Applied clinimetrics in critical care

COLOFON
Applied clinimetrics in critical care
PhD Thesis, Tilburg University, Tilburg

Author	Saskia Rijkenberg
Cover	Wim Rijkenberg (*August 1, 1918 – † January 12, 1992*)
Lay-out	S. Rijkenberg
Print	Gildeprint

The printing of this thesis was financially supported by:
Stichting Invensief Intensive Care Unit OLVG, Stichting Wetenschappelijk Onderzoek OLVG

Applied clinimetrics in critical care

Proefschrift ter verkrijging van de graad van doctor
aan Tilburg University
op gezag van de rector magnificus,
prof. dr. E.H.L. Aarts,
in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie
in de aula van de Universiteit
op vrijdag 12 januari 2018 om 14.00 uur
door
Saskia Rijkenberg
geboren op 4 december 1977 te Zaandam

Promotiecommissie

Promotores:	prof. dr. P.H.J. van der Voort
	prof. dr. B.J.M. van der Meer
Overige leden:	prof. dr. K. Brinkman
	prof. dr. J. van der Palen
	prof. dr. A.M.G.A. de Smet
	prof. dr. H.C.W. de Vet
	dr. L. Vloet
	prof. dr. D.F. Zandstra

CONTENTS

CHAPTER I

INTRODUCTION

S. Rijkenberg

INTRODUCTION

Measurements are the backbone of intensive care (IC) medicine, as they are needed for diagnostic purposes, for continuous guidance of therapy and ultimately for prognostication. From admission to discharge or end of life care; all phases in the medical treatment of a critically ill patient require measurements. The centers for Medicare & Medicaid Services defined critical care and critical illness as

> "*the direct delivery by a physician(s) of medical care for a critically ill or critically injured patient. A critical illness or injury acutely impairs one or more vital organ systems such that there is a high probability of imminent or life-threatening deterioration in the patient's condition.*" [1]

Generally, in order to stabilize a critically ill patient's vital functions and prevent further clinical deteriorating, early intensive care unit (ICU) admission is often recommended [2]. However, IC resources are expensive and scarce, and therefore an appropriate utilization of ICU beds is crucial [3]. Recently, the process of early awareness and identification of patients who are eligible for admission to the ICU became more prominent in so called *track and trigger systems*. The calculation of these early warning scores is, among other parameters, based on the measurement of vital parameters such as heart rate, blood pressure, respiratory rate and level of consciousness [4].

Adequate treatment of critical illness and organ failure requires knowledge of the underlying cause(s). Measurements – for instance, laboratory tests, blood cultures and X-rays – form an important part of this diagnostic process [5].

Measurements are also used as part of the decision making process as to the medical interventions that are appropriate for the treatment, or to evaluate the effects of these interventions. Additionally, the consequences of critical illness and its medical treatment such as agitation, delirium, cardiac conduction disorders and hyperglycemia; require structural measurements, monitoring and treatment as well.

Discharge from the ICU to a lower level of care is recommended when a patient's physiologic status has stabilized and when there no longer is a need for advanced life-supportive care and monitoring [2;3]. Critically ill patients need continuous reevaluation to identify those who no longer need ICU care [3].

Unfortunately, cure is not always the result of advanced life-supportive care. Life supportive care ends when an irreversible life-threatening clinical situation occurs or when a patient refuses further treatment. In these situations, intensive care transitions into palliative care – the effective management of pain and discomfort, which also involves structural measurements. Since all aspects of critical care medicine include measurements, the quality of the measurement instruments is important. For critically ill patients, there are numerous clinical measurement instruments available, although several have been insufficiently validated.

It is therefore a challenge to find an appropriate instrument for a specific purpose [5]. But how do critical care physicians and nurses determine whether a measurement tool or device is appropriate, valid (accurate) and reliable for their clinical practice? The answer to this question can be found in the science of clinical epidemiology.

Clinical epidemiology and evidence-based practice

Clinical epidemiology is originated in the science of epidemiology; the study of the incidence, distribution and control of disease in specified human populations [6]. Clinical epidemiology applies epidemiological methods and principles in medicine and nursing in order to gather information to answer clinical questions for (individual) patients. It has the aim to develop and practice methods of clinical observation that will result in valid conclusions by avoiding systematic error and the influence of chance [6].

An approach to apply the science of clinical epidemiology in clinical practice is the use of evidence-based medicine/practice (EBP) [6]. The term EBP was introduced in the late 1980s at the McMaster Medical University in Canada and was used for a new teaching method in medicine. David Sackett et al. clarified its definition in 1996 [7].

> *''Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of Evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. By individual clinical expertise we mean the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice"* [8].

In the last two decades EBP has been widely spread among both the medical and nursing professions and provides a solid basis for clinical practice guidelines. It is generally believed that the application of EBP potentially leads to both an improved quality of care and cost reduction and is therefore internationally embraced by healthcare professionals [9;10]. Despite the positive attitude towards EBP, its implementation is still challenging and unsatisfactory since many barriers, such as a lack of time and skills, persist [10]. The ICU is an environment where a safe and effective treatment is highly important since critically ill patients are at high risk for complications, harm and death. In order to deliver this complex medical treatment, critical care physicians and nurses must be well aware of current clinical practice guidelines. Additionally, they have to be able to critically assess scientific research and choose the best available treatment in their clinical practice for the individual patient [11]. The application of EBP requires, among organizational facilitators, a positive attitude towards EBP as well as the necessary skills and knowledge of the methods of clinical epidemiology [10].

Clinimetrics

Clinimetrics is a discipline within clinical epidemiology that is dedicated to the quality of clinical measurements and was introduced by clinical epidemiologist Alvan R. Feinstein in 1987 [12]. Clinimetrics comprises the quality of clinical measurement instruments as well as the quality of the actual measurements [5]. The quality of a measurement tool depends on the psychometric properties of the tool, whereas the quality of the measurement itself depends on the patient, the assessor and the environment. Clinical measurements include a wide range of instruments such as scales to assess the quality of life or pain but also laboratory tests [7]. This thesis focuses on clinical measurements, which measure the health status of a patient at a single moment in order to discriminate between a changed health status (pain, prolongation of the interval between the Q wave and the T wave in the electrocardiogram (QTc) and dysglycemia).

Quality assessment of studies on measurement properties

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) aim to improve the selection of health measurement scales. The COSMIN-standards have been developed in an international Delphi study and can be used as a guideline to design studies that evaluate the psychometric properties of health measurement instruments. The COSMIN standards have been incorporated in a checklist for assessing the quality of studies on measurement properties (COSMIN checklist) and are summarized in the COSMIN-taxonomy and definitions [5;13-15].

The studies in this thesis are mostly designed according to the COSMIN-standards. Figure 1 presents an adapted version of the taxonomy. The original taxonomy of the COSMIN group contains three separate quality domains: reliability, validity and responsiveness [15]. The introduction of the term responsiveness has led to debate whether this is a separate domain or part of the domain validity and/or reliability. According to Streiner and Norman, responsiveness is considered as part of validity in this thesis. After all, the ability to measure change can be interpreted as a part of the construct validity [16]. Table 1 provides an overview of the definitions of clinimetric properties that are used in this thesis. The definitions that are used to assess the accuracy of continuous glucose measurement devices are summarized in chapter 4.2.

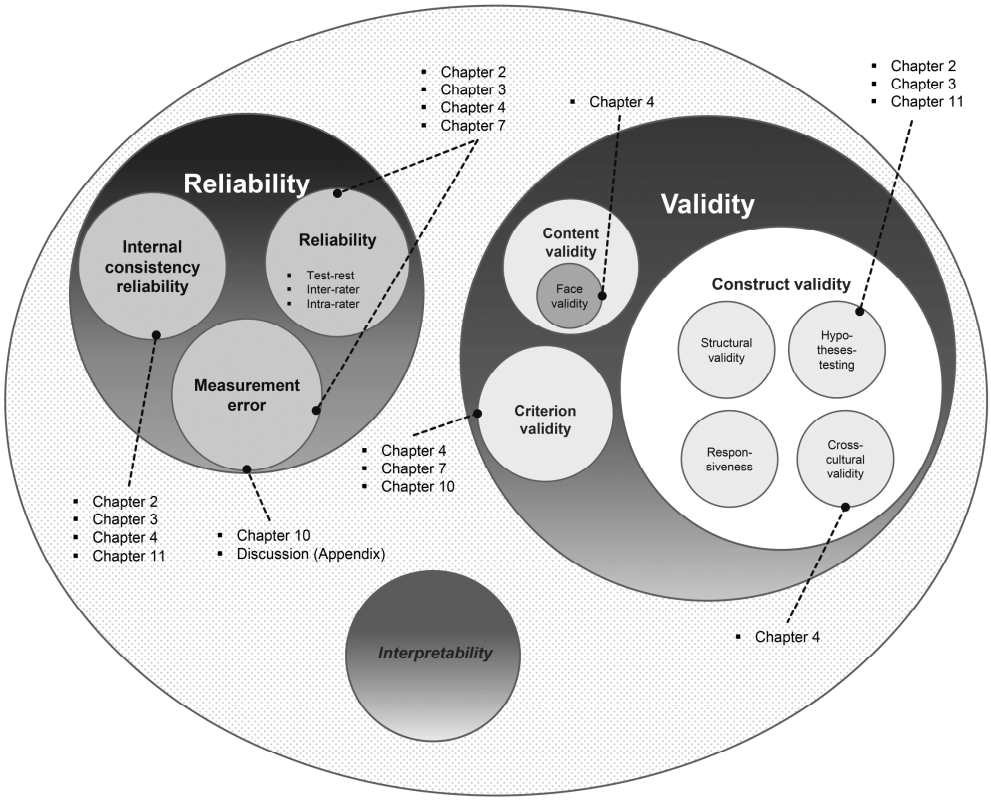


Figure 1. Quality of a measurement scale

Adapted from Mokkink et al. [14] with permission from the authors

2 Pain measurement in mechanically ventilated critically ill patients: Behavioral Pain Scale (BPS) versus Critical-Care Pain Observation Tool (CPOT)

3 Pain measurement in mechanically ventilated patients after cardiac surgery: Comparison of the Behavioral Pain Scale (BPS) and the Critical-Care Pain Observation Tool (CPOT)

4 Validation of the Dutch version of the Critical-Care Pain Observation Tool

7 Validation of continuous QTc measurement in critically ill patients

10 Accuracy and reliability of a continuous glucose measurement device in critically ill patients

11 Psychometric properties of the adapted McColl questionnaire: attitudes and knowledge towards evidence-based practice among nurses

Table 1. Clinimetric properties of measurement scales

PROPERTY	DEFINITION	
Reliability *(**Measurement error**)*		**Extent to which measurements obtained under different circumstances yield consistent and reproducible results**[1] ***Reflects the amount of random and systematic error in a measurement (not the result of differences between patients)***[1]
	Test-retest	Extent to which a scale produces reproducible scores over time (construct/patient is the same)[2]
	Intra-rater	Degree of agreement between measurements made by the same raters of the same construct (on different moments)[2]
	Inter-rater	Degree of agreement between two or more raters of the same construct/patient[2]. Proportion of the total variance in the measurements which is due to 'true'* differences between patients
Internal consistency		**Extent to which the items of a (sub)scale measure the same construct/ amount of correlation among the items of a (sub)scale**[3]

Table1 (*continued*)

PROPERTY	DEFINITION	
Validity		**Degree to which a scale measures the construct what was intented**[1]
Content validity (including *face validity*)		**Extent to which the items of a scale represent the construct to be measured**
Construct validity	Hypothesis testing	Extent to which the scores of a scale are consistent with hypotheses (including discriminant validity)
	Structural validity	Extent to which the scores of a scale are an adequate reflection of the dimensionality of the construct to be measured
	Cross-cultural validity	Extent to which the performance of the items on a translated/culturally adapted scale are an adequate reflection of the performance of the items of the original scale
	Responsiveness[1]	Extent to which the scale can measure a meaningful or clinically important change in a clinical state[1]
Criterion validity		**Extent to which the scores of a scale correlate with the 'gold standard'**
Interpretability[#]	Extent to which one can assign qualitative meaning to a scale's scores or change in scores	

Adapted from Mokkink 2010a with permission of the authors. Only definitions with a reference have been altered from the original definitions of Mokkink et al.

[1] Streiner and Norman [16]

[2] Gelinas et al. 2013 [17]

[3] Vink et al. 2017 [18]

* Strictly: the average score that would be obtained if the scale were given an in infinite number of times.

It refers only to the consistency of the score, not its accuracy. A true (high) pain score of an ICU patient hopefully changes after treatment with analgesics [16].

[#] Interpretability is not considered as a measurement property according to Mokkink, L.B. et al.

Table 2. Reliability coefficients

RELIABILITY	PARAMETER	DEFINITION		FORMULA
Inter-rater		**Proportion of the total variance in the measurements which is due to 'true'* differences between patients**		
	Intraclass correlation coefficient (ICC)	**One-way ANOVA**	**Model 1 (random)** - each subject is assessed by a different set of randomly selected raters from a larger population[1].	**ICC (1,1)** $\frac{BMS - WMS}{BMS + (k-1)WMS}$ *(single measures)*
				ICC (1,k) $\frac{BMS - WMS}{BMS}$ *(average measures)*
		Two-way ANOVA	**Model 2 (random)** – each subject is assessed by each rater. Raters have been randomly selected[1]. The random error and systematic error are both included in the denominator of the formula (*agreement*)[2].	**ICC (2,1)** $\frac{BMS - EMS}{BMS = (k-1)EMS + k(RMS – EMS)/n}$ *(single measures)*
				ICC (2,k) $\frac{BMS - EMS}{BMS + (RMS – EMS)/n}$ *(average measures)*

Table 2 (*continued*)

	Model 3 (mixed) – each subject is assessed by each rater. The raters are the only raters of interest (fixed)[1]. Only the random error is included in the denominator (*consistency*)[2].	**ICC (3,1)** $\frac{BMS - EMS}{BMS + (k-1)EMS}$ *(single measures)* **ICC (3,k)** $\frac{BMS - EMS}{BMS}$ *(average measures)*

[1] Shrout and Fleiss 1979 [35]

[2] de Vet et al. 2017 [36]

BMS = between-patient mean square (true patient variability)

WMS = within-patient mean square (measurement error)

BMS – WMS = variance due to differences between patients (σ^2 patients)

EMS = residual variance (random error variance), partly due to the unique combination of patients and raters [37]

RMS = variance due to systematic differences between raters (systematic error) [37]

k = number of raters

n = sample of n patients

Single measures = a single measurement is used to assess pain

Average measures = the average (mean) pain score of multiple pain assessments is used to assess pain

AIMS OF THIS THESIS

This thesis aims to assess the validity, reliability and accuracy of a number of clinical measurement instruments which; I. are recommended by clinical guidelines, II. are part of the standard patient monitor but not yet commonly used in critical care III. have potential benefits in safety, efficacy and cost effectiveness. How valid and reliable are measurement instruments used in intensive care medicine?

The aims of this thesis were fourfold:

1. To compare the clinimetric characteristics of clinical pain assessment tools for adult mechanically ventilated patients who cannot communicate
    - To translate the most appropriate clinical pain assessment tool into Dutch and perform a cross cultural validation
2. To assess the diagnostic accuracy of a continuous QTc measurement algorithm on a patient monitor in critically ill patients
    - To determine the frequency and variability of QTc interval prolongation in critically ill patients
3. To assess the accuracy and reliability of subcutaneous continuous glucose monitoring devices in critically ill patients
    - To assess the efficacy of a subcutaneous CGM system-guided blood glucose regulation in comparison with frequent Point of Care (POC) blood glucose-guided regulation in critically ill patients.
4. To assess the psychometric properties of the Dutch version of the McColl questionnaire concerning EBP among nurses

OUTLINE OF THIS THESIS

Part I. Pain measurement in mechanically ventilated critically ill patients

A patient's self-report of pain is considered as the gold standard in the assessment of pain [19]. Critically ill patients are often unable to communicate effectively due to severe illness and associated treatment. Clinical guidelines recommend the use of behavioral pain assessment instruments (the Behavioral Pain Scale (BPS) or Critical-Care Pain Observation (CPOT)) when critically ill mechanically ventilated patients are unable to self-report their pain [20;21]. Which pain assessment instrument; the BPS or the CPOT, has the best psychometric properties in critically ill mechanically ill patients? *(Chapter 2).* Which pain assessment instrument; the BPS or the CPOT, has the best psychometric properties in mechanically ventilated patients after elective cardiac surgery? *(Chapter 3)*. These chapters report the psychometric properties of the BPS and CPOT in a mixed ICU in a teaching hospital. The findings of this research resulted in a cross-cultural validation of the Dutch version of the CPOT *(Chapter 4)*. The presence of a delirium could interfere with the behavioral pain scores and therefore be a confounding factor. The study of Kanji et. al. addressed this problem and assessed the validity and reliability of the CPOT in adult critically ill patients with a delirium [22]. Can the CPOT be

used to assess pain in delirious ICU patients? The commentary in *Chapter 5* strives to answer this question.

Part II. Continuous QTc measurement in critically ill patients

Critically ill patients are at risk for prolongation of the interval between the Q wave and the T wave in the electrocardiogram (corrected QT [QTc]) [23-26]. It is well recognized that a prolonged ventricular repolarization, reflected on the electrocardiogram (ECG) as a prolonged QTc interval, is associated with ventricular arrhythmias [27;28]. The monitoring of the QTc interval in critically ill patients is usually performed by intermittent electrocardiogram (ECG), but it is unknown how frequently in between the ECGs a prolonged QTc occurs. Does intermittent QTc measurement underestimates the occurrence of QTc prolongation in critically ill patients? Recently, continuous QTc measurement has become available. The aim of *chapter 6* was to determine the frequency of QTc interval prolongation and the variability of QTc prolongation in critically ill patients. The software to continuously measure the QTc interval has been available since 2008. However, to our knowledge this continuous QTc measurement had not been validated yet in critically ill patients. *Chapter 7* reports elements of the diagnostic accuracy of continuous QTc measurement in critically ill patients.

Part III. Subcutaneous continuous glucose measurement in critically ill patients

Stress-induced hyperglycemia occurs in more than 90% of critically ill patients and is related to adverse outcomes (29). Recent guidelines recommend that critically ill patients with hyperglycemia should be monitored and treated with a target range of 100 to 150 mg/dL [30] or 140 to 180 mg/dL [31]. At the same time, hypoglycemia and high glucose variability should be avoided because it is also associated with adverse outcomes. Currently, glucose control in the ICU is mostly based on intermittent measurements with handheld meters for point-of-care (POC) glucose testing. These glucose measurements are used to guide intravenous insulin administration. Consequently, the management of glycemic control in critically ill patients requires frequent glucose monitoring, which is associated with costs [32;33]. Subcutaneous continuous glucose monitoring (CGM) may have benefits in achieving glycemic control in critically ill patients. It might detect glucose fluctuations and hypoglycemic events earlier because it provides insight into glucose trends. In addition, a reduction in nurse workload could be achieved. Although CGM devices have been evaluated in critically ill patients over almost ten years, they are still not widely used in daily clinical practice. This could be due to a considerable variation in accuracy and reliability among CGM devices and studies [34]. *Chapter 9* presents an overview of the evidence regarding the clinical benefits and accuracy of CGM devices in critically ill patients. *Chapter 8 and 10* show the results of an assessment of the accuracy, efficacy, safety and cost effectiveness of a subcutaneous continuous glucose monitoring device in critically ill patients.

Part IV. Evidence-based practice in intensive care medicine

Evidence-based practice (EBP) is an approach to optimize clinical decision-making by using the best external evidence in combination with a clinician's expertise and patient preferences. Practicing EBP means life-long learning and includes five steps I. convert clinical uncertainty and practice variation into an answerable clinical question II. find the best scientific evidence III. critically appraise the evidence IV. apply the results of the critical appraisal V. evaluate the implementation of the evidence [8]. Applying EBP is necessary to determine which measurement instrument or diagnostic instrument is appropriate, valid and reliable, for example, for the assessment of pain or glycemic control in a critically ill population. This requires, among other things, education, knowledge and a positive attitude towards EBP among healthcare professionals.

The implementation of EBP is still challenging and unsatisfactory since many barriers such as a lack of time and skills persist [10]. In contrast to medicine, nursing is not an academic discipline. The majority of current nurses have not received scientific education during their nursing training although EBP has been welcomed in the nursing curriculum. Custom-made EBP implementation programs might improve EBP knowledge, skills and attitudes towards EBP in a clinical setting. A valid and reliable assessment of attitudes towards EBP, current EBP utilization and EBP knowledge is a first step to enable the design and evaluation of an EBP implementation program [10]. The McColl Questionnaire has been developed to measure attitudes towards EBP; the ability to access scientific literature; self-rated knowledge of EBP journals, websites, and the terms used in this literature [24;25]. *Chapter 11* presents the results of an analysis of the construct validity and reliability of the Dutch version of the McColl questionnaire among nurses in a teaching hospital.

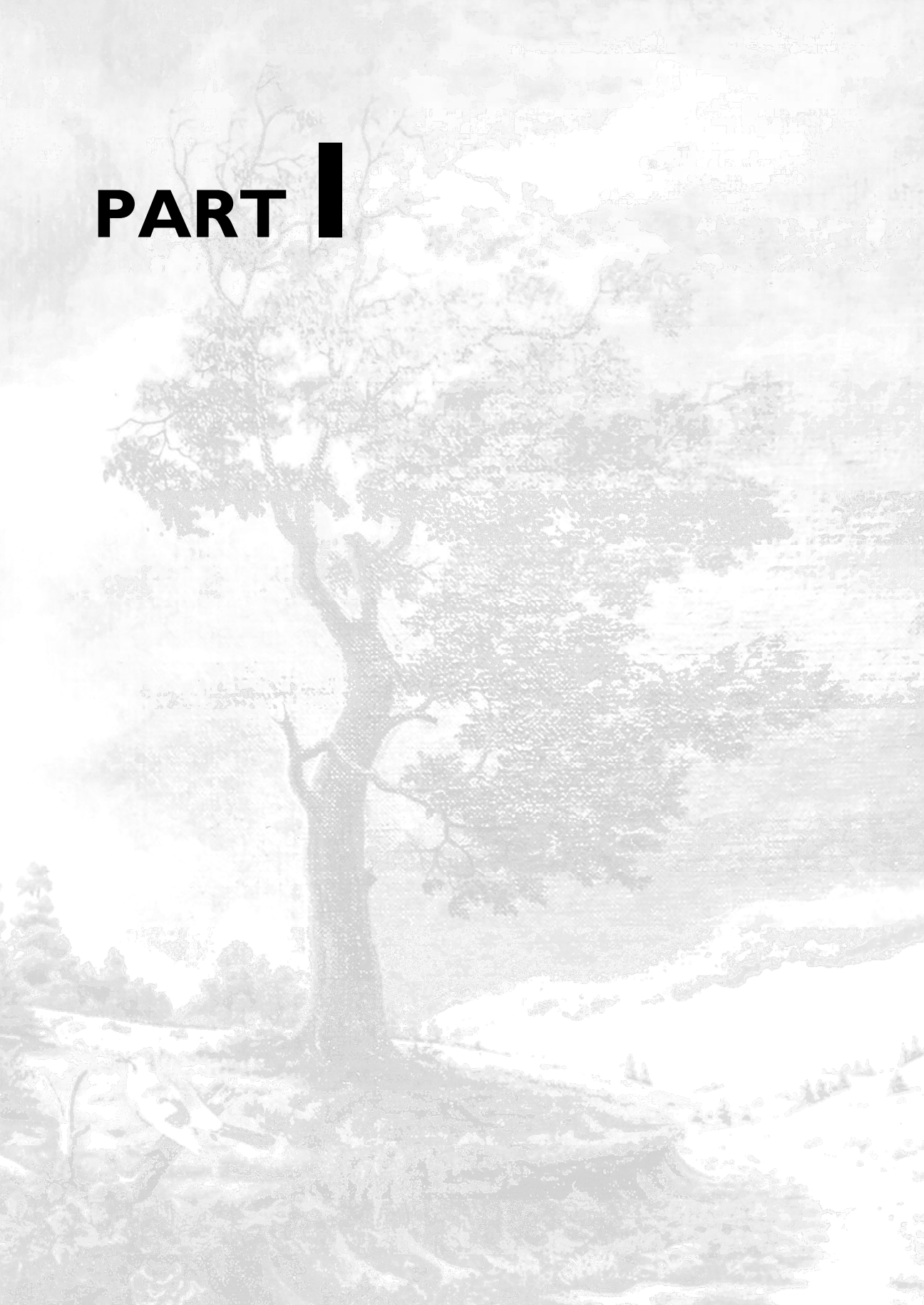
In conclusion, this thesis seeks to provide insight into the psychometric properties and applicability of a number of measurement instruments used in intensive care medicine.

