.7 %</td>
<td>99.5 %</td>
<td>99.5 %</td>
<td>99.6 %</td>
</tr>
</tbody>
</table>

Table continues on next page

### ICU admission as outcome

<table>
<thead>
<tr>
<th>Score</th>
<th>Measure</th>
<th>NEWS</th>
<th>CROS</th>
<th>CREWS</th>
<th>S-NEWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 1</td>
<td>N</td>
<td>337,949</td>
<td>29,3649</td>
<td>316,990</td>
<td>319,148</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>84 %</td>
<td>73 %</td>
<td>78 %</td>
<td>79 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>97.1 %</td>
<td>95.0 %</td>
<td>96.4 %</td>
<td>96.6 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>16.5 %</td>
<td>27.6 %</td>
<td>21.8 %</td>
<td>21.2 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>1.4 %</td>
<td>1.6 %</td>
<td>1.5 %</td>
<td>1.5 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.8 %</td>
<td>99.8 %</td>
<td>99.8 %</td>
<td>99.8 %</td>
</tr>
<tr>
<td>≥ 2</td>
<td>N</td>
<td>282,826</td>
<td>231,745</td>
<td>260,052</td>
<td>262,288</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>70 %</td>
<td>57 %</td>
<td>64 %</td>
<td>65 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>94.5 %</td>
<td>91.7 %</td>
<td>93.6 %</td>
<td>93.8 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>30.3 %</td>
<td>43.1 %</td>
<td>36.0 %</td>
<td>35.4 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>1.6 %</td>
<td>1.9 %</td>
<td>1.7 %</td>
<td>1.7 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.8 %</td>
<td>99.8 %</td>
<td>99.8 %</td>
<td>99.8 %</td>
</tr>
<tr>
<td>≥ 3</td>
<td>N</td>
<td>217,714</td>
<td>153,641</td>
<td>178,279</td>
<td>204,639</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>54 %</td>
<td>38 %</td>
<td>44 %</td>
<td>51 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>88.5 %</td>
<td>83.1 %</td>
<td>85.4 %</td>
<td>87.7 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>46.5 %</td>
<td>62.5 %</td>
<td>56.4 %</td>
<td>49.8 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>2.0 %</td>
<td>2.6 %</td>
<td>2.3 %</td>
<td>2.1 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.7 %</td>
<td>99.7 %</td>
<td>99.7 %</td>
<td>99.7 %</td>
</tr>
<tr>
<td>≥ 4</td>
<td>N</td>
<td>160,622</td>
<td>102,180</td>
<td>121,091</td>
<td>161,940</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>40 %</td>
<td>25 %</td>
<td>30 %</td>
<td>40 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>81.3 %</td>
<td>74.5 %</td>
<td>76.8 %</td>
<td>81.0 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>60.8 %</td>
<td>75.3 %</td>
<td>70.6 %</td>
<td>60.4 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>2.4 %</td>
<td>3.5 %</td>
<td>3.1 %</td>
<td>2.4 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.6 %</td>
<td>99.6 %</td>
<td>99.6 %</td>
<td>99.6 %</td>
</tr>
<tr>
<td>≥ 5</td>
<td>N</td>
<td>114,880</td>
<td>68,357</td>
<td>80,410</td>
<td>108,180</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>28 %</td>
<td>17 %</td>
<td>20 %</td>
<td>27 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>71.6 %</td>
<td>64.6 %</td>
<td>66.3 %</td>
<td>70.7 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>72.1 %</td>
<td>83.7 %</td>
<td>80.7 %</td>
<td>73.8 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>3.0 %</td>
<td>4.6 %</td>
<td>4.0 %</td>
<td>3.2 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.5 %</td>
<td>99.5 %</td>
<td>99.5 %</td>
<td>99.5 %</td>
</tr>
<tr>
<td>≥ 6</td>
<td>N</td>
<td>75,915</td>
<td>42,048</td>
<td>49,062</td>
<td>70,893</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>19 %</td>
<td>10 %</td>
<td>12 %</td>
<td>18 %</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>60.7 %</td>
<td>52.4 %</td>
<td>54.1 %</td>
<td>59.1 %</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>81.7 %</td>
<td>90.1 %</td>
<td>88.4 %</td>
<td>83.0 %</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>3.9 %</td>
<td>6.0 %</td>
<td>5.3 %</td>
<td>4.0 %</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.4 %</td>
<td>99.4 %</td>
<td>99.4 %</td>
<td>99.4 %</td>
</tr>
</tbody>
</table>