REFERENCES

1. Physicians/Nonphysicians Practitioners. Centers for Medicare & Medicaid Services: Medicare Claims Processing Manual.Baltimore, MD, Centers for Medicare & Medicaid Services; 2013.
2. Criteria voor opname en ontslag van Intensive Care afdelingen in Nederland. Revisie van de richtlijn van december 2000. 2011. Commisssie richtlijnontwikkeling van de Nederlandse Vereniging voor Intensive Care.
3. Nates JL, Nunnally M, Kleinpell R, Blosser S, Goldner J, Birriel B, et al. ICU Admission, Discharge, and Triage Guidelines: A Framework to Enhance Clinical Operations, Development of Institutional Policies, and Further Research. Crit Care Med 2016 Aug;44(8):1553-602.
4. Spagnolli W, Rigoni M, Torri E, Cozzio S, Vettorato E, Nollo G. Application of the National Early Warning Score (NEWS) as a stratification tool on admission in an Italian acute medical ward: A perspective study. Int J Clin Pract 2017 Mar;71(3-4).
5. De Vet H, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. 1 ed. Cambridge: Cambridge University Press; 2011.
6. Fletcher RH, Fletcher SW. Clinical Epidemiology The Essentials. Fourth edition ed. Philadelphia: Lippincott Williams & Wilkins; 2005.
7. Inleiding in evidence-based medicine. Vierde druk ed. Houten: Bohn Stafleu van Loghum; 2014.
8. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-2.
9. Flodgren G, Rojas-Reyes MX, Cole N, Foxcroft DR. Effectiveness of organisational infrastructures to promote evidence-based nursing practice. Cochrane Database Syst Rev 2012 Feb 15;(2):CD002212.
10. Ubbink DT, Guyatt GH, Vermeulen H. Framework of policy recommendations for implementation of evidence-based practice: a systematic scoping review. BMJ Open 2013 Jan 24;3(1).
11. Phillips C. Relationships between duration of practice, educational level, and perception of barriers to implement evidence-based practice among critical care nurses. Int J Evid Based Healthc 2015 Dec;13(4):224-32.
12. Streiner D L, Norman G R. Selecting the items. In: Health measurement scales a practical guide to their development and use. third ed. New York: Oxford University Press; 2003. p. 61-79.
13. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May;21(4):651-7.
14. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-45.
15. COSMIN. www.cosmin.nl Accessed: 1-7-2017.
16. Streiner DL, Norman GR. Validity. In: Health measurment scales a practical guide to their development and use. third edition ed. New York: Oxford University Press; 2002. p. 172-193.
17. Gelinas C, Puntillo KA, Joffe AM, Barr J. A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Seminars in Respiratory and Critical Care Medicine 2013 Apr;34(2):153-68.
18. Vink P, Lucas C, Maaskant JM, van Erp WS, Lindeboom R, Vermeulen H. Clinimetric properties of the Nociception Coma Scale (-Revised): A systematic review. Eur J Pain 2017 Jun 2.
19. Chanques G, Viel E, Constantin JM, Jung B, de Lattre S, Carr J, et al. The measurement of pain in intensive care unit: comparison of 5 self-report intensity scales. Pain 2010 Dec;151(3):711-21.
20. Baron R, Binder A, Biniek R, Braune S, Buerkle H, Dall P, et al. Evidence and consensus based guideline for the management of delirium, analgesia, and sedation in intensive care medicine. Revision 2015 (DAS-Guideline 2015) - short version. Ger Med Sci 2015;13:Doc19.
21. Barr J, Fraser GL, Puntillo K, Ely EW, Gelinas C, Dasta JF, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306.

22. Kanji S, MacPhee H, Singh A, Johanson C, Fairbairn J, Lloyd T, et al. Validation of the Critical Care Pain Observation Tool in Critically Ill Patients With Delirium: A Prospective Cohort Study. Crit Care Med 2016 Jan 16.
23. Drew BJ, Ackerman MJ, Funk M, Gibler WB, Kligfield P, Menon V, et al. Prevention of torsade de pointes in hospital settings: a scientific statement from the American Heart Association and the American College of Cardiology Foundation. Circulation 2010 Mar 2;121(8):1047-60.
24. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. BMJ 1998 Jan 31;316(7128):361-5.
25. Oude Rengerink K, Zwolsman SE, Ubbink DT, Mol BW, van DN, Vermeulen H. Tools to assess evidence-based practice behaviour among healthcare professionals. Evid Based Med 2013 Aug;18(4):129-38.
26. Pickham D, Helfenbein E, Shinn JA, Chan G, Funk M, Weinacker A, et al. High prevalence of corrected QT interval prolongation in acutely ill patients is associated with mortality: results of the QT in Practice (QTIP) Study. Crit Care Med 2012 Feb;40(2):394-9.
27. El-Sherif N, Turitto G. Torsade de pointes. Curr Opin Cardiol 2003 Jan;18(1):6-13.
28. Yap YG, Camm AJ. Drug induced QT prolongation and torsades de pointes. Heart 2003 Nov;89(11):1363-72.
29. Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, et al. Intensive insulin therapy in critically ill patients. N Engl J Med 2001 Nov 8;345(19):1359-67.
30. Jacobi J, Bircher N, Krinsley J, Agus M, Braithwaite SS, Deutschman C, et al. Guidelines for the use of an insulin infusion for the management of hyperglycemia in critically ill patients. Crit Care Med 2012 Dec;40(12):3251-76.
31. Marathe PH, Gao HX, Close KL. American Diabetes Association Standards of Medical Care in Diabetes 2017. J Diabetes 2017 Apr;9(4):320-4.
32. Aragon D. Evaluation of nursing work effort and perceptions about blood glucose testing in tight glycemic control. Am J Crit Care 2006 Jul;15(4):370-7.
33. Gartemann J, Caffrey E, Hadker N, Crean S, Creed GM, Rausch C. Nurse workload in implementing a tight glycaemic control protocol in a UK hospital: a pilot time-in-motion study. Nurs Crit Care 2012 Nov;17(6):279-84.
34. Wollersheim T, Engelhardt LJ, Pachulla J, Moergeli R, Koch S, Spies C, et al. Accuracy, reliability, feasibility and nurse acceptance of a subcutaneous continuous glucose management system in critically ill patients: a prospective clinical trial. Ann Intensive Care 2016 Dec;6(1):70.
35. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86:420-42
36. de Vet HCW, Mokkink LB, Mosmuller DG, Terwee CB. Spearman-Brown prophecy formula and Cronbach's alpha: different faces of reliability and oppurtunities for new applications. J Clin Epidemiol.2017; 45-49.
37. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability. In: Measurement in Medicine. 1 ed. Cambridge: Cambridge University Press; 2011. p. 104

PART I



Pain measurement in mechanically ventilated critically ill patients

CHAPTER 2

Pain measurement in mechanically ventilated critically ill patients: Behavioral Pain Scale (BPS) versus Critical-Care Pain Observation Tool (CPOT)

S. Rijkenberg, W. Stilma, H. Endeman, R.J. Bosman, H.M.O. Oudemans-van Straaten

This chapter in adapted form has been published in: Rijkenberg et al. Journal of critical care, 2015, 30(1):167-72

CHAPTER 3

Pain measurement in mechanically ventilated patients after cardiac surgery: Comparison of the Behavioral Pain Scale (BPS) and the Critical-Care Pain Observation Tool (CPOT)

S. Rijkenberg, W. Stilma, R.J. Bosman, N.J. van der Meer, P.H.J. van der Voort

This chapter in adapted form has been published in: Rijkenberg et al. Journal of Cardiothoracic and Vascular Anesthesia, 2017, Aug: 31(4): 1227-1234

CHAPTER 4

Validation of the Dutch version of the Critical-Care Pain Observation Tool

W. Stilma, **S.Rijkenberg**, H.M. Feijen, J.M. Maaskant, H. Endeman

This chapter in adapted form has been published in: Stilma et al. Nursing Critical Care, 2015, Dec 22.

CHAPTER 5

Can the Critical-Care Pain Observation Tool (CPOT) be used to assess pain in delirious ICU patients?

S. Rijkenberg, P.H.J. van der Voort

Rijkenberg et al. Journal of Thoracic Disease, 2016, 8(5)

There has been no financial support for these studies

CHAPTER 2

Pain measurement in mechanically ventilated critically ill patients: Behavioral Pain Scale (BPS) versus Critical-Care Pain Observation Tool (CPOT)

ABSTRACT

Introduction

The Behavioral Pain Scale (BPS) and Critical-Care Pain Observation Tool (CPOT) are behavioral pain assessment tools for uncommunicative and sedated ICU patients. This study compares the discriminant validation and reliability of the CPOT and the BPS simultaneously, in mechanically ventilated patients on a mixed adult ICU.

Materials and Methods

This is a prospective observational cohort study in 68 mechanically ventilated medical ICU patients who were unable to report pain.

Results

The BPS and CPOT scores showed a significant increase of two points between rest and the painful procedure (turning). The median BPS scores between rest and the non-painful procedure (oral care) showed a significant increase of one point, whereas the median CPOT score remained unchanged. The inter-rater reliability of the BPS and CPOT scores during turning was 0.60 and 0.62, respectively. The overall inter-rater reliability was 0.74 and 0.75 respectively.

Conclusions

The present study showed that the BPS and CPOT are reliable and valid for use in a daily clinical setting. Although both scores increased with a presumed painful stimulus, the discriminant validation of the BPS use was less supported as it increased during a non-painful stimulus. The CPOT appears preferable in this particular group of patients, especially with regard to its discriminant validation.

INTRODUCTION

Critically ill patients frequently experience pain and discomfort during ICU stay. Approximately 50% of the patients reported moderate to severe pain, both at rest and during routine procedures [1-5]. Untreated acute pain in adult ICU patients has short- and long-term physiological and psychological consequences such as postoperative myocardial infarction, insufficient sleep and the risk of developing a posttraumatic stress disorder. The consequences of inadequate control of pain are significant, but excessive use of analgesics and sedation can lead to unwanted side effects such as hypoventilation, gastrointestinal hypomotility, gastric bleeding and renal dysfunction. A systematic assessment of pain is associated with a decreased incidence of pain, use of analgesics, duration of mechanical ventilation and length of stay (LOS) on the ICU [6-9].

As a result of these findings the Society of Intensive Care Medicine (SCCM) recommends that pain should be routinely monitored in all adult ICU patients [10]. A patient's self-report of pain is considered as the gold standard in the assessment of pain [11]. However, critically ill patients are often unable to communicate effectively due to severe illness, mechanical ventilation, administration of sedatives and analgesics or a decreased level of consciousness. Vital signs appear to be less valid for pain assessment in ICU patients due to underlying disease and treatment with inotropes and vasopressors [12]. Consequently, pain assessment in patients who are unable to self-report their pain is difficult [13-15]. Therefore, the SCCM advises the use of pain assessment tools which focus mainly on behavioral indicators of pain. The Behavioral Pain Scale (BPS) [15] and Critical-Care Pain Observation Tool (CPOT) [16] are behavioral pain assessment tools for uncommunicative and sedated ICU patients. The content validation, criterion validation, discriminant validation and inter-rater reliability of the BPS and CPOT have been tested in previous studies [7,13,15,17,18]. To date there are no studies available comparing these pain assessment tools simultaneously. The aim of this study was to compare the discriminant validation and reliability of the CPOT and the BPS in mechanically ventilated patients with the purpose to find the most useful clinical pain assessment tool for patients in a mixed adult ICU.

MATERIALS AND METHODS

We performed a prospective observational cohort study with a repeated measurement design in a 20-bed mixed closed-format ICU in a teaching hospital in Amsterdam, the Netherlands. The hospital has no neurosurgical facility. The local medical ethical committee approved the study and waived the requirement for written informed consent because this study did not require any deviation from the routine standard care on the ICU.

The ICU nurses screened all patients at bedside after admission. Inclusion criteria were: critically ill patients with (1) age ≥ 18 years, (2) an expected LOS on the ICU of ≥ 12 hours, (3) mechanical ventilation, (4) an inability to self-report pain. We excluded patients who were able to self-report pain and who were admitted for elective surgery, were quadriplegic or

paralyzed due to their current condition and/or treatment, were unable to be repositioned, or patients who participated in the study during a previous admission.

Assessments of pain, agitation/sedation and delirium

The BPS has been previously tested in mechanically ventilated ICU patients of which the majority were unconscious and therefore unable to self-report pain. This scale is based on a sum of three behavioral domains: facial expression, movements of the upper limbs and compliance with ventilation. Each domain is scored from 1 (no response) to 4 (full response). The score ranges from 3 (no pain) to 12 (maximum pain) [15] (Appendix 1).

The CPOT has been developed for the assessment of pain in critically ill adult patients unable to self-report pain. This scale consists of four behavioral domains: facial expression, body movements, muscle tension and compliance with the ventilation for intubated patients or vocalization for patients without endotracheal tube. Each domain is scored between 0 and 2 and the total score ranges from 0 (no pain) to 8 (maximum pain) [16] (Appendix 1). An extensive analysis and comparison of the psychometric properties of both tools is given in a recent review of Gélinas et al. [19].

The level of agitation and sedation was assessed with the Richmond Agitation and Sedation Scale (RASS) six times daily. This system assigns a score between4 (combative) and -5 (unresponsive). A score of zero indicates an alert and calm state [20,21]. The presence of delirium was routinely assessed by the nurses and attending physician using the Confusion Assessment Method for the ICU (CAM-ICU) [22].

Data collection

We extracted demographic and clinical characteristics from the patient clinical information system (CIS) (iMD-Soft: Metavison®, Tel Aviv, Israel), including the Acute Physiology and Chronic Health Evaluation IV predicted mortality (APACHE IV PM) score [23], the Sequential Organ Failure Assessment (SOFA) score [24] and the administration of analgesics and sedatives one and four hours before the pain assessments.

Pain medication in the ICU

The intensive care physician prescribed analgesics and/or sedatives, titrated to the patient's requirements. Depending on the degree of agitation and pain, patients either received morphine as a continuous infusion (in combination with midazolam for sedation) or piritramide 2.5-5 mg intravenously (iv). Piritramide is a synthetic opioid analgesic with a strength of approximately 0.7 times that of morphine [25]. Epidural levobupivacaine/sufentanyl was continued in the ICU if an epidural catheter was inserted peri-operatively. Fentanyl iv was used for short surgical interventions in the ICU. Ketamine iv was used for the treatment of status asthmaticus or pain that was unresponsive to the previously mentioned interventions. Levels of pain were not systematically assessed and recorded until the start of the training for this study.

Study procedures

The bedside nurse screened and included patients on the day of admission and performed, together with a second nurse, the assessments. The BPS and CPOT were performed simultaneously but independently of each other in four conditions: at rest just before a non-painful procedure, during the non-painful procedure, at rest just before a painful procedure and during the painful procedure. The first assessment recorded was always the BPS. We chose turning of the patient (turning) as a painful procedure and oral care as a non-painful procedure [26]. The procedures were chosen after a literature review and during an expert group meeting with ICU nurses, an intensivist/anesthesiologist and a clinical epidemiologist. The pairs of assessing nurses were not randomized but assigned by convenience and varied across the four procedures, however, the nurse in charge of the patient was always one of them. The assessors were asked to wait for at least 20 minutes after turning, or other painful procedures, before performing the assessments of the second procedure. The timing of the procedures was adjusted to the patient's day schedule. The nurses performed all assessments on the same day between 4am and 10am, and recorded the scores in custom made study forms in the Clinical Information System (CIS).

Training of the nursing staff

All ICU nurses were trained to use the BPS and CPOT for two hours during the annual ICU training. Training material consisted of a presentation with background information of pain, the study procedures and explanation of the pain scores, the paper versions of the BPS and CPOT, training posters and an instruction video [18,27]. This was followed by 30-minute weekly training sessions on the ICU, provided by members of the study group (expert team). We also posted an explanation of the study procedures and the use of BPS and CPOT on the ICU intranet. Additionally, an instruction card was available in every patient room. We performed a trial-run of one month in which we evaluated 66 test patients in order to minimize the possible bias of a learning curve and to provide bedside teaching of the study procedures.

Data analysis

Data were analyzed with SPSS version 18.0 (SPSS, Inc., Chicago Illinois) according to a prospectively defined protocol. Inter-rater reliability of the BPS and CPOT was tested by the calculation of intraclass correlation coefficients (ICC) for all assessments (one way random) [28]. Internal consistency was assessed with Cronbach's coefficient alpha using the scores during turning, when the patient was most likely to be experiencing pain. Values between 0.70 and 0.80 are considered as acceptable, values > 0.8 as good [29,30]. The discriminant validation was examined by calculating within-patient differences in scores between the assessments using the Friedman test. This is the non-parametric alternative to the one-way ANOVA with repeated measures. In order to determine which pairs of differences between

mean ranks were significant, and thus the likely source of a significant Friedman test, we performed a post hoc analysis with a non-parametric related-sample test, the Wilcoxon Signed Rank This test is suitable for ordinal or non-normally distributed continuous data [31]. The pain scores were not normally distributed and therefore we used non-parametric statistical tests. Only patients with complete scores were suitable for analysis. We hypothesized that the score should increase during the painful procedure and remain the same during the non-painful procedure.

RESULTS

During the 4-month study period 277 medical and surgical ICU patients and patients after major surgery were admitted, 245 patients were screened, and 123 patients met the inclusion criteria. The data of 68 patients (55% of the patients meeting the inclusion criteria) were complete and suitable for further analysis (Figure 1).

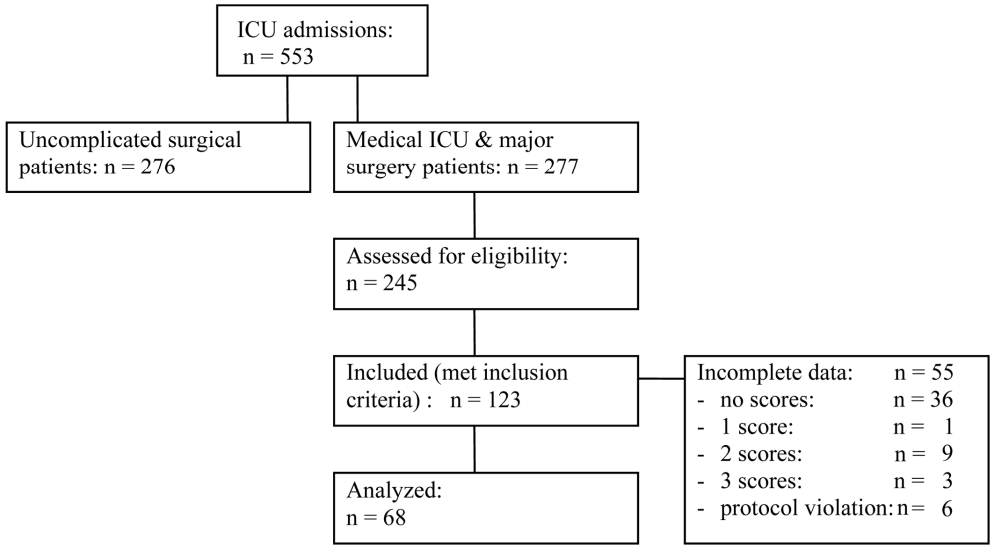


Figure 1. Flow chart

Reasons for incomplete assessments were: transfer to the ward or another hospital, deteriorating illness, death, extubation or excessive workload of the nurses. Patients with a complete data set had a significantly higher APACHE IV PM (0.23 [0.07 - 0.59] vs. 0.48 [0.24 - 0.77], p= 0.01), were older (66.79 ± 12.5 vs. 63.0 ± 17.8, p= 0.005) and had a longer ICU stay (82 [27.0 - 120.0] vs.169 [87.5 – 365.0], p= 0.00), compared to patients with incomplete data. Baseline characteristics are presented in table 1. The first assessments were performed 33.0 [20.25 - 59.75] hours after admission. All assessments were performed within 1.39 [00.51-04.31] hours. The median GCS and RASS did not significantly differ between the painful and non-painful procedures (turning and oral care). The median RASS was -2.0 [-3.0 – 0.0] during

all assessments. The median GCS was 9.0 [4.8 - 15.0] during rest I and oral care, and 9.5 [5.5 - 15.0] during rest II and turning. Approximately 30% of the patients received analgesics and/or sedatives of which midazolam and morphine were given to the majority of these patients.

Three patients received piritramide, ketamine, fentanyl or epidural levobupivacaine/sufentanyl.

Discriminant validation

The median BPS and CPOT scores of all 68 patients increased by two points from rest to the painful procedure: BPS 3.0 [3.0 - 3.0] to 5.0 [4.0 - 6.0], p= 0.000 and CPOT 0.0 [0.0 - 0.0] to 2.0 [0.0 - 3.0], p= 0.000 The median BPS scores between rest and the non-painful procedure showed a significant increase of one point, while the median CPOT score remained unchanged: BPS 3.0 [3.0 - 4.0] to 4.0 [3.0 - 4.0], p= 0.002 and CPOT 0.0 [0.0 – 0.0] to 0.0 [0.0 - 2.0], p= 0.002 (figure 2). The highest BPS and CPOT scores were 11 and 7, respectively during rest. This was in a patient not receiving any analgesic or sedative.

Reliability

The inter-rater reliability of the BPS and CPOT was recorded for all analyzed patients, with a total of 1088 assessments (68 patients x 2 raters x 4 different times x 2 scales). The ICC of the BPS scores and the ICC of the CPOT scores for all assessments showed a fair to good agreement. The ICC of the BPS (544 measurements) was 0.74 (95% CI 0.68 - 0.79), p= 0.001. The ICC of the CPOT (544 measurements) was 0.75 (95% CI 0.69 - 0.79), p= 0.001. Cronbach's alpha values indicated that the BPS and CPOT had an acceptable internal consistency during the painful procedure (turning) of 0.70 and 0.71 respectively.

Table 1. Baseline Characteristics

n= 68[1]	
Age (years)	66.79 ± 12.5
Sex (male)	41 (60.3)
APACHE IV PM[2]	0.48 [0.238 - 0.757]
SOFA[3] at moment assessments	7.0 [6.0 - 10.75]
Admission type	
Medical	48 (70.6)
Surgical	20 (23.5
Admission	
Non-elective	65 (95.6)
Elective	3 (4.4)
CPR[4]	12 (17.6)
BMI[6]	26.03 [23.09 – 31.80]
LOS (hours)	169.0 [87.5 - 365.0]
GCS[7]	
Rest I	9.0 [4.8 - 15.0]
Oral care	9.0 [4.8 - 15.0]
Rest II	9.5 [5.5 - 15.0]
Turning	9.5 [5.5 - 15.0]
CAM-ICU[8]	
Negative	16 (23.5)

Positive		5 (7.4)
Not feasible		39 (57.4)
RASS[9]		-2.0 [-3.0-0.0]
Oral care		
≤ -3		28 (41.2)
-2 – 0		30 (48.5)
≥1		7 (10.3)
Turning		
≤ -3		27 (39.7)
-2 – 0		38 (55.9)
≥1		6 (8.8)
Medication regime[10]		
1 hour before assessments		
Rest I		24 (35.3)
Oral care		24 (35.3)
Rest II		22 (32.4)
Turning		21 (30.9)
4 hours before assessments		
Rest I		23 (33.8)
Oral care		24 (35.3)
Rest II		23 (33.8)
Turning		23 (33.8)
Morphine[11]		
1 hour before assessments		
Rest I	n= 25	3.0 [2.0 - 5.0]
Oral care	n= 25	3.0 [2.0 - 5.0]
Rest II	n= 22	3.0 [2.0 - 6.0]
Turning	n= 22	3.0 [2.0 - 6.0]
Midazolam[11]		
Rest I	n= 23	1.8 [1.2 - 3.5]
Oral care	n= 23	1.8 [1.2 - 3.6]
Rest II	n= 20	1.8 [1.2 - 3.6]
Turning	n= 20	1.8 [1.2 - 3.3]

[1] Plus-minus values are means ± standard deviations (SD) other values are medians and interquartile ranges [IQR] or number (%)

[2] Acute Physiology and Chronic Health Evaluation IV predicted mortality

[3] Sepsis -relate Organ Failure Assessment scores can range from 0 to 24, with higher scores indicating more severe organ dysfunction.

[4] Patients admitted after cardiopulmonary resuscitation

[5] Body-mass index, kg/m^2

[6] Length of stay ICU [7] Glasgow Coma Scale

[8] Confusion Assessment Method for the ICU at day assessments

[9] Richmond Agitation and Sedation Scale

[10] Sedatives and/or analgesics. Sedatives included oxazepam, midazolam, haloperidol, ketamine and propofol. Analgesics included morphine, fentanyl piritramide, and levobupivacaine. Medication was administered enterally, intravenously or via an epidural catheter number (%)

[11] Milligrams per hour continuously

Table 2. Inter-rater Reliability

Assessment	BPS ICC[1]	CPOT ICC[1]
Overall	**0.74 95% CI [0.68 - 0.79]**	**0.75 95% CI [0.69 - 0.79]**
Rest I	0.70 95% CI [0.56 - 0.80]	0.72 95% CI [0.58 - 0.81]
Oral care	0.71 95% CI [0.57 - 0.81]	0.72 95% CI [0.58 - 0.82]
Rest II	0.80 95% CI [0.70 - 0.88]	0.80 95% CI [0.70 - 0.88]
Turning	0.60 95% CI [0.42 - 0.73]	0.62 95% CI [0.45 - 0.75]

[1] Intraclass correlation coefficient One-way random

Complementary analysis

After the primary analysis, the sample of 68 patients was divided into three subgroups according to their RASS scores. In a subgroup of 28 sedated patients (RASS -5, -4, -3) the median BPS and CPOT scores did not increase from rest to the non-painful procedure. However, the median BPS and CPOT scores in this subgroup increased by two points from rest to the painful procedure: BPS 3.0 [3.0 - 3.0] to 5.0 [4.0 - 6.8], p= 0.000 and CPOT 0.0 [0.0 - 0.0] to 2.0 [0.3 - 4,0], p= 0.000. In a subgroup of 33 calm patients (RASS -2, -1, 0) the median BPS scores between rest and the non-painful procedure increased by one point: BPS 3.0 [3.0 - 3.5] to 4.0 [3.0 - 4.5], p= 0.003. The median CPOT scores remained similar: CPOT 0.0 [0.0 - 0.0] to 0.0 [0.0 - 1.5], p= 0.006. A subgroup of 7 agitated patients (RASS +1) showed non-significant increases between rest and the procedures (figure 2).

DISCUSSION

This is the first prospective controlled study which has simultaneously assessed and compared discriminant validation and the reliability of two pain assessment scores, the BPS [15] and the CPOT [16] in critically ill, mechanically ventilated patients who were unable to self-report pain [29,32]. Discriminant validation of the BPS and CPOT was demonstrated by a significant increase in scores during a painful procedure (turning). However, the CPOT score remained unchanged when comparing a non-painful procedure (oral care) with rest, whereas the BPS score significantly increased during a non-painful procedure. Both tools showed a fair to good inter-rater reliability. Internal consistency of the BPS and the CPOT was acceptable.

The present study shows that both the BPS and the CPOT were able to discriminate between a non-painful procedure/rest and a painful procedure in patients who are unable to self-report pain, which is consistent with previous studies. However, contrary to previous studies, the BPS score in our study also increased significantly during the presumed non-painful procedure, whereas the CPOT did not [15,17,33]. Most of the increase in BPS score during oral care was the result of changes in facial expression and movements of the upper limbs. This increase might have been due to a reflex to touch rather than to pain. This finding is partially supported by the results of Young et al. [33]. who assessed 44 patients and also found an increase in BPS during a non-painful procedure (eye care) although this increase was not significant. Secondly,

the difference in discriminant validation of the CPOT and BPS during the non-painful stimulus could also be the result of the different number of options in each domain. For the BPS, nurses have to choose between four options compared to the three options of the CPOT. It is possible that the four options of the BPS could be less clearly distinguished than the three options of the CPOT and could therefore lead to incorrect assessment of a non-painful stimulus. Finally, Gélinas et al. described and analyzed the psychometric proportions of several pain assessment tools for use in nonverbal critically ill adults. According to their study, one of the limitations of the BPS is that operational description of some items may be interpreted differently by users [19].

This study also found that in a subgroup of sedated patients (RASS -5, -4, -3) neither of the two pain scores increased during oral care, while in calm patients (RASS -2, -1, 0) the BPS increased by one point. This could be explained by the fact that both the BPS and CPOT are developed for patients who cannot self-report their pain, mainly being deeply sedated patients [15,16,19]. Nurses have to decide in daily practice whether a patient is able to self-report pain or whether a behavioral pain assessment tool is appropriate for that individual patient. It is possible that they use an inappropriate method of assessing pain due to an incorrect assessment of a patient's abilities. An additional explanation could be that less sedated patients exhibited more behaviors than sedated patients, which might have led to errors in the pain assessment during a non-painful stimulus with the BPS due to the four options in each domain. The median BPS and CPOT scores in the small subgroup of agitated patients (RASS +1) were generally higher and increased during all procedures. Discriminating between behaviors as a result of agitation or pain might be difficult. This important issue requires further research.

The BPS and CPOT in our study showed a fair to good inter-rater reliability, 0.74 and 0.75 respectively. Previous studies generally found a higher inter-rater reliability, consistently being better during rest than during the painful procedure [13,15-17,29,34-37]. However, two studies analyzing the BPS found a lower inter-rater reliability during painful procedures [33,38]. Our inter-rater reliability could be lower because the nurses had to assess the BPS and CPOT simultaneously, which is more demanding. An additional reason may be the large number (105) of nurses on our ICU, resulting in less experience with the assessments. The number of assessors in prior studies was generally lower [15-17,35,37]. Furthermore, pairs of nurses differed in our study, only the bedside nurse was constant. In several previous studies, either one of the investigators or the physicians participated in the assessments [3,16,17,35]. Pain assessment by a large group of nurses is, however, a reflection of real life intensive care. Furthermore, the bedside nurse potentially understands the patient's reactions better because of a longer contact time. The inter-rater reliability during the painful procedure was lower than during the non-painful procedure (0.60 and 0.61 respectively). This could be due to the hypothesis that turning is a painful procedure for all ICU-patients, while it may only be uncomfortable or stressful in medical patients without wounds [26]. Behavioral changes in these patients might have been more difficult to interpret. Another explanation is that patients

could have exhibited more behaviors during turning and some of these might have been missed by the nurses during the assessments.

The internal consistency as measured with the Cronbach's alpha was acceptable for the BPS and the CPOT, and similar to other studies investigating the internal consistency of the BPS and CPOT [17,33,35,37-39].

Several studies have reported that critically ill patients frequently experience pain and discomfort and that approximately 50% of the patients report moderate to severe pain[1-5]. The BPS and CPOT in our study generally demonstrated low pain scores during the painful procedure, which is similar to previous studies using the BPS and CPOT [13,15,17,33,37]. This could be explained by the fact that these studies included sedated patients who can not self-report pain. Assessment of pain with behavioral pain scales could underestimate pain [40,41]. Gélinas et al. reported lower CPOT scores in their group of unconscious patients compared to the conscious patients [13]. They found that a higher dose of sedation and/or analgesics resulted in lower pain scores. We excluded patients who were able to self-report, which might have resulted in lower pain scores. However, in former studies between 75% and 100% of the patients received sedatives and/or analgesics, while only 34% of our patients received analgesics and/or sedatives [13,15,35,37,42,43]. This may be due to the higher proportion of medical patients in our study and the policy of aiming for the lowest acceptable dose of sedatives and analgesics. Finally, turning might not be painful for all critically ill patients, especially not for medical patients without wounds. This might explain the low increase in both scores comparing rest to turning. It is however unethical to provide an intentional painful stimulus to patients.

Limitations

Our study has potential limitations. First, the assessments were performed by bedside nurses and could therefore not be blinded. Consequently, the assessors were aware of which procedures were to be performed and they may have perceived more behavioral changes during turning because they thought that this was a painful procedure. This could have led to higher scores during the painful procedure and an overestimation of the discriminant validation. Second, the BPS was always completed first which could have affected the study data. Randomizing the order of the assessments could have strengthened the study design. Third, the number of patients included in the final analysis was relatively small which might have led to an unrepresentative group of patients. We excluded patients which had incomplete data, and this could have caused selection bias. Therefore, we compared the baseline characteristics of the 68 analyzed patients with the 209 patients excluded from the study and final analysis. We also compared the baseline characteristics of the analyzed patients (68) with the patients with incomplete data (55) (figure 1). The analyzed patients had a significantly higher Apache IV predicted mortality, were older and had a longer ICU stay which conformed to the inclusion criteria. Other patient characteristics did not significantly differ. Therefore, we

believe that our study sample reflects the group of patients for whom the BPS and CPOT are designed and that less severely ill patients were excluded.

The fourth limitation is that we initially trained the expert group and nurses in the English versions of the BPS and CPOT. During our study the BPS and CPOT had not been translated and validated in Dutch. An expert group of ICU nurses, an intensivist/ anesthesiologist and a qualified English language translator translated both tools (the short descriptions only) in Dutch in our CIS, thus language misconceptions might have occurred. We did not perform a double forward and backward translation according to the requirements of an official translation. However, the descriptions in both tools are short, clear and contain universally interpretable signs and Dutch people generally have a high standard of English so we believe that this factor has not likely affected our results.

Finally, the presence of a delirium could interfere with the behavioral pain scores and therefore be a confounding factor. Nurses at our ICU routinely assess twice daily the incidence of delirium with the CAM-ICU. A positive CAM-ICU was reported in 5/68 patients but we cannot be sure that this reflects the true incidence of delirium as the assessment of delirium in patients with decreased consciousness is extremely difficult. A valid and structured assessment of delirium performed by a psychiatrist should be taken account of during future research of behavioral pain scales.

CONCLUSION

The present study in critically ill ventilated patients unable to self-report pain shows that the CPOT and BPS both had a fair to good inter-rater reliability. However, the CPOT was superior to the BPS in assessing pain, according to the discriminant validation. Although both scores increased with a presumed painful stimulus, the BPS had a reduced discriminative performance because it increased with a non-painful stimulus as well. We therefore prefer using the CPOT for this particular group of patients. The use of a behavioral pain scale in daily intensive care practice could optimize pain treatment and use of analgesics and sedatives in critically ill patients. However, further research is needed on discrimination between discomfort, stress and pain, on the effect of analgesics and/or sedatives on the CPOT, and on the assessment of pain in restless, agitated or delirious patients.

Acknowledgements

The authors thank R. Koper, R. Peek, M. Koning, the ICU-nurses and information technology specialist A. Bianchi for their contributions to this study, as well as S.Toohey for his English revisions.

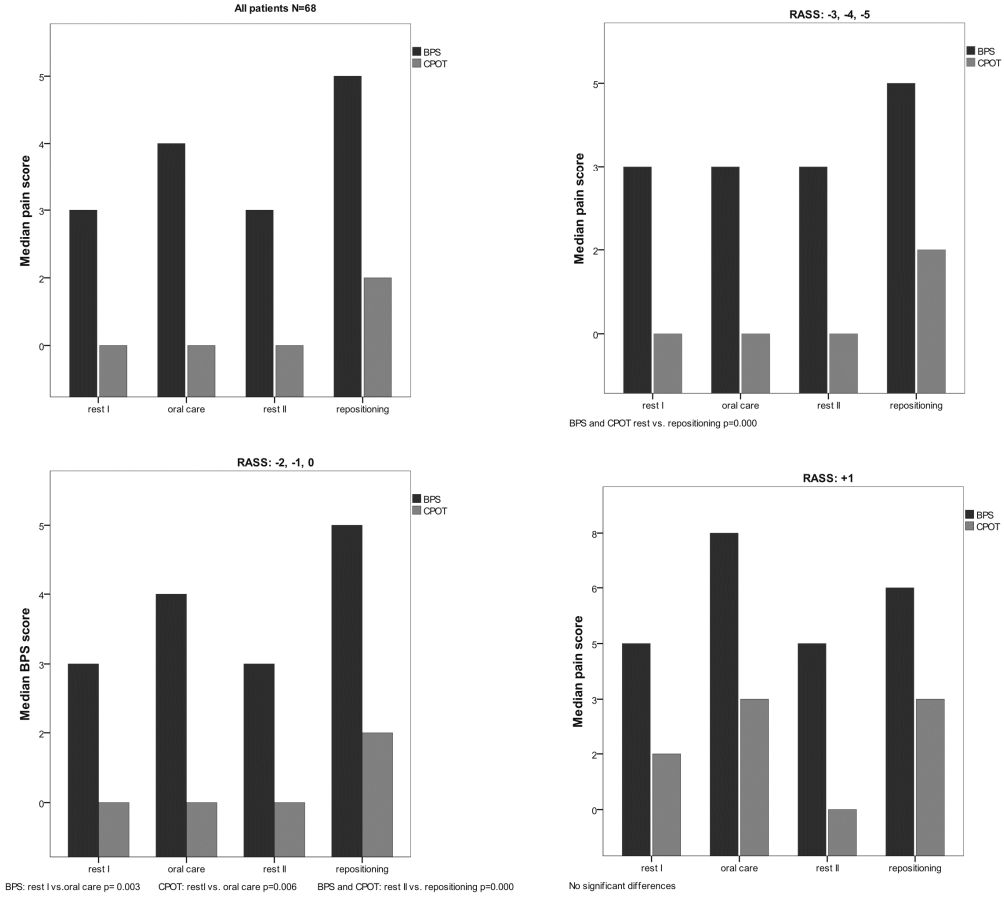


Figure 2. Results discriminant validation

All patients n= 68

BPS rest II 3.0[3.0-3.0] vs. turning 5.0[4.0-6.0], Z = -5.815, *p* = 0.000
CPOT rest II 0.0[0.0-0.0] vs. turning 2.0[0.0-3.0], Z = -5.496, *p* = 0.000
BPS rest I 3.0[3.0-4.0] vs. oral care 4.0[3.0-4.0], Z = -3.149, *p* = 0.002
CPOT rest I 0.0[0.0-0.0] vs. oral care 0.0[0.0-2.0], Z = -3.045, *p* = 0.002

REFERENCES

1. Desbiens NA, Wu AW, Broste SK, et al: Pain and satisfaction with pain control in seriously ill hospitalized adults: findings from the SUPPORT research investigations. For the SUPPORT investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatmentm. Crit Care Med 1996; 24:1953-1961.
2. Li DT, Puntillo K: A pilot study on coexisting symptoms in intensive care patients. Appl Nurs Res 2006; 19:216-219.
3. Nelson JE, Meier DE, Oei EJ, et al: Self-reported symptom experience of critically ill cancer patients receiving intensive care. Crit Care Med 2001; 29:277-282.
4. Nelson JE, Meier DE, Litke A, et al: The symptom burden of chronic critical illness. Crit Care Med 2004; 32:1527-1534.
5. Puntillo K, Pasero C, Li D, et al: Evaluation of pain in ICU patients. Chest 2009; 135:1069-1074.
6. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators: JAMA 1995; 274:1591-1598.
7. Chanques G, Jaber S, Barbotte E, et al: Impact of systematic evaluation of pain and agitation in an intensive care unit. Crit Care Med 2006; 34:1691-1699.
8. Payen JF, Chanques G, Mantz J, et al: Current practices in sedation and analgesia for mechanically ventilated critically ill patients: a prospective multicenter patient-based study. Anesthesiology 2007; 106:687-695.
9. Payen JF, Bosson JL, Chanques G, et al: Pain assessment is associated with decreased duration of mechanical ventilation in the intensive care unit: a post Hoc analysis of the DOLOREA study. Anesthesiology 2009; 111:1308-1316.
10. Barr J, Fraser GL, Puntillo K, et al: Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013; 41:263-306.
11. Chanques G, Viel E, Constantin JM, et al: The measurement of pain in intensive care unit: comparison of 5 self-report intensity scales. Pain 2010; 151:711-721.
12. Puntillo KA, Miaskowski C, Kehrle K, et al: Relationship between behavioral and physiological indicators of pain, critical care patients' self-reports of pain, and opioid administration. Crit Care Med 1997; 25:1159-1166.
13. Gelinas C, Johnston C: Pain assessment in the critically ill ventilated adult: validation of the Critical-Care Pain Observation Tool and physiologic indicators. Clin J Pain 2007; 23:497-505.
14. Hamill-Ruth RJ, Marohn ML: Evaluation of pain in the critically ill patient. Crit Care Clin 1999; 15:35-vi.
15. Payen JF, Bru O, Bosson JL, et al: Assessing pain in critically ill sedated patients by using a behavioral pain scale. Crit Care Med 2001; 29:2258-2263.
16. Gelinas C, Fillion L, Puntillo KA, et al: Validation of the critical-care pain observation tool in adult patients. Am J Crit Care 2006; 15:420-427.
17. Ahlers SJ, van der Veen AM, van Dijk M, et al: The use of the Behavioral Pain Scale to assess pain in conscious sedated patients. Anesth Analg 2010; 110:127-133.
18. Chanques G, Payen JF, Mercier G, et al: Assessing pain in non-intubated critically ill patients unable to self report: an adaptation of the Behavioral Pain Scale. Intensive Care Med 2009; 35:2060-2067.
19. Gelinas C, Puntillo KA, Joffe AM, et al: A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Semin Respir Crit Care Med 2013; 34:153-168.
20. Sessler CN, Gosnell MS, Grap MJ, et al: The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med 2002; 166:1338-1344.

21. Robinson BR, Berube M, Barr J, et al: Psychometric analysis of subjective sedation scales in critically ill adults. Crit Care Med 2013; 41:S16-S29.
22. Ely EW, Inouye SK, Bernard GR, et al: Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). JAMA 2001; 286:2703-2710.
23. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006; 34:1297-1310.
24. Vincent JL, Moreno R, Takala J, et al: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996; 22:707-710.
25. Punt C, Kreutz M, Dekkers P, et al: Comparison of intravenous boluses of piritraminde and morphine. Did we use the correct ratio of analgetic potency? 29 ed. 2014. p. 196.
26. Puntillo KA, Morris AB, Thompson CL, et al: Pain behaviors observed during six common procedures:results from Thunder Project II. Crit Care Med 2004; 32:421-427.
27. http://nursingpathways.kp.org/national/learning/webvideo/resources/cpot/.
28. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86:420-428.
29. Gelinas C, Puntillo KA, Joffe AM, et al: A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Semin Respir Crit Care Med 2013; 34:153-168.
30. Streiner D.L, Norman G.R. Health measurement scales a practical guide to their development and use. third edition edn. Hamilton: Oxford university press; 2003.
31. Leech NL, Barret KC, Morgan GA. Repeated-measures and mixed anovas. IBM SPSS for Intermediate Statistics Use and Interpretation.New York: Taylor & Francis Group; 2011. p. 175-84.
32. Pudas-Tahka SM, Axelin A, Aantaa R, et al: Pain assessment tools for unconscious or sedated intensive care patients: a systematic review. J Adv Nurs 2009; 65:946-956.
33. Young J, Siffleet J, Nikoletti S, et al: Use of a Behavioural Pain Scale to assess pain in ventilated, unconscious and/or sedated patients. Intensive Crit Care Nurs 2006; 22:32-39.
34. Ahlers SJ, van Gulik L, van der Veen AM, et al: Comparison of different pain scoring systems in critically ill patients in a general ICU. Crit Care 2008; 12:R15.
35. Aissaoui Y, Zeggwagh AA, Zekraoui A, et al: Validation of a behavioral pain scale in critically ill, sedated, and mechanically ventilated patients. Anesth Analg 2005; 101:1470-1476.
36. Gelinas C, Arbour C, Michaud C, et al: Implementation of the critical-care pain observation tool on pain assessment/management nursing practices in an intensive care unit with nonverbal critically ill adults: a before and after study. Int J Nurs Stud 2011; 48:1495-1504.
37. Nurnberg DD, Saboonchi F, Sackey PV, et al: A preliminary validation of the Swedish version of the Critical-Care Pain Observation Tool in adults. Acta Anaesthesiol Scand 2011; 55:379-386.
38. Juarez P, Bach A, Baker M, et al: Comparison of two pain scales for the assessment of pain in the ventilated adult patient. Dimens Crit Care Nurs 2010; 29:307-315.
39. Paulson-Conger M, Leske J, Mai dL C, et al: Comparison of two pain assessment tools in nonverbal critical care patients. Pain Manag Nurs 2011; 12:218-224.
40. Prkachin KM, Solomon PE, Ross J: Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. Can J Nurs Res 2007; 39:88-106.
41. Prkachin KM, Rocha EM: High levels of vicarious exposure bias pain judgments. J Pain 2010; 11:904-909.
42. Gelinas C, Fillion L, Puntillo KA: Item selection and content validity of the Critical-Care Pain Observation Tool for non-verbal adults. J Adv Nurs 2009; 65:203-216.
43. Gelinas C, Harel F, Fillion L, et al: Sensitivity and specificity of the critical-care pain observation tool for the detection of pain in intubated adults after cardiac surgery. J Pain Symptom Manage 2009; 37:58-67.

APPENDIX I

Behavioral Pain Scale

Item	Description	Score
Facial expression	Relaxed	1
	Partially tightened	2
	Fully tightened	3
	Grimacing	4
Upper limbs	No movement	1
	Partially bent	2
	Fully bent with finger flexion	3
	Permanently retracted	4
Compliance with Ventilation	Tolerating movement	1
	Coughing but tolerating ventilation for majority of the time	2
	Fighting ventilator	3
	Unable to control ventilation	4

APPENDIX II

Critical-Care Pain Observation Tool

Indicator	Description		Score
Facial expression	No muscular tension observed	Relaxed, neutral	0
	Presence of frowning, brow lowering, orbit tightening, and levator contraction	Tense	1
	All of the above facial movements plus eyelid tightly closed	Grimacing	2
Body movements	Does not move at all	Absence of movements	0
	Slow, cautious movements, touching or rubbing the pain site, seeking attention through movements	Protection	1
	Pulling tube, attempting to sit up, moving limbs/thrashing, not following commands, striking at staff, trying to climb out of bed	Restlessness	2
Muscle tension evaluation by passive flexion and extension of upper extremities	No resistance to passive movements	Relaxed	0
	Resistance to passive movements	Tense, rigid	1
	Strong resistance to passive movements, inability to complete them	Very tense or rigid	2
Compliance with the ventilator	Alarms not activated, easy ventilation	Toleration ventilator or Movement	0
	Alarm stop spontaneously	Coughing but tolerating	1
	Asynchrony: blocking ventilation, alarms frequently activated	Fighting ventilator	2

CHAPTER 3
Pain measurement in mechanically ventilated patients after cardiac surgery: Comparison of the Behavioral Pain Scale (BPS) and the Critical-Care Pain Observation Tool (CPOT)

ABSTRACT

Introduction

The Behavioral Pain Scale (BPS) and Critical-Care Pain Observation Tool (CPOT) are behavioral pain assessment tools for sedated and unconscious critically ill patients. The aim of this study was to compare the reliability, internal consistency and discriminant validation of the BPS and the CPOT simultaneously, in mechanically ventilated patients post-cardiac surgery.

Materials and Methods

Design: A prospective observational cohort study.
Setting: A 20-bed closed-format ICU with mixed medical, surgical and cardiac surgery patients in a teaching hospital in Amsterdam, the Netherlands
Participants: 72 intubated and mechanically consecutive ventilated patients post-cardiac surgery who were not able to self-report pain.