Table continues on next page
<table>
<thead>
<tr>
<th>Score</th>
<th>Measure</th>
<th>NEWS</th>
<th>CROS</th>
<th>CREWS</th>
<th>S-NEWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 7</td>
<td>N</td>
<td>48,658</td>
<td>25,571</td>
<td>28,729</td>
<td>44,274</td>
</tr>
<tr>
<td>%</td>
<td>Sensitivity</td>
<td>12%</td>
<td>6%</td>
<td>7%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>61.7%</td>
<td>42.5%</td>
<td>47.8%</td>
<td>56.4%</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>5.3%</td>
<td>7.0%</td>
<td>7.0%</td>
<td>5.3%</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.5%</td>
<td>99.4%</td>
<td>99.4%</td>
<td>99.5%</td>
</tr>
<tr>
<td>≥ 8</td>
<td>N</td>
<td>29,711</td>
<td>15,527</td>
<td>16,745</td>
<td>25,420</td>
</tr>
<tr>
<td>%</td>
<td>Sensitivity</td>
<td>7%</td>
<td>4%</td>
<td>4%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>50.3%</td>
<td>33.2%</td>
<td>36.5%</td>
<td>43.5%</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>7.1%</td>
<td>9.0%</td>
<td>9.1%</td>
<td>7.2%</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.4%</td>
<td>99.3%</td>
<td>99.3%</td>
<td>99.4%</td>
</tr>
<tr>
<td>≥ 9</td>
<td>N</td>
<td>16,528</td>
<td>8,776</td>
<td>9,331</td>
<td>13,912</td>
</tr>
<tr>
<td>%</td>
<td>Sensitivity</td>
<td>4%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>36.9%</td>
<td>24.2%</td>
<td>26.1%</td>
<td>31.9%</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>9.4%</td>
<td>11.6%</td>
<td>11.7%</td>
<td>9.6%</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>99.3%</td>
<td>99.2%</td>
<td>99.2%</td>
<td>99.3%</td>
</tr>
</tbody>
</table>

### Appendix 4: Calibration plots

On the following pages are shown calibration plots of NEWS total scores from patients in different age groups and with different primary diagnoses. As 48-hour mortality was used as the end point, the plots can be seen as concerning NEWS regarded as model for the estimation of short-term risk of death. Records from admissions with one of the 60 most frequent primary diagnoses where at least one NEWS record was followed by 48-hour mortality was included. Predicted risks were calculated for each NEWS total score as the number of records with that score associated with 48-hour mortality divided by the total number of records with that score.
Calibration plots, different age groups

**Plot 1:**
Age: 20–29 years
N = 80,785, died: 8 (0%)

**Plot 2:**
Age: 30–39 years
N = 104,220, died: 62 (0.001%)

**Plot 3:**
Age: 40–49 years
N = 176,999, died: 209 (0.001%)

**Plot 4:**
Age: 50–59 years
N = 304,057, died: 797 (0.003%)

**Plot 5:**
Age: 60–69 years
N = 544,502, died: 2,891 (0.005%)

**Plot 6:**
Age: 70–79 years
N = 638,297, died: 4,910 (0.008%)

**Plot 7:**
Age: 80–89 years
N = 525,570, died: 6,746 (0.013%)

**Plot 8:**
Age: 90–99 years
N = 165,597, died: 2,763 (0.017%)
Calibration plots, different primary diagnoses (p. 1/7)

Plot 9: R10.0: Acute abdomen
N = 32,281, died: 202 (0.006%)

Plot 10: J18.9: Pneumonia, unspecified organism
N = 139,846, died: 2,027 (0.014%)

Plot 11: I48.9: Unspecified atrial fibrillation and atrial flutter
N = 49,053, died: 276 (0.006%)

Plot 12: Z03.9: Obs due to suspicion of disease or condition, unspecified
N = 84,520, died: 819 (0.01%)

Plot 13: Z03.4: Obs due to suspicion of myocardial infarction
N = 27,261, died: 56 (0.002%)

Plot 14: Z03.8: Encounter for observation for other suspected diseases and conditions ruled out
N = 55,413, died: 675 (0.012%)

Plot 15: Z03.5: Obs due to suspicion of other cardiovascular condition
N = 24,986, died: 166 (0.007%)

Plot 16: J96.0: Acute respiratory failure
N = 99,931, died: 2,539 (0.025%)
Calibration plots, different primary diagnoses (p. 2/7)

Plot 17: Z50.8: Encounter for other rehabilitation
N = 141,048, died: 269 (0.002%)

Plot 18: J44.1: Chronic obstructive pulmonary disease with (acute) exacerbation
N = 44,663, died: 454 (0.01%)

Plot 19: N39.0: Urinary tract infection, site not specified
N = 35,253, died: 179 (0.005%)

Plot 20: I63.9: Cerebral infarction, unspecified
N = 59,811, died: 413 (0.007%)

Plot 21: I50.9: Heart failure, unspecified
N = 43,453, died: 506 (0.012%)

Plot 22: I21.4: Non−ST elevation (NSTEMI) myocardial infarction
N = 30,648, died: 95 (0.003%)

Plot 23: C50.9: Malignant neoplasm of breast of unspecified site
N = 12,526, died: 86 (0.007%)

Plot 24: J15.9: Unspecified bacterial pneumonia
N = 47,706, died: 835 (0.018%)
Calibration plots, different primary diagnoses (p. 3/7)

Plot 25: J96.9: Respiratory failure, unspecified
N = 78,027, died: 1,320 (0.017%)