Results

Two nurses assessed the BPS and CPOT simultaneously and independently at four moments – during: rest, a non-painful procedure (oral care), rest and a painful procedure (turning). Both scores showed a significant increase of two points between rest and turning. The median BPS score of nurse1 showed a significant increase of one point between rest and the non-painful procedure (oral care), whereas both median CPOT scores did not change. The inter-rater reliability of the BPS and CPOT showed substantial agreement of 0.74 of all assessments. During rest I & II values ranged from 0.24 to 0.46. The ICCs of the BPS and CPOT during turning were 0.75 and 0.62, respectively. Cronbach's alpha values for the BPS were 0.62 (nurse1) and 0.59 (nurse2), compared to 0.65 and 0.58 for the CPOT.

Conclusions

The BPS and CPOT are reliable and valid pain assessment tools in a daily clinical setting. However, the discriminant validation of both scores seems less satisfactory in sedated or agitated patients and this topic requires further investigation.

INTRODUCTION

Procedural pain and pain at rest is common in critically ill patients and is a considerable stressor. It has both short- and long-term psychological and physiological consequences and has a negative effect on recovery [1-5]. Severe pain and a number of other adverse experiences have been linked to the development of posttraumatic stress disorder related symptoms post-ICU [6]. Post-cardiac surgery patients in the ICU are prone to experience procedural pain due to chest tubes and wounds. Recent research determined an association between cardiac surgery and the development of chronic post-surgical pain [5].

Because of these adverse effects of pain, clinical practice guidelines recommend individualized goal directed pain management for ICU patients and a routine monitoring of pain with a validated scale appropriate for the patient's level of consciousness [1;2]. A patient's self-report of pain is acknowledged as the gold standard in the assessment of pain [7]. However, self-assessment of pain in ICU patients is often hampered due to mechanical ventilation, treatment with sedatives and analgesics or a decreased level of consciousness caused by severe illness or delirium. In these nonverbal critically ill patients, pain can be monitored with behavioral pain assessment tools such as the Behavioral Pain Scale (BPS) and Critical-Care Pain Observation Tool (CPOT) [1;2;8-12].

Psychometric properties of the BPS and CPOT have been tested and reviewed previously but, due to inclusion of a small number of nonverbal patients, only limited information is available about critically ill patients post-cardiac surgery who are unable to rate their pain [8;9;11-26]. This specific group of patients most likely differs from general ICU patients because they are post-anesthesia and underwent specific surgical procedures. Surgical ICU patients indicate the surgical site as the most painful location during rest, whereas medical patients most likely indicate pain in their limbs and back [27]. Additionally, there are no studies available to date comparing the BPS and CPOT simultaneously measured in exclusively post-cardiac surgery mechanically ventilated patients without communication capabilities in a daily clinical setting.

We aimed in this study to compare the inter-rater reliability, internal consistency and discriminant validation of the BPS and the CPOT in mechanically ventilated patients post cardiac surgery unable to self-report pain.

MATERIALS AND METHODS

We conducted a prospective observational cohort study with a repeated measurements design. In this design, the patients are their own comparison. The setting is a 20-bed closed-format ICU with mixed medical, surgical and cardiac surgery patients in a teaching hospital in Amsterdam, the Netherlands. The local medical ethical committee approved the study and waived the requirement for written informed consent because of its observational design according to Dutch and European regulations. The ICU nurses screened all patients after admission for eligibility using a digital screening log. Mechanically ventilated patients after cardiac surgery who were (1) ≥ 18 years, (2) unable to self-report pain, (3) expected to stay on the ICU ≥ 12 hours and (4) were able to physically respond to stimuli were included in the study. We excluded patients who were unable to be repositioned, were paralyzed or

quadriplegic due to their medical condition and/or treatment; or patients who participated in the study during an earlier admission to our ICU.

Assessments of pain, sedation/agitation and delirium

The BPS is a validated observational pain scale for unconscious mechanically ventilated patients and is based on the sum score concerning three behavioral items: facial expression, movements of the upper limbs and compliance with ventilation. Each item is scored from 1 (no response) to 4 (full response). The total BPS score ranges from 3 (no pain) to 12 (maximal pain) - see the appendix for a complete description of the items. The selection of items is established from a literature review and a questionnaire among ICU nurses. The psychometric properties of the BPS have been tested in various subsets of critically ill patients (i.e., medical, post-operative and trauma) [9].

The CPOT is a validated observational pain scale for the assessment of pain in both intubated and non-intubated critically ill adult patients incapable to self-report their pain. The scale is constructed from literature review, retrospective reviews of common pain characteristics in patients' medical files and consultation of ICU nurses and physicians. The CPOT is based on the sum of four behavioral items: facial expression, body movements, muscle tension and compliance with ventilation for intubated patients (or vocalization for patients without an endotracheal tube). Each item is scored between 0 and 2 and the total CPOT score ranges from 0 (no pain) to 8 (maximal pain) [10-12] - see the appendix for a complete description of the items.

The Richmond Agitation and Sedation Scale (RASS) was performed six times daily to assess the level of agitation and sedation. The RASS ranges from +4 (combative) to -5 (unresponsive) and a score of zero indicates an alert and calm state [28]. The presence of delirium was routinely assessed three times daily by the bedside nurses and the attending intensive care physician using the Confusion Assessment Method for the ICU (CAM-ICU) [29] . The RASS and CAM-ICU have been part of the routine care since 2006 and nurses have been frequently trained in the use of these tools.

Data collection

We extracted clinical characteristics and demographic data from the patient clinical information system (CIS) (iMD-Soft: Metavison®, Tel Aviv, Israel), along with the Sequential Organ Failure Assessment (SOFA) score, the Acute Physiology and Chronic Health Evaluation IV predicted mortality (APACHE IV PM), and the administration of analgesics and sedatives in the four hours preceding the pain assessments.

Intraoperative and postoperative treatment

All patients received general anesthesia, which was tailored to the specific condition of each patient. The anesthesia protocol was based on analgesia, hypnosis, amnesia, skeletal muscle

relaxation and inhibition of sensory and autonomic reflexes. This balanced anesthesia has a focus on the reduction of stress by the administration of a relatively high dose of opioids.
Cardio protective interventions followed the local protocol and were individualized for each patient. Patients were treated with antegrade of retrograde cold blood cardioplegia or crystalloid cardioplegia, which was delivered intermittently or continuously. Rectal temperature off bypass had to be >36°Celcius before transfer to the ICU. Intraoperative medication was administrated by continuous infusion and/or intermittent bolus administration and was always discontinued before the patient was transferred to the ICU.
In the ICU, patients are permitted to be awake, unless this interferes with their ICU treatment. The attending physician, who is present 24/7, prescribes analgesics and/or sedatives titrated to the patient's needs. In awake patients the level of pain is assessed with a numerical rating scale (NRS) and if necessary patients either received morphine as a continuous infusion (in combination with midazolam for sedation) or a bolus injection piritramide 2.5-5 mg intravenously (iv). Piritramide is a synthetic opioid analgesic with a strength of approximately 0.7 times that of morphine [30]. Fentanyl iv was used for short surgical interventions in the ICU. Pain in patients unable to self-report pain was not systematically assessed and reported until the start of this study [31].

Study procedures

The bedside ICU nurses screened all patients for eligibility within the first hour after cardiac surgery. The timing of the study procedures was adjusted to the patient's treatment schedule and had to be performed in one day between 4 pm and 10 am. Inclusion was only allowed when patients were able to physically respond to stimuli and unable to rate their pain. The bedside nurse and an ICU nurse colleague assessed the BPS and CPOT simultaneously but independently of each other at four pre-defined moments: at rest just before a non-painful procedure (1), during the non-painful procedure (2), at rest just before a painful procedure (3) and during the painful procedure (4). The first assessment was always the BPS. These procedures were selected based on a literature review, which was discussed during an expert group meeting between ICU nurses, an intensivist/anesthesiologist and a clinical epidemiologist. We chose oral care as a non-painful procedure and turning of the patient (repositioning) as a painful procedure [4;32]. The pairs of assessing nurses were composed of the nurse in charge of the patient and a second available nurse. As a consequence, the combination of nurses could differ during the four procedures. The nurses were instructed to wait for a minimum of 20 minutes after turning or other painful procedures like tracheal suctioning, before performing the assessments of the following procedure. The nurses recorded the pain scores immediately after the assessments in a separate study form within the clinical information system (CIS).

Training of the nursing staff

The study team, consisting of ICU nurses, an intensivist/anesthesiologist, a clinical epidemiologist and a qualified English language translator, provided a custom made BPS and CPOT course for the complete nursing staff during the routine ICU training days. Training material consisted of a PowerPoint presentation, paper versions of the BPS and CPOT, training posters and an instruction video [33). This was followed by weekly training sessions to become familiar with the study procedures and the scores. Additionally, an instruction card was made available in all patient rooms. We performed a trial-run on 66 patients in order to minimize the possible bias of a learning curve and to provide bedside teaching of the study procedures.

Data analysis

Data were analyzed with Statistical Package for the Social Sciences (SPSS) version 20.0 (SPSS, Inc., Chicago Illinois) according to a prospectively defined analysis plan. Inter-rater reliability of the BPS and CPOT was examined by calculation of intraclass correlation coefficients (ICC) for all assessments (one way random) [34]. ICC values between 0.21 to 0.40 are considered as fair agreement, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial and above 0.80 as almost perfect [35]. Internal consistency was examined with Cronbach's coefficient alpha, using the scores during turning when the patient was most likely to be experiencing pain. Values between 0.70 and 0.80 are considered as acceptable, values > 0.80 as good [36]. We assessed the Item-Total Statistics to see whether the value increased by deleting a domain. The discriminant validation was assessed by calculating within-patient differences in pain scores between the procedures (restI vs. oral care and restII vs. turning) using the Friedman test. In order to determine which pairs of differences between mean ranks were significant, and thus the likely source of a significant Friedman test, we performed a post-hoc analysis with the Wilcoxon Signed Rank test [37]. To gain insight into the pain scores during different levels of sedation and agitation, we divided the scores into three subgroups (sedated: RASS -5, -4, -3, calm: RASS -2, -1, 0 and agitated: RASS ≥ +1) according to the RASS at the measurement (rest) just before the procedures (oral care and turning).

Only patients with complete scores were included in the analysis. We hypothesized that the BPS and CPOT scores should significantly increase during the painful procedure and remain similar during the non-painful procedure compared to rest. The alpha is adjusted to 0.01 because multiple comparisons were made on the same subjects.

A sample size of 68 subjects with two observations per subject achieves 90% power to detect an Intraclass correlation of 0.95 under the alternative hypothesis when the intraclass correlation under the null hypothesis is 0.90 using an F-test with a significance level of 0.05 [38-40]. We therefore aimed to include 68 patients with complete assessments performed by two nurses.

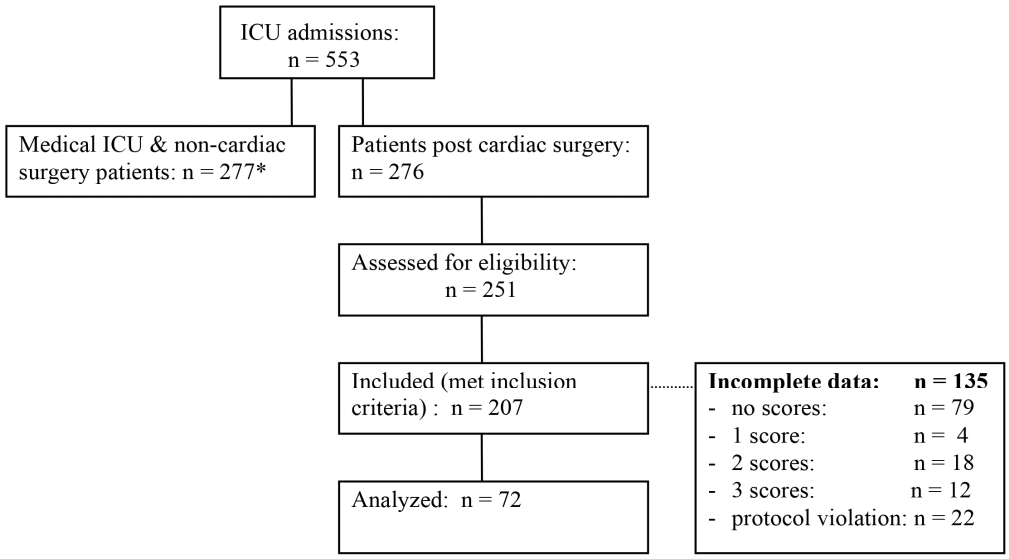


Figure 1. Flowchart

*Separately analyzed and published before [31]

RESULTS

Over the 4-month study period, 276 patients were admitted at the ICU after cardiothoracic surgery, 251 patients were screened and 207 patients were eligible for inclusion. The data of 72 patients (35% of the patients meeting the inclusion criteria) were complete and used for the analysis of the psychometric proportions of the BPS and CPOT (Figure 1). Reasons for incomplete data were: extubation, transfer to the ward, post-surgical complications or excessive workload of the nurses. Patients included in the analysis were significantly older (70.0 ± 12.0 vs. 65.5 ± 11.3; p= 0.027) and had a longer ICU stay (24.5 [21.3 - 47.8] vs. 22.0 [18.0 - 38.8] hours; p= 0.001), compared to patients with incomplete data (n= 135).

The first assessments were performed 223 [148 - 311] minutes after admission. All assessments were completed within 198 [95 - 384] minutes from the start of the first assessment. The median GCS and RASS did not significantly differ between rest and the procedures (oral care and turning). Approximately 36% of the patients received analgesics (morphine and piritramide) and/or sedatives (oxazepam, midazolam, diazepam and haloperidol) in the four-hour ICU period before the assessments, of which piritramide and diazepam were given to the majority of these patients. No patients received epidural analgesia using bupivacaine or opioids one and four hours before the assessments on the ICU. Two patients were treated with a continuous infusion of morphine combined with midazolam. All baseline characteristics are presented in table 1.

Reliability

The inter-rater reliability of the BPS and CPOT was recorded for all analyzed patients, with a total of 1,152 assessments (72 patients x 2 raters x 4 different times x 2 scales). The ICCs of the BPS and CPOT for all assessments showed a fair to good agreement. The ICC values of the BPS and CPOT (both 544 measurements) were equal: 0.74 (95% CI 0.68 - 0.79), p= 0.001. During rest, the ICC values of both pain scores were lower than during the procedures (Table 2).

Cronbach's alpha values for the BPS of both nurses during turning were 0.62 (nurse1) and 0.59 (nurse2). The values of Cronbach's alpha for the CPOT of nurse1 and nurse2 were 0.65 and 0.58 respectively. Exclusion of a domain did not result in higher values.

Discriminant validation

The median BPS and CPOT scores of both nurses increased significantly by two points from restII to the painful procedure (turning). The median BPS score between restI and the non-painful procedure (oral care) of nurse1 showed a significant increase of a half point, while the median BPS score of nurse2 and the CPOT scores of both nurses remained unchanged: BPS nurse1 3.0 [3.0 - 3.0] to 3.5 [3.0 - 5.0] p= 0.00 (Table 3). This increase during oral care was mainly caused by higher scores in the domains facial expression and movements of the upper limbs. The highest BPS scores were 9 during turning (five times) and rest (once). The highest CPOT score was 6 during turning.

The BPS and CPOT in the subgroup of 38 calm patients (RASS -2, -1, 0) did not show an increase from restI to oral care. The BPS and CPOT of both nurses between restII and turning increased significantly by 1.5 (CPOT nurse2) and two points (BPS both nurses and CPOT nurse1).

In a subgroup of 30 sedated patients (RASS -5, -4, -3), the median BPS scores of both nurses and the median CPOT score of nurse1 increased significantly by one point from rest1 to oral care. Between rest2 and turning, the BPS and CPOT showed significant increases of two points (nurse1) and 2.5 points (nurse2). The subgroup of four agitated patients (RASS ≥ +1) was too small to show significant differences between the measurements (Table 3).

Table 1. Baseline characteristics

n= 72[1]	
Age (years)	70.0 ± 12.0
Sex (male)	52 (72.2)
Elective admission	71 (98.6)
Surgery[2]	
CABG	31 (43.1)
CABG + valve surgery	20 (27.8)
CABG + other cardiac surgery	2 (2.8)

Valve surgery		18 (25)
Other cardiac surgery		1 (1.4)
Full sternotomy		70 (97.2)
Mini-sternotomy		2 (2.8)
Re-operation		13 (18.1)
Off pump		2 (2.8)
ECC (minutes)[3]		103.0 [84.5 - 144.5]
Aortic cross clamp (minutes)		69.0 [54.0 - 93.5]
Drains		46 (25.3)
Pleural		
Mediastinal		38 (20.9)
Substernal		14 (7.7)
Pericardial		86 (47.3)
ASA[4]		
ASA II		2 (2.9)
ASA III		55 (78.6)
ASA IV		12 (18.6)
EuroSCORE[5]		4.5 [2.1-7.9]
APACHE IV PM[6]		0.015 [0.005 - 0.037]
BMI[8]		26.1 [24.0 - 29.7]
LOS ICU (hours)[9]		24.5 [21.3 - 47.8]
GCS[10]		
Rest I & Rest II		3.0 [3.0 - 4.0]
Oral care & Turning		3.0 [3.0 - 6.0]
CAM-ICU[11]		
Negative		40 (55.3)
Positive		3 (4.2)
Not feasible		16 (22.2)
RASS[12]		
Rest1		-2.0 [-3.0 - -1.0]
Oral care		-2.0 [-3.0 - -1.0]
Rest2		-1.0 [-2.0 - -1.0]
Turning		-1.0 [-2.0 - -0.3]
Medication regime ICU[13]	no./total no. (%)	
1 hour before assessments		
Rest1		12 (16.7)
Oral care		11 (15.3)
Rest2		7 (9.7)
Turning		6 (8.3)
4 hours before assessments		
Rest1		27 (37.5)
Oral care		26 (36.1)
Rest2		26 (36.1)
Turning		24 (33.3)

[1] Plus-minus values are means ± standard deviations (SD) other values are medians and interquartile ranges [IQR]or number (%(

[2] Coronary artery bypass graft surgery/ Valve surgery: one or more valves (aortic valve Replacement, mitral valve replacement, tricuspid valve replacement)/ Other cardiac surgery: aortic prosthesis surgery.

[3] Duration of extra corporeal circulation in minutes

[4] ASA physical status classification system

[5] Logistic European System for Cardiac Operative Risk Evaluation (0 - 100%)

[6] Acute Physiology and Chronic Health Evaluation IV predicted mortality

[7] Sepsis-related Organ Failure Assessment scores can range from 0 to 24, with higher scores indicating more severe organ dysfunction.

[8] Body-mass index. kg/m^2
[9] Length of stay Intensive Care Unit
[10] Glasgow Coma Scale
[11] Confusion Assessment Method for the ICU at day assessments
[12] Richmond Agitation and Sedation Scale
[13] Analgesics and/or sedatives. Analgesics included morphine and piritramide.
Sedatives included oxazepam, midazolam, diazepam and haloperidol. Medication was administered enterally and intravenously. no= number of patients
[14] Total milligrams 4 hours before assessments

Table 2. Inter-rater reliability

Assessment	BPS ICC[1] (95%CI[2])	CPOT ICC (95%CI)
Overall	**0.744 [0.687–0.791]**	**0.743 [0.687 – 0.791]**
RestI	0.239 [0.010 – 0.444]	0.344 [0.124 – 0.531]
Oral care	0.712 [0.577 – 0.810]	0.829 [0.741 – 0.890]
RestII	0.398 [0.186 – 0.575]	0.465 [0.264 – 0.628]
Turning	0.750 [0.629 – 0.836]	0.623 [0.456 – 0.746]

[1] Intraclass correlation coefficients (one way random) [34]
[2] Confidence interval

Table 3. Discriminant validation

BPS	n= 72	Sedated[1]	Calm[2]	Agitated[3]
Nurse 1				
RestI	3.0 [3.0 – 3.0]	3.0 [3.0 – 3.0]	3.0 [3.0 – 3.0]	3.5 [3.0 – 4.0]
Oral care	3.5 [3.0 – 5.0]	4.0 [3.0 – 5.0]	3.0 [3.0 – 4.0]	5.5 [3.5 – 6.0]
p-Value[6]	*0.001*	*0.001*	*0.002*	*0.109*
RestII	3.0 [3.0 – 4.0]	3.0 [3.0 – 3.0]	3.0 [3.0 – 4.0]	4.5 [3.3 – 5.8]
Turning	5.0 [4.0 – 6.0]	5.0 [4.0 – 6.0]	5.0 [4.0 – 5.0]	4.5 [4.0 – 5.8]
p-Value[7]	*0.001*	*0.001*	*0.001*	*0.564*
p-Value 4 moments[8]	*0.001*	*0.001*[9]	*0.001*[9]	*0.099*[9]
Nurse 2				
RestI	3.0 [3.0 – 3.0]	3.0 [3.0 – 3.0]	3.0 [3.0 – 3.3]	3.0 [3.0 – 3.8]
Oral care	3.0 [3.0 – 4.0]	4.0 [3.0 – 5.0]	3.0 [3.0 – 4.0]	6.5 [3.9 – 7.0]
p-Value[6]	*0.001*	*0.001*	*0.005*	*0.102*
RestII	3.0 [3.0 – 4.0]	3.0 [3.0 – 3.0]	3.0 [3.0 – 3.8]	4.0 [3.3 – 4.8]
Turning	5.0 [4.0 – 6.0]	5.5 [4.0 – 6.0]	5.0 [4.0 – 5.8]	4.5 [3.3 – 5.0]
p-Value[7]	*0.001*	*0.001*	*0.001*	*0.655*
p-Value 4 moments[8]	*0.001*	*0.001*[9]	*0.001*[9]	*0.090*[9]

Table 3 *(continued)*

CPOT	n= 72	Sedated[1]	Calm[2]	Agitated[3]
Nurse 1				
RestI	0.0 [0.0 – 0.0]	0.0 [0.0 – 0.0]	0.0 [0.0 – 0.0]	0.5 [0.0 – 1.0]
Oral care	0.0 [0.0 – 1.0]	1.0 [0.0 – 2.0]	0.0 [0.0 – 0.3]	2.5 [0.5 – 4.5]
p-Value[6]	*0.001*	*0.001*	*0.009*	*0.109*
RestII	0.0[0.0 – 0.0]	0.0 [0.0 – 0.0]	0.0 [0.0 – 0.0]	1.5 [0.3 – 3.5]
Turning	2.0[0.0 – 3.0]	2.0 [0.3 – 3.0]	2.0 [0.0 – 3.0]	2.0 [1.5 – 4.5]
p-Value[7]	*0.001*	*0.004*	*0.001*	*0.564*
p-Value 4 moments[8]	*0.001*	*0.001*[9]	*0.001*[9]	*0.116*[9]
Nurse 2				
RestI	0.0 [0.0 – 0.0]	0.0 [0.0 – 0.0]	0.0 [0.0 – 0.0]	0.0 [0.0 – 1.5]
Oral care	0.0 [0.0 – 1.0]	0.0 [0.0 – 2.0]	0.0 [0.0 – 1.0]	3.5 [0.8 – 4.8]
p-Value[6]	*0.001*	*0.006*	*0.008*	*0.109*
RestII	0.0 [0.0 – 0.0]	0.0[0.0 – 0.0]	0.0 [0.0 – 0.0]	1.0 [0.3 – 3.3]
Turning	2.0 [0.3 – 3.0]	2.5[1.0 – 3.0]	1.5 [0.0 – 3.0]	2.0 [0.3 – 3.8]
p-Value[7]	*0.001*	*0.001*	*0.001*	*0.655*
p-Value 4 moments[8]	*0.001*	*0.001*[9]	*0.001* [9]	*0.172*[9]

Values are expressed as medians with interquartile range
1. RASS -5, -4, -3
2. RASS -2, -1, 0
3. RASS 1, 2, 3, 4
4. Nurse 1
5. Nurse 2
6. RestI vs. Oral care analyzed with the Wilcoxon Signed Ranks Test. RASS category of moment Rest1
7. RestII vs. Turning analyzed with the Wilcoxon Signed Ranks Test. RASS category of moment RestII
8. Analyzed with the Friedman Test
9. RASS category of moment RestI
Number of patients RASS category of moment RestI sedated n=30 calm n=38 agitated n=4
Number of patients RASS category of moment RestII sedated n=16 calm n=52 agitated n=4

DISCUSSION

We have shown that both pain assessment tools showed a substantial inter-rater reliability for all assessments. The internal consistency during a painful procedure (turning) was poor to questionable. Discriminant validation was demonstrated by a significant increase in pain scores during turning.