Plot 26: E86.9A: Dehydration
N = 28,003, died: 186 (0.007%)

Plot 27: I25.9: Chronic ischemic heart disease, unspecified
N = 22,710, died: 70 (0.003%)

Plot 28: I20.9: Angina pectoris, unspecified
N = 13,460, died: 15 (0.001%)

Plot 29: R10.4: Other or unspecified abdominal pain
N = 9,415, died: 61 (0.006%)

Plot 30: Z50.9: Encounter for rehabilitation, unspecified
N = 59,240, died: 192 (0.003%)

Plot 31: N30.0: Acute cystitis
N = 21,484, died: 53 (0.002%)

Plot 32: A41.9B: Urosepsis
N = 38,886, died: 371 (0.01%)

Predicted risk vs. Observed risk for different primary diagnoses.
Plot 33: K59.0: Constipation
N = 14,065, died: 53 (0.004%)

Plot 34: Z04.9: Encounter for examination and observation
for unspecified reason
N = 27,462, died: 216 (0.008%)

Plot 35: A46.9: Erysipelas, unspecified
N = 21,146, died: 35 (0.002%)

Plot 36: R07.4: Chest pain, unspecified
N = 6,257, died: 36 (0.006%)

Plot 37: C34.9: Malignant neoplasm of unspecified part of
bronchus or lung
N = 20,869, died: 238 (0.011%)

Plot 38: C61.9: Malignant neoplasm of prostate
N = 12,918, died: 152 (0.012%)

Plot 39: A41.9: Sepsis, unspecified organism
N = 46,022, died: 1,277 (0.028%)

Plot 40: J44.9: Chronic obstructive pulmonary disease,
unspecified
N = 20,581, died: 294 (0.014%)
Plot 41:  R29.8A: Symptom in the nervous system, unspecified  
N = 11,151, died: 37 (0.003%)  

Plot 42:  D64.9: Anemia, unspecified  
N = 14,370, died: 98 (0.007%)  

Plot 43:  R06.0: Dyspnea  
N = 15,521, died: 219 (0.014%)  

Plot 44:  Z03.1: Obs due to suspicion of cancer  
N = 14,845, died: 113 (0.008%)  

Plot 45:  A09.9: Gastroenteritis or colitis of unspecified cause  
N = 8,917, died: 28 (0.003%)  

Plot 46:  I20.0: Unstable angina  
N = 13,473, died: 5 (0%)  

Plot 47:  I10.9: Essential (primary) hypertension  
N = 8,322, died: 17 (0.002%)  

Plot 48:  G45.9: Transient cerebral ischemic attack, unspecified  
N = 9,757, died: 2 (0%)
Calibration plots, different primary diagnoses (p. 6/7)

Plot 49:  
R55.9: Syncope or collapse  
N = 7,800, died: 1 (0%)  

Plot 50:  
N18.9: Chronic kidney disease, unspecified  
N = 17,107, died: 70 (0.004%)  

Plot 51:  
R42.9: Vertigo, unspecified  
N = 7,424, died: 21 (0.003%)  

Plot 52:  
Z03.3: Obs due to suspicion of neurological disease  
N = 9,537, died: 28 (0.003%)  

Plot 53:  
S72.0: Fracture of head and neck of femur  
N = 37,647, died: 340 (0.009%)  

Plot 54:  
I35.0: Nonrheumatic aortic (valve) stenosis  
N = 14,050, died: 112 (0.008%)  

Plot 55:  
F10.2: Alcohol dependence  
N = 8,229, died: 10 (0.001%)  

Plot 56:  
K57.3: Diverticular disease of large intestine without perforation or abscess  
N = 7,622, died: 18 (0.002%)
Calibration plots, different primary diagnoses (p. 7/7)

Plot 57: C67.9: Malignant neoplasm of bladder, unspecified
N = 12,026, died: 65 (0.005%)

Plot 58: I21.3: ST elevation (STEMI) myocardial infarction of unspecified site
N = 8,266, died: 90 (0.011%)

Plot 59: J96.1: Chronic respiratory failure
N = 10,647, died: 241 (0.023%)

Plot 60: I21.9: Acute myocardial infarction, unspecified
N = 14,049, died: 204 (0.015%)

Plot 61: K40.9: Unilateral inguinal hernia, without obstruction or gangrene
N = 3,830, died: 6 (0.002%)

Plot 62: E86.9: Volume depletion, unspecified
N = 14,584, died: 72 (0.005%)

Plot 63: K56.7: Ileus, unspecified
N = 23,071, died: 296 (0.013%)
References


27. Time to Intervene? A review of patients who underwent cardiopulmonary resuscitation as a result of an in-hospital cardiorespiratory arrest. National Confidential Enquiry into Patient Outcome and Death; 2012


78. Makary MA, Daniel M. Medical error—the third leading cause of death in the US. BMJ. 2016;353:i2139.

The studies were included in the printed version of the thesis, but are not included here – they are available from the individual publishers, e.g. through PubMed. See page 7 for study titles and journals.