This is the first prospective cohort study which has simultaneously assessed the reliability and discriminant validation of the BPS [9] and CPOT [8;10-12] in intubated mechanically ventilated patients post-cardiac surgery who were unable to self-report pain in a daily clinical setting. Although the BPS and CPOT in our study showed a substantial inter-rater reliability for all assessments, this is lower than numerous previous studies [8;14;15;17;18;22;23;25;41;42].This could partially be the result of considerably lower ICC values during rest (fair to moderate) compared to the procedures (moderate to almost perfect), which is contrary to several previous studies [8;10;21-23;26;43]. ICC values are assumed to be susceptible to between subject variability and a lack of subject variability can cause incorrect low ICC values [34;44]. A

minority of the validation studies have only assessed patients post-cardiac surgery, which were likely more homogeneous populations in terms of diagnosis, age, severity of illness, pain and level of sedation and analgesia than a general ICU population or a mixed medical/surgical population [10;12;13;16;21]. A lack of variability in pain scores during rest could also be the result of post-surgical pain. Patients might have been afraid to move, which could have resulted in fewer movements and as a consequence lower pain scores during rest. The opposite could also have been occurred; due to residual analgesia, patients could have been enough sedated to not experience pain during rest [45]. Our inter-rater reliability could also be lower because the nurses had to assess the BPS and CPOT simultaneously, which is more demanding. An additional reason may be the large number of nurses in our ICU, resulting in less experience with the assessments. Furthermore, the pairs of nurses differed in our study; only the bedside nurse was consistently part of the assessment team. In a majority of the studies with a higher inter-rater reliability, either one of the investigators or physicians participated in the assessments, or the number of assessing nurses was smaller [9;10;14;15;17;18;25;41;42]. Pain assessment by a large group of nurses is, however, a reflection of real life intensive care.

The values of Cronbach's alpha for both scores were <0.70, which indicates that the correlation between the behavioral domains is insufficient to be considered reliable to measure a single construct (pain) [36]. Values for Cronbach's alpha in previous studies assessing the reliability of the CPOT were between 0.31 to 0.81 [17;22;23;31]. Cronbach's alpha in studies assessing the BPS ranged from 0.63 to 0.77 [14;15;19;26; 42]. Although the coefficients in former studies are generally higher, these values should be interpreted with caution. Several studies did not calculate Cronbach's alpha for the sample size but used all paired assessments, which potentially leads to higher coefficient values [15;17;46]. Furthermore, Cronbach's alpha is affected by the number of items and will be higher when a questionnaire is constituted of more items [36]. Finally, an Cronbach's alpha calculated for item response options lower than 7 points will be an underestimation and therefore would benefit of a different approach when calculating the internal consistency [47] . The results of the discriminant validation during painful procedures in the total sample are consistent with previous studies. However, the BPS of nurse1 showed a significant increase of half a point during the presumed non-painful procedure in the total sample.

In a subgroup of sedated patients (RASS -5, -4, -3) the BPS scores of both nurses and the CPOT score of nurse2 increased significantly by one point during the non-painful procedure. Four studies assessed the BPS during a non-painful procedure [9;13;26;31] and two of these also showed a significant increase of the BPS during the non-painful procedure [9]. Several previous studies assessed the CPOT during a non-painful procedure in unconscious patients and none of these showed a similar increase [8;16;18;20;22;23;31]. Only two of these studies presented the CPOT scores during different levels of sedation and agitation [8;31]. A possible explanation could be that nurses have more difficulty assessing pain in patients with higher levels of sedation, thus obscuring response to painful stimuli.

The BPS showed more significant increases during the non-painful procedure than the CPOT. A reflex to stress rather than to pain might be the cause of these increases. Coughing and

straining might also be a reflex due to movement of the endotracheal tube during oral care. Another explanation could be the different number of options in the domains of the BPS and CPOT. The BPS has four options compared to the three options of the CPOT, which might be more difficult to rate. Additionally, the BPS demands the assessment of ventilator waveforms and asynchrony, which could be difficult while observing the patient at the same time [17]. However, the increases during the non-painful procedures did not reach the established cut-off scores for the presence of pain (BPS >4 and CPOT >2) [8;11;18;22-24;48]. This finding needs further investigation since the cut-off for the BPS is established in only one study [48]. The cut-off values of previous BPS and CPOT studies are presented in the appendix.

The median BPS and CPOT scores in the small and underpowered subgroup of agitated patients (RASS +1) were generally higher and increased during all procedures. Discriminating between behaviors as a result of agitation or pain might be difficult. This important issue requires further research since the incidence of delirium after cardiac surgery varies between 3 - 55% and the overall incidence in critically ill patients varies between 30 - 50% [49] .

Finally, our patients received substantially less sedation and analgesics than patients in most previous studies. Since a restricted sedation policy is recommended in recent guidelines, it is to be expected that current and future critically ill patients will receive less sedation [1]. The development of behavioral pain assessment tools, determination of the psychometric properties of these tools and the critical appraisal of research articles should take this development into account.

Limitations

Our study has a number of limitations. First, since the assessors are trained ICU nurses, they were aware of which procedure was potentially painful. Therefore, they might have observed more behavioral changes during turning which could have resulted in higher scores and an inflation of the discriminant validation.

Second, the number of analyzed patients is relatively small compared to the initial number of included patients and this could have induced selection bias. The short postoperative ICU stay of our study patients limited the period in which they could meet the inclusion criteria. This was reflected by the significantly longer ICU stay of the 72 analyzed patients in comparison with the 135 unanalyzed patients.

Finally, we did not officially translate and validate the translated pain scores before the start of the study. The study team, including a qualified English language translator, translated the short descriptions of both tools into Dutch in our CIS, thus language misinterpretations might have occurred. However, the descriptions in both tools are short, clear and contain universally interpretable signs, so we believe it is unlikely that this limitation affected our results.

CONCLUSION

This study in mechanically ventilated patients after cardiac surgery, who were unable to self-report pain, showed that the BPS and CPOT are reliable and valid in a daily clinical setting.

However, the discriminant validation of both scores seems less satisfactory in sedated or agitated patients and this topic requires further investigation.

Acknowledgements

The authors thank R. Koper, R. Peek, M. Koning, the ICU-nurses and information technology specialist A. Bianchi for their contributions to this study, as well as S.Toohey for his English revisions.

REFERENCES

1. Baron R, Binder A, Biniek R, Braune S, et al. Evidence and consensus based guideline for the management of delirium, analgesia, and sedation in intensive care medicine. Revision 2015 (DAS-Guideline 2015) - short version 1. Ger Med Sci 2015;13:Doc19.
2. Barr J, Fraser GL, Puntillo K, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306.
3. Puntillo K, Pasero C, Li D, et al. Evaluation of pain in ICU patients. Chest 2009 Apr;135(4):1069-74.
4. Puntillo KA, Max A, Timsit JF, et al. Determinants of procedural pain intensity in the intensive care unit. The Europain(R) study. Am J Respir Crit Care Med 2014 Jan 1;189(1):39-47.
5. Peng Z, Li H, Zhang C, et al. A retrospective study of chronic post-surgical pain following thoracic surgery: prevalence, risk factors, incidence of neuropathic component, and impact on qualify of life. PLoS One 2014;9(2):e90014.
6. Granja C, Gomes E, Amaro A, et al. Understanding posttraumatic stress disorder-related symptoms after critical care: the early illness amnesia hypothesis. Crit Care Med 2008 Oct;36(10):2801-9.
7. Merskey H. The taxonomy of pain 3. Med Clin North Am 2007 Jan;91(1):13-20, vii.
8. Gelinas C, Johnston C. Pain assessment in the critically ill ventilated adult: validation of the Critical-Care Pain Observation Tool and physiologic indicators. Clin J Pain 2007 Jul;23(6):497-505.
9. Payen JF, Bru O, Bosson JL, et al. Assessing pain in critically ill sedated patients by using a behavioral pain scale. Crit Care Med 2001 Dec;29(12):2258-63.
10. Gelinas C, Fillion L, Puntillo KA, et al. Validation of the critical-care pain observation tool in adult patients. Am J Crit Care 2006 Jul;15(4):420-7.
11. Gelinas C, Fillion L, Puntillo KA. Item selection and content validity of the Critical-Care Pain Observation Tool for non-verbal adults. J Adv Nurs 2009 Jan;65(1):203-16.
12. Gelinas C, Harel F, Fillion L, et al. Sensitivity and specificity of the critical-care pain observation tool for the detection of pain in intubated adults after cardiac surgery. J Pain Symptom Manage 2009 Jan;37(1):58-67.
13. Ahlers SJ, van Gulik L, van der Veen AM, et al. Comparison of different pain scoring systems in critically ill patients in a general ICU. Crit Care 2008;12(1):R15.
14. Ahlers SJ, van der Veen AM, van Dijk M, et al. The use of the Behavioral Pain Scale to assess pain in conscious sedated patients. Anesth Analg 2010 Jan 1;110(1):127-33.
15. Aissaoui Y, Zeggwagh AA, Zekraoui A, Abidi K, Abouqal R. Validation of a behavioral pain scale in critically ill, sedated, and mechanically ventilated patients. Anesth Analg 2005 Nov;101(5):1470-6.
16. Boitor M, Fiola JL, Gelinas C. Validation of the Critical-Care Pain Observation Tool and Vital Signs in Relation to the Sensory and Affective Components of Pain During Mediastinal Tube Removal in Postoperative Cardiac Surgery Intensive Care Unit Adults. J Cardiovasc Nurs 2015 Mar 30.
17. Chanques G, Pohlman A, Kress JP, et al. Psychometric comparison of three behavioural scales for the assessment of pain in critically ill patients unable to self-report. Crit Care 2014;18(5):R160.
18. Echegaray-Benites C, Kapoustina O, Gelinas C. Validation of the use of the Critical-Care Pain Observation Tool (CPOT) with brain surgery patients in the neurosurgical intensive care unit. Intensive Crit Care Nurs 2014 Oct;30(5):257-65.
19. Juarez P, Bach A, Baker M, et al. Comparison of two pain scales for the assessment of pain in the ventilated adult patient. Dimens Crit Care Nurs 2010 Nov;29(6):307-15.

20. Kanji S, MacPhee H, Singh A, et al. Validation of the Critical Care Pain Observation Tool in Critically Ill Patients With Delirium: A Prospective Cohort Study. Crit Care Med 2016 Jan 16.
21. Keane KM. Validity and reliability of the critical care pain observation tool: a replication study. Pain Manag Nurs 2013 Dec;14(4):e216-e225.
22. Li Q, Wan X, Gu C, et al. Pain assessment using the critical-care pain observation tool in Chinese critically ill ventilated adults. J Pain Symptom Manage 2014 Nov;48(5):975-82.
23. Nurnberg DD, Saboonchi F, Sackey PV,et al. A preliminary validation of the Swedish version of the Critical-Care Pain Observation Tool in adults. Acta Anaesthesiol Scand 2011 Apr;55(4):379-86.
24. Stilma W, Rijkenberg S, Feijen HM, et al. Validation of the Dutch version of the critical-care pain observation tool. Nurs Crit Care 2015 Dec 22.
25. Vazquez CM, Pardavila Belio MI, Lucia MM, et al. Evaluation of pain during posture change in patients with invasive mechanical ventilation. Enfermeria Intensiva 2009 Jan;20(1):2-9.
26. Young J, Siffleet J, Nikoletti S,et al. Use of a Behavioural Pain Scale to assess pain in ventilated, unconscious and/or sedated patients. Intensive Crit Care Nurs 2006 Feb;22(1):32-9.
27. Chanques G, Sebbane M, Barbotte E, et al. A prospective study of pain at rest: incidence and characteristics of an unrecognized symptom in surgical and trauma versus medical intensive care unit patients. Anesthesiology 2007 Nov;107(5):858-60.
28. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med 2002 Nov 15;166(10):1338-44.
29. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). JAMA 2001 Dec 5;286(21):2703-10.
30. Punt C, Kreutz M, Dekkers P, et al. Comparison of intravenous boluses of piritraminde and morphine. Did we use the correct ratio of analgetic potency? Eur J Anaesthesiol 29, 196. 2014.
31. Rijkenberg S, Stilma W, Endeman H, et al. Pain measurement in mechanically ventilated critically ill patients: Behavioral Pain Scale versus Critical-Care Pain Observation Tool. J Crit Care 2015 Feb;30(1):167-72.
32. Puntillo KA, Morris AB, Thompson CL, et al. Pain behaviors observed during six common procedures: results from Thunder Project II. Crit Care Med 2004 Feb;32(2):421-7.
33. CPOT instruction video made by nursing pathways. http://nursingpathways.kp.org Accessed: April 2010
34. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-8.
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-74.
36. Streiner D.L, Norman G.R. Health measurement scales a practical guide to their development and use. third edition ed. Hamilton: Oxford university press; 2003.
37. Leech NL, Barret KC, Morgan GA. Repeated-measures and mixed anovas. IBM SPSS for Intermediate Statistics Use and Interpretation. New York: Taylor & Francis Group; 2011, p. 175-84.
38. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. Stat Med 2001 Nov 15;20(21):3205-14.
39. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.
40. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998 Jan 15;17(1):101-10.
41. Buttes P, Keal G, Cronin SN, S et al. Validation of the critical-care pain observation tool in adult critically ill patients. Dimens Crit Care Nurs 2014 Mar;33(2):78-81.
42. Chen J, Lu Q, Wu XY, et al. Reliability and validity of the Chinese version of the behavioral pain scale in intubated and non-intubated critically ill patients: Two cross-sectional studies. Int J Nurs Stud 2016 May 27;61:63-71.
43. Olsen BF, Rustoen T, Sandvik L, et al. Implementation of a pain management algorithm in intensive care units and evaluation of nurses' level of adherence with the algorithm. Heart Lung 2015 Nov;44(6):528-33.
44. Lee KM, Lee J, Chung CY, et al. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. Clin Orthop Surg 2012 Jun;4(2):149-55.

45. Gelinas C, Puntillo KA, Joffe AM, et al. A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Seminars in Respiratory and Critical Care Medicine 2013 Apr;34(2):153-68.
46. Paulson-Conger M, Leske J, Mai dL C, et al. Comparison of two pain assessment tools in nonverbal critical care patients. Pain Manag Nurs 2011 Dec;12(4):218-24.
47. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research & Evaluation 2012;17(3):1-13.
48. Payen JF, Chanques G, Mantz J, et al. Current practices in sedation and analgesia for mechanically ventilated critically ill patients: a prospective multicenter patient-based study. Anesthesiology 2007 Apr;106(4):687-95.
49. van den Boogaard M, Schoonhoven L, van der Hoeven JG, et al. Incidence and short-term consequences of delirium in critically ill patients: A prospective observational cohort study. Int J Nurs Stud 2012 Jul;49(7):775-83.

APPENDIX

Behavioral Pain Scale and Critical-Care Pain Observation Tool (appendix chapter 2)

Statistical analysis of cut-off values CPOT and BPS

CPOT Study/ Setting	Participants	Procedure	Cut- off	Sensitivity	Specificity	Positive predictive value	LR[1]	AUC[2]	Analysis
Gelinas et al. 2007[a]/ ICU	n= 55 (30 conscious) ventilated adults	Nociceptive (turning)	3	66.7%	83.3%	85.7%	NA[3]	0.789	ROC[4]
Gelinas et al. 2009[b]/ ICU	n= 99 ventilated and extubated adults post cardiac surgery	Nociceptive (turning)	2	86.1%	77.8%		+: 3.9 -: 0.2	0.838	ROC
		Non nociceptive (pre turning)	1	47.2%	82.9%		+: 2.8 -: 0.6	0.677	
		Non nociceptive (post turning)	1	63.0%	97.4%		+: 24.6 -: 0.4	0.836	
Nurnberg et al. 2011[c] / ICU	n= 40 ventilated medical/ surgical adults (N=10 able to rate pain)	Unclear	3	NA	NA	NA	NA	NA	Logistic regression
Li et al. 2014[d]/ ICU	n= 63 conscious ventilated adults	Nociceptive (turning)	2	80.8%	73.3%	NA	NA	0.899	ROC
		Non nociceptive (blood pressure)	2	88.2%	74.6%			0.902	
Echegaray et al. 2014[e]/ Neuro ICU	n= 43 conscious elective brain surgery	Nociceptive (turning)	1.5	76.9%	73.3%	NA	NA	0.864	ROC
Stilma et al. 2015[f]/ ICU	n= 108 ventilated medical/ surgical adults able to self report	Nociceptive (turning) & Non nociceptive (rest)	2	39%	85%	NA	NA	0.651	ROC
		Nociceptive	2	51.1%	76.2%	NA	NA	0.698	

Joffe et al. 2016[g] ICU	n= 79 Brain injury adults (n= 28 able to self report)	Nociceptive (turning)	2	90%	67%	NA	NA	0.72	ROC
Severgnini et al. 2016[h] ICU	n= 101 conscious critically ill patients (n= 41)	Nociceptive (turning)	2	76.5%	70.8%	NA	NA	0.80	ROC
		Non nociceptive (pre turning)	2	25%	91.3%			0.57	
		Non nociceptive (after turning)	2	33.3%	60.8%			0.50	

BPS Study/ Setting	**Participants**	**Procedure**	**Cut-off**	**Sensitivity**	**Specificity**	**Positive predictive value**	**LR[1]**	**AUC[2]**	**Analysis**
Payen et al. 2007[i] ICU	n= 1381 adults ICU	Nociceptive (endotracheal suctioning/ mobilization)	4						
Severgnini et al. 2016 ICU	n= 101 ICU patients conscious (n= 41)	Nociceptive (turning)	5	62.8%	91.7%	NA	NA	0.83	ROC
		Non nociceptive (pre turning)	5	79.2%	61.2%			0.71	
		Non nociceptive (after turning)	5	62.5%	60.8%			0.60	

1. Likelihood ratio +: positive -: negative 2. Area under the curve 3. Not assessed 4. Receiver operating characteristic curve analysis

a. Gelinas C Johnston C: Pain assessment in the critically ill ventilated adult: validation of the Critical-Care Pain Observation Tool and physiologic indicators. Clin J Pain 2007; 23:497-505
b. Gelinas C et al. Sensitivity and specificity of the critical-care pain observation tool for the detection of pain in intubated adults after cardiac surgery. J Pain Symptom Manage 2009 Jan;37(1):58-67
c. Nurnberg DD, et al.: A preliminary validation of the Swedish version of the Critical-Care Pain Observation Tool in adults. Acta Anaesthesiol Scand. 2011; 55:379-386.
d. Li Q et al. Pain assessment using the critical-care pain observation tool in Chinese critically ill ventilated adults. J Pain Symptom Manage 2014 Nov;48(5):975-82.
e. Echegaray-Benites et al. Validation of the use of the Critical-Care Pain Observation Tool (CPOT) with brain surgery patients in the neurosurgical intensive care unit. Intensive Crit Care Nurs 2014 Oct;30(5):257-65
f. Stilma W et al. Validation of the Dutch version of the critical-care pain observation tool. Nurs Crit Care 2015 Dec 22
g. Joffe AM et al. Validation of the Critical-Care Pain Observation Tool in brain-injured critically ill adults. J Crit. Care 2016, 36 76–80
h. Severgnini P et al. Accuracy of Critical Care Pain Observation Tool and Behavioral Pain Scale to assess pain in critically ill conscious and unconscious patients: prospective, observational study. Journal of Intensive Care 2016, 4:68
i. Payen JF et al: Current practices in sedation and analgesia for mechanically ventilated critically ill patients: a prospective multicenter patient-based study. Anesthesiology 4 2007, Vol.106, 687/69

CHAPTER 4

Validation of the Dutch version of the Critical-Care Pain Observation Tool

ABSTRACT

Introduction

Systematic assessment of pain is necessary for adequate treatment of pain. Patient self-reported pain is a superior assessment but is of limited use for intubated patients in the intensive care unit. For these patients, the critical-care pain observation tool (CPOT) has been developed. The aim of this cross-sectional observational study was to perform a validation of the Dutch CPOT.

Materials and Methods

The Dutch translation of the CPOT was used. Clinimetric characteristics were analyzed in a cross-sectional design. Internal consistency (Cronbach's alpha) was tested by collecting CPOT scores in patients at rest and during turning. Inter-rater reliability was tested by collecting CPOT scores simultaneously by two different nurses who were blinded to each other's scores. Criterion validity (area under the curve, sensitivity and specificity) of the Dutch CPOT(index test) was analyzed using patient self-reported pain (reference test).

Results

Cronbach's alpha was 0.56. During rest, the inter-rater reliability was 0.38 95% confidence interval (CI): 0.20 – 0.53. During turning, the inter-rater reliability was 0.56 (95% CI: 0.42-0.68; area under the curve=0.65, 95% CI 0.57 – 0.73). At a threshold CPOT score of 2, the sensitivity and specificity were 39% and 85%, respectively.

Conclusions

The Dutch CPOT is available for pain assessment in intubated patients unable to self-report. Inter-rater reliability is moderate. At the threshold, a CPOT score of 2, the sensitivity was 39% and the specificity of 85%. The CPOT is easy to use for systematic assessment of pain. Additional information about the threshold is valuable for use in daily practice.

INTRODUCTION

Pain assessment in intensive care patients is a challenge. Pain is reported in nearly 50% of intubated adult intensive care patients, and nurses underestimate patient pain in 35 – 55% of the cases [33,25,31]. Systematic assessment of pain is necessary for adequate treatment and is associated with decreased pain and complications, such as inadequate sleep, disorientation and prolonged sedation [11,6,33]. Patient self-reported pain is a superior method of assessment, but this method is of limited use for most intensive care patients because they are intubated, severely ill or have a decreased level of consciousness [22,4].

Different pain assessment tools have been created for these patients [23,28]. Five scales include facial expressions and behavioral aspects, and three scales include physiological indicators as well [29]. Physiological indicators have been found to be unreliable for pain assessment in intensive care patients due to illness, medication and the tendency to adapt to the presence of pain [28]. Consequently, guidelines advise the use of the critical-care pain observation tool (CPOT) or the behavioral pain scale (BPS) for intubated intensive care patients [3,36]. We selected the CPOT because it has demonstrated high inter-rater reliability [15,16]. Additionally, the CPOT is an easy-to-use tool that provides uniform pain assessment in intensive care patients unable to self-report [12]. Thus far, the CPOT is available in English, French, Spanish, Swedish and Korean [13-16,21,24,39] but has not been officially translated into Dutch or validated in a Dutch population. In addition to the value of this study to Dutch critical care, it is also of relevance internationally as the concept of using CPOT as a pain assessment tool is increasingly being adopted both within and outside Europe as is evidenced by the number of languages into which it has been translated.

MATERIALS AND METHODS

The aim of this study is to perform a cross-cultural validation of the Dutch version of the CPOT in intubated adult intensive care patients.

Design

To perform a cross-cultural validation of the Dutch version of the CPOT, we used a cross-sectional design. This design makes is possible to compare the CPOT and the reference standard simultaneously, which is essential for pain assessment in a dynamic environment like the intensive care unit (ICU).

Setting and sample

The study was performed in a 22-bed mixed ICU in the Netherlands. Approximately 1500 patients are admitted to the ICU yearly, 40% of whom undergo cardio-thoracic surgery. The medical staff consists of senior and junior intensivists and intensive care nurses. Pain assessment is determined at the bedside based on a variety of observations by health care professionals, including patient behavior and self-reports. The policy is to keep patients awake as much as possible unless this interferes with necessary mechanical ventilation or any other

treatment. As the majority of the patients are awake, it is often possible to communicate with them, even if they are on mechanical ventilation. All patients admitted to the ICU in the first 6months of 2013 were eligible for inclusion in this study. Inclusion criteria were as follows: (a) minimum age of 18 years, (b) intubated for mechanical ventilation, (c) awake and able to answer simple YES/NO questions and (d) able to move their arms and locate pain. Exclusion criteria were as follows: (a) use of neuromuscular blocking agents, (b) being investigated for brain death, (c) impossibility of turning patient due to instability or treatment procedures, (d) presence of delirium and (e) Glasgow Coma Scale score below 9. We aimed to include 100 patients according to the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) manual [38].

Data collection tools and methods

CPOT

The CPOT has been developed for the assessment of pain in intubated adult intensive care patients. It was first validated in patients who had undergone cardio-thoracic surgery and later in a patient group with a medical diagnosis such as sepsis The scale includes four behavioral domains: (a) facial expression, (b) body movements, (c) muscle tension and (d) compliance with the ventilator for intubated patients or vocalization for non-intubated patients. All domains are scored by observing the patient, and muscle tension is examined by passive flexion of the patient's underarm [13-16]. Each domain is scored as 0, 1 or 2. The total score ranges from 0 (no pain) to 8 (maximum score). The CPOT only indicates the presence of pain, not its severity. Initially, the suggested threshold for the presence of pain was a CPOT score >3 [13]. Later, a threshold CPOT score >2 was adopted [15,16]. The validity of these thresholds could not be confirmed [24]. The CPOT is presented in Appendix A.

Dutch translation

According to the COSMIN checklist, an official translation includes: (a) a double forward and backward translation, (b) establishment of face validity of the items and comprehensiveness in a multidisciplinary expert committee, (c) testing of the final translation in practice and (d) documentation of the process [38].To create a numerical method to compare the translations, translators were asked to rate the difficulty of translating the items of the CPOT. Rating scores were divided as follows: (a) impossible to do, (b) extremely difficult, (c) moderately difficult, (d) a little bit difficult and (e) not at all difficult.

Training nurses

Before the start of the study, training was organized by four nurses who had read three articles explaining the CPOT [13-16] watched a video instruction [8] and had exercised assessments of the CPOT and discussed difficulties during a whole day. These expert nurses trained a team of 125 nurses in the Dutch CPOT through a 90-min training session. In addition, the expert nurses provided ongoing bedside support. The complete Dutch CPOT

and instructions were available at the bedside and on the intranet.

Clinimetric characteristics

Three clinimetric characteristics of the CPOT were determined according to the COSMIN criteria [38]. First, the internal consistency of the CPOT was analyzed. We collected CPOT scores when the patient was at rest and while turning the patient, as turning has been proven to be a painful procedure [30]. Second, inter-rater reliability was calculated, for which the CPOT scores (again during rest and while turning) were collected simultaneously by two different nurses who were blinded to each other's scores. Finally, criterion validity of the index test, the Dutch CPOT, was analyzed by using patient self-reports of pain as a reference test [4,7,22]. Self-reports were obtained by asking the question 'Do you have pain?', and the patient could answer by moving his/her head 'YES' or 'NO' [22].

Measurements and study procedures

Demographic characteristics of the patients, administered analgesics and sedatives, type of admission, length of stay (LOS), severity of illness scores, Acute Physiology and Chronic Health Evaluation (APACHE II and IV) [20] and Sepsis-related Organ Failure Assessment (SOFA) [40] were extracted from the clinical information
system [19].

In order to measure the level of awareness, we used the Richmond Agitation-Sedation Scale (RASS) [34]. Self-reports required a minimum RASS score of –2. When RASS scores were below –3, the neurological level of functioning was measured by the Glasgow Coma Scale [32]. The RASS score was measured six times a day. To rule out delirium, which could interfere with the reliability of the CPOT, the presence of delirium was tested at least twice a day using the Confusement Assessment Method [9,10]. Patients with a positive CAM-ICU result were excluded from the analysis. All measurements were part of daily care.

The order and instructions for study measurements appeared automatically in the clinical information system when the patient was eligible for the study according to the inclusion and exclusion criteria [19]. The timing of the measurements depended on care procedures and was planned by the nurse. When the patient had not experienced a painful procedure for 30 min, the CPOT was first scored simultaneously by two nurses blinded to each other's scores. Subsequently, the nurse who was taking care of the patient asked and registered the patient's self-report. In a direct sequence, the patient was turned. During turning, the CPOT scores were assessed and registered in the same way. Measurement procedures were only performed once per patient. All measurements were recorded bedside in the clinical information system.

Data analysis

Internal consistency was calculated by Cronbach's alpha of the CPOT scores at rest and during turning. The scores collected by the nurse who had taken care of the patient were used for this analysis. Reliability analysis was performed by the one-way random agreement intraclass correlation coefficient (ICC) because sample measurements were performed by a random set

of the 125 ICU nurses [35]. We analyzed the CPOT scores when the patient was at rest and during turning, separately.

Criterion validity was analyzed by calculating the sensitivity and specificity for the CPOT scores using self-reports as the reference standard. We determined the sensitivity and specificity for the threshold scores of 2 and 3. A receiver operating curve (ROC) was made for the CPOT scores, and 95% confidence interval (CI) was calculated. Missing data were excluded from the analysis as were data that were not obtained in accordance with the instructions. All data were collected anonymously between November 2012 and August 2013 and analyzed.

All statistical analysis was performed using SPPS software (PAWS statistics version 18.0, Chicago, IL, USA).

Ethical considerations

The Medical Ethics Review Board of our hospital gave this study its approval (MEC nr WO 12.078) and waived the requirement for written informed consent because the study did not require any deviation from the routine standard care.

RESULTS

Dutch translation of the CPOT

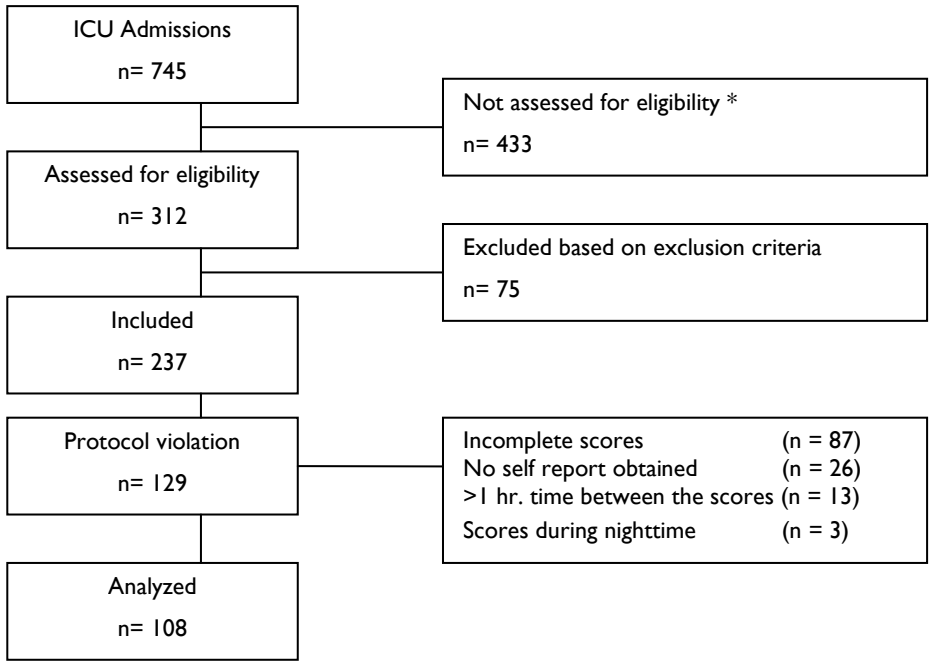
With the permission of the original author, we translated the English CPOT into Dutch (email and contract with Gelinas dated 9 October 2012). A double forward translation was performed by two professional translators, one of whom had a clinical background as an intensive care nurse. A synthesis of the translations was performed with the translators, according to common consensus, and documented. A backward translation was made by one professional translator unaware of the original CPOT and with mother tongue as English. Translators did not find the translation difficult. Ratings varied between 3 (moderately difficult) and 5 (not at all difficult) with a slight difference in the mean scores: translator 1 (PR) 4.5, translator 2 (LvdW) 4.3 and translator 3 (DL) 4.5.

All the translations and face validity were discussed by an expert committee, resulting in a final translation. The expert committee consisted of health professionals, methodologists and language professionals. All discussions and changes were documented. The final translation and the documentation of the translation process were sent to the original author. The Dutch translation was pre-tested by 20 intensive care nurses on 20 patients to check the interpretation and ease of comprehension. No changes were necessary (Appendix).

Study sample

One hundred eight patients were included in the study (Figure 1). Patients were admitted for medical (25%) or surgical (75%) reasons. The level of awareness (RASS scores) varied among patients: light sedation (24%), drowsy (36%), calm and alert (34%) and restless (7%). The

median severity of illness as measured by the APACHE II score was 16 (range, 6 – 34), and the median SOFA score was 6 (range, 6 – 14). Before measurement procedures, 6% of the patients had received sedation and 12% had received analgesics. Baseline characteristics are presented in Table 1.



* Due to high workload and difficulty to organize the required study measurements. The possibility of selection bias was checked. There was a only significant difference between groups in length of stay (LOS).

Figure 1. Flow chart

Clinimetric characteristics of the CPOT

Internal consistency, analyzed by Cronbach's alpha, was 0.56 for the CPOT scores during rest and turning. The items 'facial expression', 'movement' and 'muscle tension' are correlated with the final CPOT score because deleting one of these items resulted in a lower Cronbach's alpha. The item 'compliance with the ventilator' is less correlated with the final score. The corrected item's total correlation was 0.15, and Cronbach's alpha increased to 0.60 when this item was deleted. The results of the internal consistency are presented in Table 2.

Reliability analysis was performed using the one-way random ICC during rest and turning [35]. During rest, the ICC was 0.38 (95% CI: 0.20 – 0.53), indicating fair inter-rater reliability. During turning, the ICC was 0.56 (95% CI: 0.42 – 0.68), indicating moderate inter-rater reliability.

Criterion validity analysis of the CPOT was based on patient self-reports. During 216 measurements, pain was reported in 65 cases. The Receiver Operating Characteristic (ROC) of the scores during rest and turning produced an Area under the curve (AUC) of 0.65 (95% CI: 0.57 – 0.73).We calculated a sensitivity of 39% and a specificity of 85% for a threshold CPOT score of 2. For a threshold CPOT score of 3, the sensitivity and specificity were 20% and 92%, respectively. We conclude that based on the ROC (Figure 2), a CPOT score of 2 could be chosen as the threshold.

DISCUSSION

The internal consistency presented as Cronbach's alpha was 0.56. This is congruent with previous research that reported an internal consistency of the CPOT with a Cronbach's alpha between 0.31 and 0.81 [24]. Our results revealed a low correlation of the item 'compliance with the ventilator'. Although Streiner advises a minimum correlation of 0.20 for an item to be included [37], we do not suggest the deletion of this item because it is part of the original scale. These results could not be compared with other studies because the correlation of the items with the final score has not been published previously [15-17].

We observed a fair to moderate ICC, which is lower than the results in other [13-16,24]. Some studies have used a fixed or limited set of raters, which might be an explanation for the higher ICCs. Our results can be explained by the fact that we analyzed the CPOT scores for tests performed by various different nurses from a larger team of 125 nurses. Although the results of our study demonstrate lower reliability, they are more generalizable because the study settings reflect daily practice at an ICU. Another explanation for the lower ICCs might be the fact that most CPOT scores were 0. This lack of variance might have hampered the ability to determine higher ICCs. Also, ICCs might have been higher if we had assessed all the nurses before the study on their ability to score the CPOT. Differences in CPOT observations by nurses and self-reports could also be influenced by other aspects. For example, studies have mentioned that nurses underestimate their patients' pain [2, 27]. On the other hand, high levels of vicarious exposure biases judgments of pain [26]. The influence of these aspects could not be excluded from this analysis.

Criterion validity of the CPOT at different thresholds was analyzed. The sensitivity of the CPOT is 20% at a threshold score of 3 and 39% at a threshold score of 2. Gélinas et al. report a sensitivity of 66.7% at a threshold score of 3, which is higher than our results [13]. Another publication by the same author reported a sensitivity of 86.1% at a threshold score of 2 [15,16]. It is remarkable that we observed low sensitivity rates of the CPOT. For use in daily clinical practice, patient self-reports will be superior. The CPOT should only be used when patients are not able to self-report. Nurses must remain vigilant for other situations that could lead to higher CPOT scores.

Limitations

In the translation process, we used a double forward but a single backward translation, which might be considered a limitation. The value of a double backward translation has been discussed, and the quality of the translator is considered the most important component [1,18]. Criteria for the quality of the translators are as follows: linguistically competent, fully briefed, experienced in the field and able to comment on their own version [1,18]. We have fulfilled all these criteria.

A second limitation of this study might be the number of patients excluded because of protocol violations. Reasons for uncompleted or unperformed scores could be (a) the high workload during the study period, which might have hampered the organization of the study measurements, (b) an insufficiently short timeframe between the arrival at the ICU and the extubation of the patient and (c) study instructions that might have been too complicated for the team of 125 nurses, despite the availability of information and ongoing support given by the expert nurses. We did not experience any technical problems with the computerized data entry of the CPOT score during the study. We did analyze the possibility of selection bias by comparing baseline characteristics of the analyzed patients with the excluded patients. Differences were observed only in the length of stay (LOS). Median LOS was 26 h in this group compared with a median 34 h in the included patient group (p= 0.009). This difference is consistent with the assumption that during the first 24 h, nurses were not able to complete the measurements for the study because of their workload and the short timeframe between the arrival of the patient and the extubation.

Implications and recommendations for practice

Further research is needed to study the inter-rater reliability and internal consistency of the CPOT. Additionally, further research on patients who are unable to self-report is needed. According to the original exclusion criteria, patients with a positive CAM-ICU or delirium were excluded. This limits the practical value of the CPOT because there is an incidence of delirium of up to 87% in ICU patients [10]. As the reliability of the CAM-ICU varies [10], it is possible that some patients were missed in the diagnosis of delirium, which influenced the study results. We suggest that further research of the use of the CPOT is needed for this patient group.In order to enhance pain assessment in the ICU, further development of education concerning pain and pain assessment in the ICU is necessary to maintain awareness among the staff. Additionally, the development of computer assistance in CPOT scores could support ongoing pain measurement during several shifts. Our results reflect an ICU where most patients are awake during mechanical ventilation. In an ICU where sedation is more common, the CPOT scores and the threshold could be different.

CONCLUSION

A Dutch version of the CPOT is available for daily clinical practice. We observed fair reliability in measurements when a patient was at rest and moderate reliability during turning. We observed a sensitivity of 39% and a specificity of 85% at a CPOT score threshold of 2. We recommend further research to study the validity in patients with delirium or patients who are not able to self-report.

Acknowledgements

The authors are extremely grateful for the support and effort of A. Bianchi, R.J. Bosman, M. Koning, R. Koper, R. Peek, P. Rep, L. van der Wouw and all the ICU nurses in our team.

Table 1. Baseline characteristics

Variables (n= 108)		
Age	median (range)	70 (24 - 91)
Male gender		73 (68.2)
Type of admission		
Medical		27 (25.2)
Surgery		80 (74.8)
Apache IV PM[1]	median (range)	0.4 (0 - 1)
Apache II PM[1]	median (range)	16 (6 - 34)
SOFA[2]	median (range)	6 (0 - 14)
LOS in hours[3]	median (range)	34 (15 - 1162)
RASS[4]		
Score 1 (restless)		7 (6.5)
Score 0 (calm and alert),		34 (31.8)
Score -1 (drowsy)		39 (36.4)
Score -2 (light sedation),		24 (22.4)
Analgesics administered 1 hr. before	no./total no. (%)	12 (11.1)
Sedatives administered 1 hr. before	no./total no. (%)	6 (5.6)

[1] Apache IV or II PM: Acute Physiology And Chronic Health Evaluation
[2] SOFA: Sepsis-Related Organ Failure Assessment
[3] LOS: Length of Stay
[4] RASS: Richmond Agitation Sedation Scale

Table 2a. Internal consistency

	Item-total correlation		
CPOT domains	**Scale mean if item deleted**	**Corrected item-total correlation**	**Cronbach's alpha if item deleted**
Face	.54	.411	.417
Movement	.61	.343	.483
Muscle tension	.51	.461	.373
Compliance ventilator	.70	.151	.599

Table 2b. Internal consistency

	Inter-item correlation			
CPOT domains	**Face**	**Movement**	**Muscle tension**	**Compliance ventilator**
Face	1.00	.230	.414	.177
Movement	.230	1.00	.361	.091
Muscle tension	.414	.361	1.00	.066
Compliance ventilator	.177	.091	.066	1.00

Table 3. Inter-rater reliability

Assessment	**ICC[1] (95%CI[2])**
Rest	0.38 [0.20 – 0.53]
Turning	0.56 [0.42 – 0.68]

[1] Intraclass correlation coefficients (one way random) [2] Confidence interval

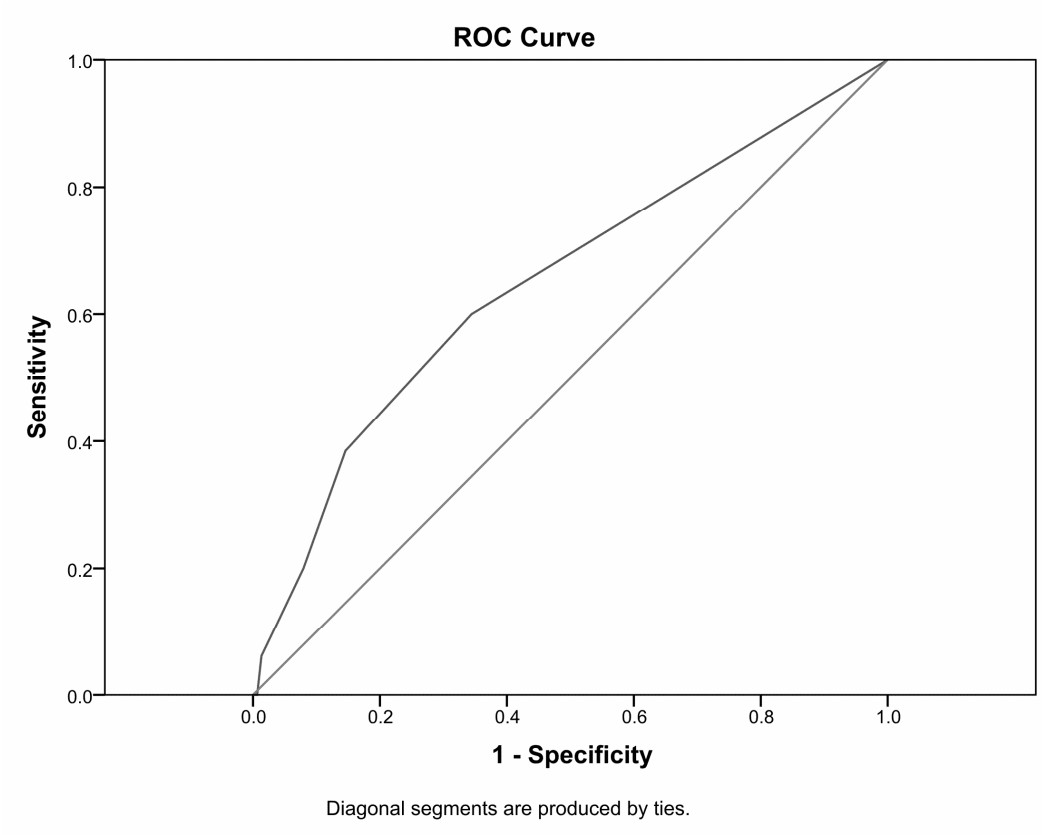


Figure 2. ROC Curve

REFERENCES

1. Acquadro C, Conway K, Hareendran A, Aaronson N. (2008). Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health; 11: 509-521.
2. Aslan FE, Badir A, Selimen D. (2003). How do intensive care nurses assess patients' pain? Nursing in Critical Care; 8: 62-67.
3. Barr J, Fraser GL, Puntillo K, Ely EW, Gelinas C, Dasta JF, Davidson JE, Devlin JW, Kress JP, Joffe AM, Coursin DB, Herr DL, Tung A, Robinson BR, Fontaine DK, Ramsay MA, Riker RR, Sessler CN, Pun B, Skrobik Y, Jaeschke R. (2013). Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Critical Care Medicine; 41: 263-306.
4. Breivik H, Borchgrevink PC, Allen SM, Rosseland LA, Romundstad L, Hals EK, Kvarstein G,
5. Stubhaug A. (2008). Assessment of pain. British Journal of Anaesthesia; 101: 17-24.
6. Chanques G, Jaber S, Barbotte E, Violet S, Sebbane M, Perrigault PF, Mann C, Lefrant JY, Eledjam JJ. (2006). Impact of systematic evaluation of pain and agitation in an intensive care unit. Critical Care Medicine; 34: 1691-1699.
7. Chanques G, Viel E, Constantin JM, Jung B, de Lattre S, Carr J, Cisse M, Lefrant JY, Jaber S. (2010). The measurement of pain in intensive care unit: comparison of 5 self-report intensity scales. Pain; 151: 711-721.
8. CPOT (2006). CPOT instruction video make by nursing pathways. http://nursingpathways.kp.org/national/learning/webvideo/resources/cpot/.
9. van Eijk MM, van Marum RJ, Klijn IA, de Wit N, Kesecioglu J, Slooter AJ. (2009). Comparison of delirium assessment tools in a mixed intensive care unit. Critical Care Medicine; 37: 1881-1885.
10. van Eijk MM, van den BM, van Marum RJ, Benner P, Eikelenboom P, Honing ML, van der HB, Horn J, Izaks GJ, Kalf A, Karakus A, Klijn IA, Kuiper MA, de Leeuw FE, de Man T, van der Mast RC, Osse RJ, de Rooij SE, Spronk PE, van der Voort PH, van Gool WA, Slooter AJ. (2011). Routine use of the confusion assessment method for the intensive care unit: a multicenter study. American Journal Respiratory and Critical Care Medicine; 184: 340-344.
11. Fraser GL, Riker RR. (2001). Monitoring sedation, agitation, analgesia, and delirium in critically ill adult patients. Critical Care Clinics; 17: 967-987.
12. Gelinas C. (2010). Nurses' evaluations of the feasibility and the clinical utility of the Critical-Care Pain Observation Tool. Pain Management Nursing; 11: 115-125.
13. Gelinas C, Johnston C. (2007). Pain assessment in the critically ill ventilated adult: validation of the Critical-Care Pain Observation Tool and physiologic indicators. Clinical Journal of Pain; 23: 497-505.
14. Gelinas C, Fillion L, Puntillo KA, Viens C, Fortier M. (2006). Validation of the critical-care pain observation tool in adult patients. American Journal of Critical Care; 15: 420-427.
15. Gelinas C, Fillion L, Puntillo KA. (2009). Item selection and content validity of the Critical-Care Pain Observation Tool for non-verbal adults. Journal of Advanced Nursing; 65: 203-216.
16. Gelinas C, Harel F, Fillion L, Puntillo KA, Johnston CC. (2009). Sensitivity and specificity of the critical-care pain observation tool for the detection of pain in intubated adults after cardiac surgery. Journal of Pain and Symptom Management; 37: 58-67.
17. Gelinas C, Puntillo KA, Joffe AM, Barr J. (2013). A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Seminars in Respiratory and Critical Care Medicine; 34: 153-168.
18. Guillemin F, Bombardier C, Beaton D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. Journal of Clinical Epidemiology; 46: 1417-1432.
19. iMD-Soft; Metavision, Tel Aviv, Israel. 2013.
20. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. (1985). APACHE II: a severity of disease classification system. Critical Care Medicine; 13: 818-829.
21. Kwak EM, Oh H. (2012). Validation of a Korean translated version of the Critical Care Pain Observation Tool (CPOT) for ICU patients. Journal of Korean Academy of Nursing; 42: 76-84.
22. Kwekkeboom KL, Herr K. (2001). Assessment of pain in the critically ill. Critical Care Nursing Clinics of North America; 13: 181-194.

23. Li D, Puntillo K, Miaskowski C. (2008). A review of objective pain measures for use with critical care adult patients unable to self-report. The Journal of Pain; 9: 2-10.
24. Nurnberg DD, Saboonchi F, Sackey PV, Bjorling G. (2011). A preliminary validation of the Swedish version of the Critical-Care Pain Observation Tool in adults. Acta Anaesthesiologica Scandinavica; 55: 379-386.
25. Payen JF, Bosson JL, Chanques G, Mantz J, Labarere J. (2009). Pain assessment is associated with decreased duration of mechanical ventilation in the intensive care unit: a post Hoc analysis of the DOLOREA study. Anesthesiology; 111: 1308-1316.
26. Prkachin KM, Rocha EM. (2010). High levels of vicarious exposure bias pain judgments. The Journal of Pain; 11: 904-909.
27. Prkachin KM, Solomon PE, Ross J. (2007). Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. Canadian Journal of Nursing Research; 39: 88-106.
28. Pudas-Tahka SM, Axelin A, Aantaa R, Lund V, Salantera S. (2009). Pain assessment tools for unconscious or sedated intensive care patients: a systematic review. Journal of Advanced Nursing; 65: 946-956.
29. Puntillo KA, White C, Morris AB, Perdue ST, Stanik-Hutt J, Thompson CL, Wild LR. (2001). Patients' perceptions and responses to procedural pain: results from Thunder Project II. American Journal of Critical Care; 10: 238-251.
30. Puntillo KA, Morris AB, Thompson CL, Stanik-Hutt J, White CA, Wild LR. (2004). Pain behaviors observed during six common procedures: results from Thunder Project II. Critical Care Medicine; 32: 421-427.
31. Puntillo K, Pasero C, Li D, Mularski RA, Grap MJ, Erstad BL, Varkey B, Gilbert HC, Medina J, Sessler CN. (2009). Evaluation of pain in ICU patients. Chest; 135: 1069-1074.
32. Rowley G, Fielding K. (1991). Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. Lancet; 337: 535-538.
33. Sessler CN, Wilhelm W. (2008). Analgesia and sedation in the intensive care unit: an overview of the issues. Critical Care; 12 Suppl 3: S1.
34. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA, Tesoro EP, Elswick RK. (2002). The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. American Journal of Respiratory and Critical Care Medicine; 166: 1338-1344.
35. Shrout PE, Fleiss JL. (1979). Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin; 86: 420-428.
36. Spijkstra JJ, Horn J, Gielen-Wijffels SEMJ, Burger D, van den Berg B, Snellen FTF. (2013) Revised guideline analgesia and sedation for adult intensive care patients, Utrecht 2010. (in Dutch). http://nvic.nl/sites/default/files/richtlijn%20sedatie.pdf (accessed 19/07/13).
37. Streiner D.L, Norman G.R. (2008). Selecting the items., Health measurement scales. Hamilton: Oxford university press. p 87.
38. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research; 21: 651-657.
39. Vazquez CM, Pardavila Belio MI, Lucia MM, Aguado LY, Margall Coscojuela MA, Asiain Erro MC. (2009). Evaluation of pain during posture change in patients with invasive mechanical ventilation. Enfermeria Intensiva; 20: 2-9.
40. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, Reinhart CK, Suter PM, Thijs LG. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Medicine; 22: 707-710.

APPENDIX Nederlandse Critical-Care Pain Observation Tool (CPOT-NL)

Indicator	Beschrijving	Score	
Gezichtsuitdrukking	Geen spierspanning waargenomen	Ontspannen, neutraal	0
	Aanwezig zijn van fronsend voorhoofd, naar beneden getrokken wenkbrauwen, dichtknijpen van de ogen en contractie van de gezichtsspieren of een andere verandering (bijv. openen van ogen of tranende ogen tijdens pijnlijke procedures)	Gespannen	1
	Alle bovenstaande gezichtsbewegingen plus ogen stijf dichtgeknepen (de patiënt heeft mogelijk de mond open of bijt op de endotracheale tube)	Grimassen	2
Lichaamsbewegingen	Beweegt helemaal niet (betekent niet noodzakelijkerwijs dat pijn afwezig is) of normale houding (bewegingen zijn niet gericht naar de plek waar de pijn zit of met het doel deze te beschermen)	Beweging afwezig of normale houding	0
	Langzame, voorzichtige bewegingen, aanraken of wrijven over pijnlijke plek, aandacht vragen met bewegingen.	Beschermend	1
	Trekken aan tube, pogingen om rechtop te zitten, bewegen van ledematen/woelen, niet opvolgen van opdrachten, slaan naar personeel, proberen uit bed te klimmen.	Rusteloosheid/ agitatie	2
Spierspanning (Beoordeling middels passief buigen en strekken van bovenste ledematen wanneer patiënt in rust is of wanneer patiënt wordt gedraaid)	Geen weerstand tegen passieve bewegingen	Ontspannen	0
	Weerstand tegen passieve bewegingen	Gespannen, rigide	1
	Sterke weerstand tegen passieve bewegingen of niet in staat om ze te voltooien.	Zeer gespannen of rigide	2
Acceptatie van beademing (geïntubeerde patiënten)	Alarmen gaan niet af, ongehinderde beademing	Accepteert beademing of beweging	0
	Hoesten, alarmen gaan mogelijk af maar stoppen vanzelf	Hoesten, maar acceptatie	1
	Asynchronie: ademt asynchroon, alarmen gaan regelmatig af	Vecht tegen beademing	2

CHAPTER 5

Can the Critical-Care Pain Observation Tool (CPOT) be used to assess pain in delirious ICU patients?

Critically ill patients frequently experience both procedural pain and pain at rest. Chest tube removal, tracheal suctioning, wound care, turning and arterial line insertion have been shown to be the most painful procedures [1;2].

Untreated acute pain in adult ICU patients can lead to short- and long-term physiological and psychological complications such as postoperative myocardial infarction, insufficient sleep and posttraumatic stress disorder [3-6]. Practice guidelines recommend an individualized and goal directed pain management. This includes a systematic assessment of pain with a validated pain scale appropriate to the patient's level of consciousness. Pain assessment in critically ill patients is a challenge due to mechanical ventilation, severe illness, administration of sedatives and analgesics or a decreased level of consciousness. When a patient's self-report is unachievable, validated behavioral pain scales are advised for the assessment of pain in this particular group of patients [6;7].

Two independent systematic reviews compared the psychometric proportions of pain assessment scores for intensive care patients who are unable to self-report pain [8;9]. The Critical-Care Pain Observation Tool (CPOT) and Behavioral Pain Scale (BPS) received the best scores in their quality assessments and both scores are recommended in recent clinical practice guidelines for the assessment of pain in nonverbal critically ill adults [10-12]. The CPOT was developed for the assessment of pain in critically ill patients. The scale consists of four behavioral domains: facial expression, body movements, muscle tension and compliance with the ventilation for intubated patients or vocalization for extubated patients. Patient's behavior in each domain is scored between 0 and 2. The possible total score ranges from 0 (no pain) to 8 (maximum pain). The CPOT cutoff score was >2 during nociceptive procedures [7;13].

A limitation of the CPOT is the lack of sufficient research in delirious critically ill patients . Delirium is a common complication in ICU patients and the incidence of delirium after cardiac surgery varies between 3-55% [14]. The overall incidence in critically ill patients is on average 30-50% [15]. Self-report of pain in this vulnerable group of patients is complicated because of the limited communication, the variable level of consciousness and a potential different presentation of pain. As a consequence, validation of a behavioral pain scale like the CPOT in delirious critically ill patients is warranted [8]. *Kanji et. al.* addressed this problem and investigated the validity and reliability of the CPOT in adult critically ill patients with a delirium [16]. They included 40 ICU patients in which delirium was positively assessed with the confusion Assessment Method-ICU (CAM-ICU) and excluded patients who were unable to show a reliable physical response to pain. The authors thoroughly evaluated several important psychometric properties of the CPOT like the discriminant validation, the inter-rater

reliability, and the internal consistency. Discriminant validation is the assessment of the ability of a scale to discriminate between different conditions or groups. Pain scales are often tested by comparing the score between a painful and non-painful procedure. The inter-rater reliability is the degree of agreement between different raters on different occasions [8;17]. The authors chose a non-invasive blood pressure measurement as a non-painful procedure and repositioning, endotracheal suctioning of a dressing change as the painful procedures. The mean difference between baseline and painful procedures was 3.13 ± 1.56; p= 0.001. The inter-rater reliability was based on 120 paired assessments between one of two members of the study team and an independent nurse who was not familiar with the patient. The authors tested the inter-rater reliability by the calculation of weighted kappa coefficients, spearman correlation coefficients and intraclass correlation coefficients (ICC) for the individual domains and the overall CPOT score. All coefficients had substantial to almost perfect agreement for the individual domains and the overall CPOT score. Kanji et. al. concluded that their study indicates that the CPOT is a valid and reliable tool for the detection of pain in non-comatose, delirious adult ICU patients.

Although this study was meticulously designed and executed, a firm conclusion on the use of the CPOT in delirious patients cannot be made yet. In this study a point of concern is the lack of data about the severity of delirium, the subtype of delirium and the relation between the Richmond Agitation-Sedation Scale (RASS) and CPOT score. The DSM-V subdivides delirium in three subtypes: 1. Hyperactive form 2. Hypoactive form and 3. Mixed form. The hyperactive form is characterized by increased vigilance, restlessness, aggression and intense emotions, such as anger or anxiety. The hypoactive form is characterized by reduced alertness, sparse speech and apathy. In patients suffering from the mixed form, hyperactive and hypo - active periods alternate with each other. *Peterson et al.* defined the three subtypes according to the RASS scores [18]. A hyperactive delirium was present when the RASS was persistently positive (+1 to +4). Pain and agitation may interfere in delirious patients resulting in a higher CPOT due to agitation instead of pain. In addition, the interference of sedation needs further investigation [9]. *Kanji et al.* reported a median RASS of 0 with a range from -3 to +3 which shows that they included a number of patients with anxious or apprehensive movements (RASS +1), patients with frequent non-purposeful movements or patient-ventilator dyssynchrony (RASS +2) or patients pulled on tube(s) and had aggressive behavior toward staff (RASS +3). All four domains of the CPOT may potentially have been affected by high RASS scores, which might result in inappropriate high CPOT scores. These high CPOT scores may lead to additional use of analgesics were anti-delirium medication would be more appropriate. A recent study about the validity of the CPOT and BPS showed in a subgroup of seven agitated patients (RASS +1) non-significant increases in CPOT scores between rest and the painful procedure but no difference at all between the non-painful procedure and the painful procedure. The baseline CPOT score in this small subgroup was also higher than patients with RASS < +1 (19). Although this was a very small sample it is a signal that the validity of the CPOT in patients with a hyperactive delirium and/or RASS > +1 requires further investigation.

In contrast to previously performed research, *Kanji et. al* reported the inter-rater reliability of the four domains of the CPOT instead of the inter-rater reliability of the different procedures (painful versus non-painful or rest). A drawback of this method is that it does not comply with daily ICU practice since the CPOT is used as the sum of four domains during different occasions like tracheal suctioning or rest. The inter-rater reliability of the CPOT in delirious patients during different procedures is therefore still unknown.

In this study and several previous studies, either one of the investigators or the physicians participated in the assessments. However, in daily practice a large group of nurses assess pain in the intensive care. In addition, the bedside nurse potentially interprets the patient's reactions better because of a longer contact time. Hence, more raters should be used in the assessment of inter-rater reliability in future studies [10]. Finally, there are at least six versions of the ICC and they can give different results when applied in the same data [17;20]. The authors did not report which model of ICC was used in the analysis and thus it is unclear whether they used the appropriate ICC model.

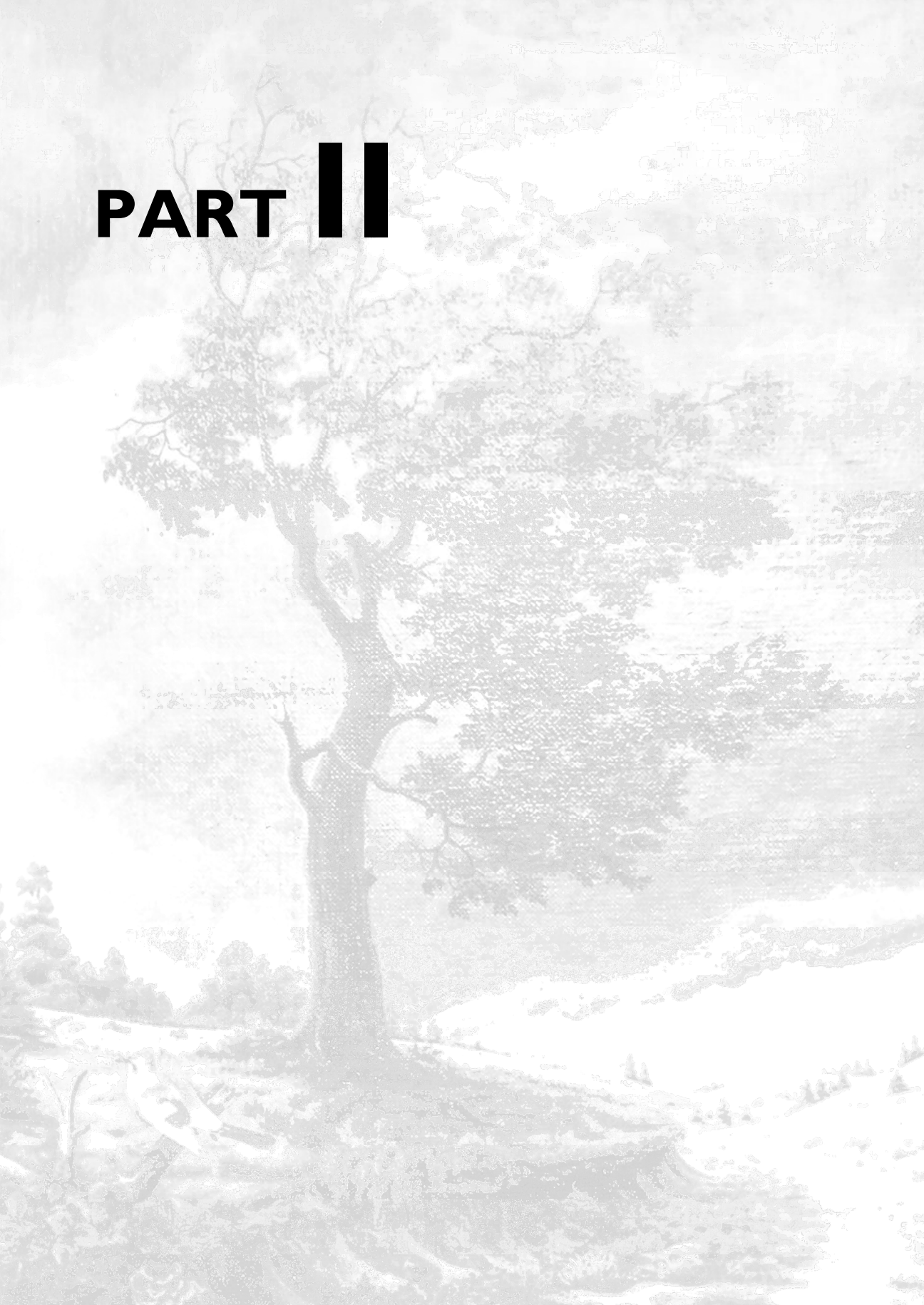
In conclusion, the study of *Kanji et al.* is an important first step in the validation of the CPOT in critically ill patients with a delirium. However, assessment of the inter-rater reliability of the CPOT should reflect daily practice in IC. Studies with a larger sample of delirious patients, and sufficient subsets of the three subtypes of delirium and RASS > +1, are obligatory before we can conclude that the CPOT is a valid and reliable pain assessment tool in ventilated critically ill patients suffering from a delirium.

REFERENCES

1. Puntillo KA, Max A, Timsit JF, Vignoud L, Chanques G, Robleda G, et al. Determinants of procedural pain intensity in the intensive care unit. The Europain(R) study. Am J Respir Crit Care Med 2014 Jan1;189 (1):39-47.
2. Bruce EA, Howard RF, Franck LS. Chest drain removal pain and its management: a literature review. J Clin Nurs 2006 Feb;15(2):145-54.
3. Chanques G, Jaber S, Barbotte E, Violet S, Sebbane M, Perrigault PF, et al. Impact of systematic evaluation of pain and agitation in an intensive care unit. Crit Care Med 2006 Jun;34(6):1691-9.
4. Granja C, Gomes E, Amaro A, Ribeiro O, Jones C, Carneiro A, et al. Understanding posttraumatic stress disorder-related symptoms after critical care: the early illness amnesia hypothesis. Crit Care Med 2008 Oct;36(10):2801-9.
5. Payen JF, Bosson JL, Chanques G, Mantz J, Labarere J. Pain assessment is associated with decreased duration of mechanical ventilation in the intensive care unit: a post Hoc analysis of the DOLOREA study. Anesthesiology 2009 Dec;111(6):1308-16.
6. Baron R, Binder A, Biniek R, Braune S, Buerkle H, Dall P, et al. Evidence and consensus based guideline for the management of delirium, analgesia, and sedation in intensive care medicine. Revision 2015 (DAS-Guideline 2015) - short version.Ger Med Sci 2015;13:Doc19.
7. Barr J, Fraser GL, Puntillo K, Ely EW, Gelinas C, Dasta JF, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306.
8. Gelinas C, Puntillo KA, Joffe AM, Barr J. A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Seminars in Respiratory and Critical Care Medicine 2013 Apr;34(2):153-68.

9. Pudas-Tahka SM, Axelin A, Aantaa R, Lund V, Salantera S. Pain assessment tools for unconscious or sedated intensive care patients: a systematic review. J Adv Nurs 2009 May;65(5):946-56.
10. Gelinas C, Fillion L, Puntillo KA, Viens C, Fortier M. Validation of the critical-care pain observation tool in adult patients. Am J Crit Care 2006 Jul;15(4):420-7.
11. Barr J, Fraser GL, Puntillo K, Ely EW, Gelinas C, Dasta JF, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306.
12. Payen JF, Bru O, Bosson JL, Lagrasta A, Novel E, Deschaux I, et al. Assessing pain in critically ill sedated patients by using a behavioral pain scale. Crit Care Med 2001 Dec;29(12):2258-63.
13. Stilma W, Rijkenberg S, Feijen HM, Maaskant JM, Endeman H. Validation of the Dutch version of the critical-care pain observation tool. Nurs Crit Care 2015 Dec 22.
14. Ouimet S, Kavanagh BP, Gottfried SB, Skrobik Y. Incidence, risk factors and consequences of ICU delirium. Intensive Care Med 2007 Jan;33(1):66-73.
15. van den Boogaard M, Schoonhoven L, van der Hoeven JG, van AT, Pickkers P. Incidence and short-term consequences of delirium in critically ill patients: A prospective observational cohort study. Int J Nurs Stud 2012 Jul;49(7):775-83.
16. Kanji S, MacPhee H, Singh A, Johanson C, Fairbairn J, Lloyd T, et al. Validation of the Critical Care Pain Observation Tool in Critically Ill Patients With Delirium: A Prospective Cohort Study. Crit Care Med 2016 Jan 16.
17. Streiner D.L, Norman G.R. Health measurement scales a practical guide to their development and use. third edition ed. Hamilton: Oxford university press; 2003.
18. Peterson JF, Pun BT, Dittus RS, Thomason JW, Jackson JC, Shintani AK, et al. Delirium and its motoric subtypes: a study of 614 critically ill patients. J Am Geriatr Soc 2006 Mar;54(3):479-84.
19. Rijkenberg S, Stilma W, Endeman H, Bosman RJ, Oudemans-van Straaten HM. Pain measurement in mechanically ventilated critically ill patients: Behavioral Pain Scale versus Critical-Care Pain Observation Tool. J Crit Care 2015 Feb;30(1):167-72.
20. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-8.

PART II



Continuous QTc measurement in critically ill patients

CHAPTER 6

Corrected QT-interval prolongation and variability in intensive care patients

E. Hoogstraaten, **S. Rijkenberg**, P.H.J. van der Voort

Hoogstraaten et al. Journal of Critical Care, 2014, 29(5)

CHAPTER 7

Validation of continuous QTc measurement in critically ill patients

G. Janssen, **S. Rijkenberg**, P.H.J. van der Voort

This chapter in adapted form has been published in:
Janssen et al. Journal of Electrocardiology, 2016, 49(1); 81-86

There has been no financial support for these studies

CHAPTER 6
Corrected QT-interval prolongation and variability in intensive care patients

ABSTRACT

Introduction

Critically ill patients are at risk for prolongation of the interval between the Q wave and the T wave in the electrocardiogram (corrected QT [QTc]). Corrected QT prolongation is probably a dynamic process. It is unknown how many patients have a QTc prolongation during their intensive care stay and how variable QTc prolongation is.

Materials and methods

In a prospective cohort study, continuous 5-minute QTc measurements of 50 consecutive patients were collected. A prolonged QTc interval was more than 500 milliseconds for at least 15 minutes. The QT variance and variability index was used to evaluate QTc variation.

Results

Fifty-two percent of included patients had a prolonged QTc interval. In a single patient, 0.2% to 91.3% of the QTc intervals over time were prolonged. The use of erythromycin and amiodarone was associated with the mean QTc (p= 0.02 and p= 0.006, respectively). The Acute Physiology and Chronic Health Evaluation IV and Sequential Organ Failure Assessment scores were significantly higher in patients with a prolonged QTc interval (30.8 vs. 8.6 and 7 vs. 5.5, respectively). Eighty-four percent of all patients received at least 1 QTc prolonging drug. The QT variance and QTc variance were significantly higher in patients with a prolonged QTc (p= 0.019 and p= 0.001, respectively).

Conclusions

Continuous QTc monitoring showed a prolonged QTc interval in 52% of intensive care patients. Severity of illness and QT and QTc variances are higher in these patients.

INTRODUCTION

It is well recognized that a prolonged ventricular repolarization, reflected on the electrocardiogram (ECG) as a prolonged corrected QT (QTc) interval, is associated with ventricular arrhythmias [1–3]. This is also true in the hospital setting and, in particular, the intensive care unit (ICU) [4–7]. There is no definite consensus about the normal limits of the QTc interval; however, a QTc more than 500 milliseconds is associated with a higher occurrence of torsades de pointes (TdP) [4]. Some authors use lower limits, for example, more than 450 milliseconds for men and more than 460 milliseconds for women [8]. Prolongation of the QTc interval can be both congenital and acquired [9,10]. The acquired form is reversible and has a still growing list of causes. Risk factors for an acquired prolonged QTc in the ICU population seem to be similar to those in the ambulatory population, including older age, female sex, low body mass index (BMI; anorexia nervosa), bradycardia, heart disease (especially ischemia and left ventricular hypertrophy), electrolyte abnormalities (hypokalemia, hypomagnesemia, hypocalcemia), liver and kidney impairment, subarachnoid hemorrhage, and the use of QTc-prolonging medication [4,11,12]. This QTc-prolonging medication consists of both cardiac (antiarrhythmic) and noncardiac drugs [4,12–15], among them being the, in critically ill patients, widely used amiodarone, sotalol, erythromycin, and haloperidol. The presence of QTc prolongation in intensive care patients is suspected to be high because this population often has several risk factors at the same time and receives a lot of medication [6,7]. The monitoring of the QTc interval in critically ill patients is usually performed by intermittent ECG, but it is unknown how often in between the ECGs a prolonged QTc is present. Recently, continuous QTc measurement has become available. This study was undertaken to determine the frequency of QTc interval prolongation. We hypothesized that the QTc interval, by using the continuous measurement, would show a considerable variability.

MATERIALS AND METHODS

Study population

The study was designed as a prospective observational single center cohort study and was conducted at the 22-bed ICU of the Onze Lieve Vrouwe Gasthuis in Amsterdam. Patients of 18 years and older who were admitted to our ICU between May 15, 2013, and June 25, 2013, and with a length of stay more than 24 hours were subsequently included. Eligible patients who were already staying in the ICU when the study started and who met the inclusion criteria were also included. All patients where followed up until their discharge from the ICU or until the arbitrarily chosen maximum follow-up time of 14 days. Patients in whom an accurate continuous QTc measurement was impossible to perform were excluded. This was the case when a prolonged QRS complex (>120 milliseconds), the use of a ventricular pacemaker, a ventricular bigeminy or trigeminy rhythm, and insufficient continuous QTc data (<500 measurements per 24 hours) were present. The hospital's medical-ethical committee approved the protocol for this study and waived informed consent because the study was observational and concerned only monitoring of standard variables and therapy.

Data collection

Continuous QTc interval measurements

All patients were monitored by a 5-electrode ECG, enabling to obtain 7 leads (I, II, III, aVR, aVL, aVF, V1). The novel continuous QT/QTc measurement software was developed by Philips Healthcare (Philips, Amsterdam, the Netherlands) and approved for clinical use in 2008 [6]. This software determined the QT interval per minute by constructing a root-mean-squared waveform, meaning a combined beat from all 7 leads. Every five minutes, a mean of these QT measurements was calculated and displayed. The Bazett formula was used to calculate QTc: $QTc = QT/\sqrt{RR}$ [4]. RR is the interval between two R waves in the ECG. Fridericia [16] described an alternative formula to correct QT for heart rate. The difference with the Bazett formula is that the Fridericia formula uses the cube root instead of the square root of the RR interval. Because the cube root would lead to longer QTc intervals compared with the square root and, therefore, with a higher incidence of QTc prolongation, we conservatively choose to use the Bazett formula. All the QTc interval data were stored in a clinical information system (MetaVision; iMDsoft, Tel Aviv, Israel). A prolonged QTc interval was defined as more than 500 milliseconds during 15 or more consecutive minutes, equal for men and women.

Risk factors for QTc interval prolongation

Data concerning the known risk factors for QTc prolongation were collected from the clinical information system and included sex, age, BMI, electrolyte disturbances (daily lowest values of potassium, magnesium, and calcium), bradycardia, and use of QTc-prolonging medication. This medication consisted of haloperidol, sotalol, amiodarone, ketanserin, ciprofloxacin, and erytromycin. In addition, liver and kidney functions (daily highest values) were determined. The Charlson comorbidity index, Acute Physiology and Chronic Health Evaluation (APACHE) IV, and daily Sequential Organ Failure Assessment (SOFA) scores were obtained to determine severity of disease.

QT variability

The variability of the QT interval over time was determined by using the QT variance and the QT variability index (QTVI), calculated as $QTVI = \log_{10}\{[QTv/(QTm^2)]/[RRv/(RRm^2)]\}$. This formula represents the log-ratio between the QT and RR variances (QTv and RRv), each normalized by the squared mean of the QT and RR (QTm and RRm) of a chosen time series [17]. A higher (closer to zero) calculated index indicates a higher QT variability and therefore a higher risk of ventricular arrhythmia.

Statistical analysis

Data were analyzed with SPSS version 18.0 (SPSS, Inc, Chicago, Ill). Descriptive statistics were used to provide means, medians, ranges, SDs (±), interquartile ranges (IQRs), and variances for continuous data. Statistical differences between patients with a prolonged QTc interval and patients without prolonged QTc interval were calculated using the Chi-square test for independent categorical variables and the independent-samples t test for continuous variables.

When data showed a skewed distribution, the nonparametric Mann-Whitney test was used. Results were statistically significant at $p \leq 0.05$.

RESULTS

During the study period, 163 patients were admitted to the ICU. Seventy-six patients were admitted for less than 24 hours; 1 patient was younger than 18 years; and 13 patients had insufficient QTc monitoring data. Of the remaining 73 patients, 14 were excluded because of a QRS duration more than 120 milliseconds, 2 were excluded because of bigeminy/trigemy rhythm, and 7 were excluded because of a ventricular pacemaker. The final sample of 50 analyzed patients consisted of 35 men (70%) and 15 women (30%) with a mean age of 63 (±14) years. The median length of stay was 3.5 days, with an IQR of 2 to16 days and a range of 1 to 69 days.

Occurrence of QTc prolongation

None of the patients had a known form of congenital prolonged QTc syndrome. A total of 221405 minutes of continuous QT-interval monitoring data were obtained, with a median of 2892 minutes per patient, ranging from 709 to 19614 minutes. The mean Bazett QTc interval was 457 (±33) milliseconds, ranging from 301 to 669 milliseconds. The mean QTc over their follow-up was significantly longer in patients who were treated with erythromycin (p= 0.02) and with amiodarone (p= 0.006) compared with those without this medication. Corrected QT-interval prolongation, defined as more than 500 milliseconds during an interval of least 15 minutes, occurred in 26 patients (52%).

Prolonged QTc vs non-prolonged QTc

Differences in baseline characteristics for the prolonged QTc group vs. the non-prolonged QTc group are shown in Table 1. Univariate analyses revealed no significant differences between the 2 groups concerning sex, age, BMI, length of stay, type of admission, mechanical ventilation time, and Charlson comorbidity index scores. The APACHE IV predicted mortality score, and the SOFA score was significantly higher in the patients with a prolonged QTc interval. Electrolytes and liver and kidney function were not significantly different between groups. Cardiac arrhythmia was more common in the prolonged QTc group but did not reach a level of significance. The incidence of TdP was 1 (3.8%) of 26 in patients with prolonged QTc interval.

In patients with a prolonged QTc, the proportion of QTc values that were prolonged in a single patient varied from 0.2% to 91.3%, with a median of 13% (IQR, 2% - 45%). The time of onset of prolongation of the QTc interval ranged from 15 to 2030 minutes after admission. Four patients had a prolonged QTc at the time of admission, which were all normalized at the end of follow-up, indicating that it was not a (so far undiscovered) congenital type of prolonged QTc.

Among the 26 patients with a prolonged QTc interval, 24 (92%) received a QTc-prolonging drug compared with 18 (75%) in the non-prolonged QTc group, as shown in Table 2 (*p*-value, not significant). According to patients' charts, in two cases, haloperidol was discontinued because a prolonged QTc interval was noticed. For no other patients, adjustments in medication were made because of prolonged QTc interval.

QTc variation: variance and variability index

The mean monitoring time (number of values) in the prolonged QTc and non-prolonged QTc groups was not different (respectively 2884 IQR [1359 – 7176] and 2892 IQR [1610 – 4854], *p*= 0.88). The variance of both the QT and the QTc over the monitoring time was significantly prolonged in the patients with a prolonged QTc (Figure 1). The RR-interval (heart rate) variances did not differ between the 2 groups with 3774 IQR [1482 – 8729] in the prolonged QTc patients vs. 3368 IQR [1791 – 6214] in the normal QTc patients (*p*= 0.85). The QTVI showed a normal distribution for the patients without QTc prolongation but a skewed distribution for the group of patients with QTc prolongation (Table 3). Because the QTVI was calculated by the log ratio between the QT- and RR-interval variances, this larger QTVI in the prolonged QTc group was due to the larger QT variance (the numerator in the equation), rather than a decreased RR variance (the denominator in the equation).

Table 1. Baseline characteristics of included patients

Variables	All n= 50	Prolonged QTc n= 26 (52%)	Non-prolonged QTc n= 24 (48%)	*p*-value
Sex (male)	35 (70)	20 (77)	15 (63)	0.266
Age (years)	63 ± 14.	63.7 ± 14.1	62.1 ± 14.4	0.684
BMI[1] (kg/m^2)	26.1 [24 - 29.4]	26.1 [23.3 - 31.2]	26.1 [24.3 - 28.3]	0.528
APACHE[2] IV pm (%)	17.4 [3.5 - 36]	30.8 [5.2 - 44.4]	8.6 [3.5 - 24.9]	0.045
SOFA[3] day one	6 [5 - 7]	6.5 [5.0 - 8.3]	5 [4.3 - 6]	0.048
SOFA highest score	6 [5 - 7]	7 [5 - 9]	5.5 [5 - 7]	0.032
Charlson comorbidity index	1 [0 - 2]	1 [0 - 3]	1 [0 - 1.8]	0.854
LOS[4] (days)	3.5 [2 -15.8]	4.3 [2 - 18.9]	3 [2 - 10.5]	0.553
Mechanical ventilation (hours)	43 [19.5 - 143]	57 [29 - 215]	37 [15 - 114]	0.383
Admission type				
Medical	28 (56)	15 (57.7)	13 (54.2)	
Surgical	22 (44)	11 (42.3)	11 (45.8)	0.802
Reason for admission				
Cardiac surgery	20 (40)	9 (34.6)	11 (45.8)	
General surgery	3 (6)	2 (7.7)	1 (4.2)	

Internal medicine	5 (10)	3 (11.5)	2 (8.3)	
Neurology	2 (4)	1 (3.8)	1 (4.2)	
Infectious disease	8 (16)	5 (19.2)	3 (12.5)	
Respiratory insufficiency	8 (16)	3 (11.5)	5 (20.8)	
Cardiopulmonary resuscitation	4 (8)	3 (11.5)	1 (4.2)	
Potassium lowest value	3.8 ± 0.4	3.7 ± 0.3	3.8 ± 0.5	0.260
Calcium lowest value	2.1 ± 0.2	2.0 ± 0.2	2.1 ± 0.3	0.346
Magnesium lowest value	0.84 ± 0.20	0.84 ± 0.22	0.84 ± 0.18	0.951
Creatinine highest value	95 [74 - 137]	98 [76 - 147]	94 [70 - 133]	0.303
ALAT[5] highest value	43 [29 - 73]	43 [30 - 70]	45 [25 - 89]	0.946
Bradycardia	5 (10)	3 (12)	2 (8)	1.000
VT[6]	6 (12)	4 (15)	2 (8)	0.669
PVC[7] (>2/h)	11 (22)	7 (27)	4 (17)	0.382
TdP[8]	1 (2)	1 (4)	0	1.000

Plus-minus values are means ± SDs; other values are medians and IQR Data are mean (SD), median (IQR) or n (%)

[1] BMI: Body mass index
[2] APACHE: Acute Physiology and Chronic Health Evaluation, pm indicates predicted mortality
[3] SOFA: sepsis-related organ failure assessment score
[4] LOS: length of stay
[5] ALAT: alanine transferase
[6] VT: ventricular tachycardia
[7] PVC: premature ventricular complex
[8] TdP: torsades de pointes

Table 2. Use of QTc-prolonging medication

Medication	Prolonged QT (n= 26)	Non-prolonged QTc (n=24)	*p*-value
Haloperidol	11 (42)	10 (42)	0.963
Daily dosage (mg)	4.2 [2.3 - 6.0]	3.6 [2.4 - 9.5]	0.520
Sotalol	6 (23)	3 (13)	0.467
Daily dosage (mg)	63 ± 20	80 ± 0	0.199
Erytromycin	7 (27)	2 (8)	0.142
Daily dosage (mg)	321 ± 122	313 ± 88	0.927
Ciprofloxacin	8 (31)	10 (42)	0.423
Daily dosage (mg)	800 [600 - 800]	800 [800 - 800]	0.292
Amiodaron	7 (27)	2 (8)	0.142
Daily dosage (mg)	600 [450 - 725]		0.703
Ketanserin	8 (31)	4 (17)	0.243
Daily dosage (mg)	41.1± 14.1	36.3± 9.2	0.610
No medication	2 (8)	6 (25)	0.132
≥ 2 types medication	13 (50)	7 (29)	0.133

Data present absolute number of patients with a specific medication and (%).
Plus-minus values are means ± SDs; other values are medians and IQRs.

Table 3. Descriptive characteristics of QT variance, QTc variance, and QTVI in patients without QTc prolongation

		Min	Max	Mean	Median	IQR	*p*-value
QT variance	QTc <500 msec	90	1457	451	370	440	0.019
	QTc >500 msec	135	2241	810	570	897	
QTc variance	QTc <500 msec	58	651	199	158	127	0.001
	QTc >500 msec	113	1740	501	318	571	
QTVI	QTc <500 msec	-0.73	-0.07	-0.43	-0.41	0.23	0.07
	QTc >500 msec	-0.57	0.21	-0.25	-0.37	0.56	

Min: minimum
Max: maximum
IQR: interquartile range

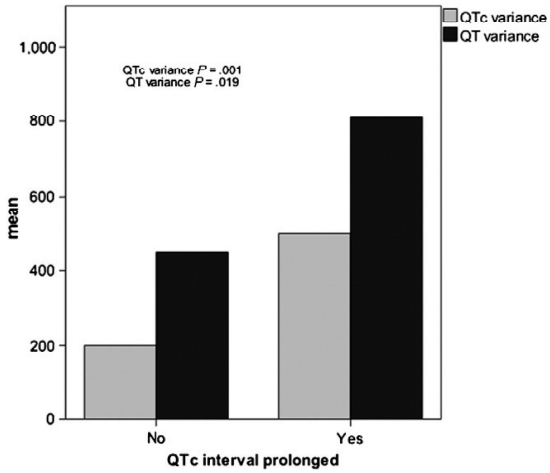


Figure 1. QT and QTc variances in the prolonged QTc group vs. non-prolonged QTc group

DISCUSSION

This prospective study showed, using a continuous measurement technique, a high incidence of QTc prolongation (52%) in a mixed medical surgical population of critically ill patients with an intensive care stay of more than 24 hours. We used the Bazett formula instead of the Fridericia formula. The Bazett formula will lead to less prolonged QTc measurements compared with the Fridericia formula. As such, we may even have underestimated the incidence of prolonged QTc. A large inter patient variance of duration of this QTc prolongation was present, ranging from 0.2% to 91.3% of the measurements. Furthermore, the QT and QTc variances were significantly higher in patients with a prolonged QTc interval, indicating more electrical instability of the ventricular conductance.

Measurement of QTc interval

Since the introduction of ECG monitoring in hospitalized patients, many improvements have been made according to the functions and accuracy of the ECG monitors [4]. Recently, a novel adjustment to the ECG monitor made it possible to continuously measure the QT interval. A study conducted on the accuracy of this algorithm proved it to be stable and accurate [18]. The continuous QT measurement is a reflection of global ventricular repolarization, which seems to be a better indicator than a single-lead (manually) measurement, partly because the leads that are used do not change during measurement period [6,18]. This also explains why this automated QTc intervals can differ somewhat from the QTc measured by the ECG using only 1 lead. Limitations of the QTc monitoring include high heart rate, as a heart rate more than 150 beats/min leads to P and T waves approaching each other too much. A wide QRS complex also confounds the QT measurements. Therefore, all patients with QRS more than 120 milliseconds were excluded in our study. Tachycardia greater than 150/min did not occur.

This study is the second study using continuous QTc measurements to evaluate the variation and prolongation of the QTc interval in critically ill patients. The other study on QTc prolongation using a continuous QTc monitor system is the QT-in-practice-study [6,19]. Their prospective observational study concluded that QTc prolongation in the ICU was common (24%) and increased the mortality risk nearly 3 times. Our study suggests that QTc prolongation in the ICU is even more common (52%). The present study added severity of disease to the risk factors of prolonged QTc, as we found the APACHE IV predicted mortality score to be significantly higher in the prolonged QTc group. Our follow-up time and sample size were insufficient to address the risk of mortality, but the high APACHE IV predicted mortality scores in the prolonged QTc group point indicates a higher mortality risk in this group. In addition, the present study measured QTc variability and QT variance, which appeared to be associated with a prolonged QTc interval.

Definition of prolonged QTc interval

A prolonged QTc interval in our study was defined as QTc more than 500 milliseconds, for at least 15 consecutive minutes. There is no consensus in the literature about what QTc interval threshold should be considered clinically important [7,14,20,21]. Several authors used the 500-millisecond cutoff [5,6,14]. Pickham et al [6] proposed the 15 consecutive minutes. Because little research so far using continuous QTc measurements has been performed, there is no evidence yet for the time that a prolonged QTc has to be present to create a higher risk of TdP. Our study presents a very high range in the time that QTc is prolonged, ranging from 0.2% to 91.3%. It needs further study to determine whether a prolonged QTc of 15 consecutive minutes is relevant or that a somewhat longer period is more associated with clinical outcome. In addition, it is still unclear what actions nurses and physicians should undertake other than be aware and reduce QT prolonging drugs. The use of erythromycin and amiodarone was associated with QTc interval but did not significantly add to the risk of a QTc prolongation more than 500 milliseconds. This may be due to the relatively low dosage of erythromycin (250 mg twice daily) or to the limited sample size.

QTc variation and the QTVI

Berger et al. [17] designed an index to describe the relative magnitude of QT-interval changes compared with heart rate variability. They validated this QTVI for beat-to-beat QT changes, measured for 256 consecutive seconds in patients with cardiomyopathy and heart failure. They found the QTVI to be elevated in these patient categories and posed QTVI to be a marker of electrical cardiac disease, which might be associated with higher risk of ventricular arrhythmias [17,22]. In 2004, a Finnish study was conducted in which the investigators evaluated whether the QTVI was increased in patients with a congenital type of prolonged QTc in comparison with healthy controls [23]. They indeed found the QTVI in 64 patients with congenital prolonged QTc compared with 32 controls to be significantly higher. Moreover, in accordance to our findings, the increased QTVI in this group was caused by a significant higher QT

variance (the numerator in the formula). The heart rate variance (the denominator in the formula) was not decreased in the QTc-prolonged patients. However, the opposite appears to exist; also, a recent study on the QTVI in patients with left ventricular dysfunction concluded that increased QTVI predicts cardiovascular mortality [24]. They found a decreased heart rate variance rather than an increased QT variance to be the cause of this QTVI elevation. In our study of intensive care patients, a significantly greater QT and QTc variance was present in the patients with prolonged QTc but not a significantly increased RR variance. The log ratio of these two variances, the QTVI, was however, not significantly different (p= 0.07; Table 3) because of the small sample size. QT variance is apparently a predictor of a prolonged QTc.

Indication for continuous QTc monitoring in ICU patients

In the present study, severity of illness was the only significant risk factor for QTc prolongation. Severity of disease may thus be an independent risk factor for QTc prolongation. With this novel continuous QTc measurement tool now available, further study should explore whether it is appropriate to monitor intensive care patients with continuous monitoring of the QTc interval, in particular when the severity of illness is high. Recent statement from the American Heart Association recommends hospitalized that patients should have continuous QTc monitoring when they receive QTc prolonging drugs, when electrolyte disturbances (potassium or magnesium) are present, when bradycardia is present, or when the reason for admission was drug overdose [7]. A previous study showed that 69% of the intensive care patients had one or more American Heart Association indications for continuous QTc monitoring [6,19]. It was therefore stated that the need of continuous QTc monitoring in intensive care patients is high [19]. In our study, considering use of QTc-prolonging medication alone, up to 84% of patients meet the indication criteria for continuous QTc monitoring. In addition, the proportion of the time that the QTc interval is prolonged varies a lot, but overall QTc prolongation is highly prevalent.

Limitations

The most important limitation of this study was the relative small sample size of 50 patients. Especially when considering types of medication, even smaller numbers of patients are compared. This probably contributed to a lack of power to observe significant differences in risk factors between the prolonged and non-prolonged QTc groups. A larger study needs to be performed to address this issue. Another issue to take in consideration is that 15 of the included patients were already present in the ICU at the start of the study and, as a consequence, were not followed up from the beginning of their admission. However, 8 (53%) of them had a prolonged QTc, which is concordant with the 52% of the entire group. The small sample limits conclusions about the relation with ventricular arrhythmias. However, the main goal of the study was to investigate whether or not patients had a prolonged QTc and not the consequences.

Another limitation is that automated measurements of the QTc measurements were used in which errors could have occurred. We did not manually over read these. In addition, the positions of the electrodes will probably have been slightly changed as a consequence of daily routine change of electrodes. This could have made minimal changes to the QT measurements. Also, we excluded patients with confounders of QTc measurements such as a widened QRS complex. This might have caused selection bias. Our conclusions can therefore only be extrapolated for patients without a widened QRS complex. Finally, when addressing the QTVI, we used this QTVI not in a beat-to-beat variance (in which it has been validated) but in a five-minute QT variance. In addition, this QTVI has not yet formally been validated for critically ill patients.

CONCLUSIONS

We have shown that prolongation of the QTc interval has a high cumulative incidence (52%) in critically ill patients when monitored by continuous QTc measurements. However, the onset of QTc prolongation, as well as the total duration of the prolonged QTc, varies. The QT and QTc variances were significantly higher in patients with prolongation of the QTc interval, indicating a greater instability of ventricular repolarization. Severity of illness is associated with a prolonged QTc, and therefore, continuous QTc measurements should be used to detect all episodes of QTc prolongation in a critically ill patient.

REFERENCES

1. El Sherif N, Turitto G. Torsade de pointes. Curr Opin Cardiol 2003;18:6–13.
2. Fabiato A, Coumel P. Torsade de pointes, a quater of a century later: a tribute to Dr. F. Desertenne. Cardiovasc Drugs Ther 1991;5:167–9.
3. Yap YG, Camm AJ. Drug induced QT prolongation and torsades de pointes. Heart 2003;89:1363–72
4. Drew BJ, Califf RM, Funk M, Kaufman ES, Krucoff MW, Laks MM, et al. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. Circulation 2004;110:2721–46.
5. Kozik TM, Wung SF. Acquired long QT syndrome: frequency, onset and risk factors in intensive care patients. Crit Care Nurse 2012;32:32–41.
6. Pickham D, Helfenbein E, Shinn JA, Chan G, Funk M, Weinacker A, et al. High prevalence of corrected QT interval prolongation in acutely ill patients is associated with mortality: results of the QT in Practice (QTIP) study. Crit Care Med 2012;40:394–9.
7. Drew BJ, Ackerman MJ, Funk M, Gibler WB, Kligfield P, Menon V, et al. Prevention of Torsade de pointes in hospital settings: a scientific statement from the American Heart Association and the American College of Cardiology Foundation. J Am Coll Cardiol 2010;55:934–47.
8. Williams ES, Thomas KL, Broderick S, Shaw LK, Velazquez EJ, Al-Khatib SM, et al. Race and gender variation in the QT interval and its association with mortality in patients with coronary artery disease: results from the Duke Databank for Cardiovascular Disease (DDCD). Am Heart J 2012;164:434–41.
9. Camm AJ, Janse MJ, Roden DM, Rosen MR, Cinca J, Cobbe SM. Congenital and acquired long QT syndrome. Eur Heart J 2000;21:1232–7.
10. Moss AJ. Long QT syndrome. JAMA 2003;289:2041–4.

11. Benoit SR, Mendelsohn AB, Nourjah P, Staffa JA, Graham DJ. Risk factors for prolonged QTc among US adults: Third National Health and Nutrition Examination Survey. Eur J Cardiovasc Prev Rehabil 2005;12:363–8.
12. Tan HL, Hou CJ, Lauer MR, Sung RJ. Electrophysiologic mechanisms of the long QT interval syndromes and torsade de pointes. Ann Intern Med 1995;122:701–14.
13. Zeltser D, Justo D, Halkin A, Prokhorov V, Heller K, Viskin S. Torsade de pointes due to noncardiac drugs: most patients have easily identifiable risk factors. Medicine (Baltimore) 2003;82:282–90.
14. Roden DM. Drug-induced prolongation of the QT interval. N Engl J Med 2004;350:1013–22.
15. www.qtdrugs.org. [Assessed 07 Jan 2014].
16. Fridericia LS. The duration of systole in the electrocardiogram of normal subjects and of patients with heart disease. 1920. Ann Noninvasive Electrocardiol 2003;8:343–51.
17. Berger RD, Kasper EK, Baughman KL, Marban E, Calkins H, Tomaselli GF. Beat-tobeat QT interval variability: novel evidence for repolarization lability in ischemic and nonischemic dilated cardiomyopathy. Circulation 1997;96:1557–65.
18. Helfenbein ED, Zhou SH, Lindauer JM, Field DQ, Gregg RE, Wang JJ, et al. An algorithm for continuous real-time QT interval monitoring. J Electrocardiol 2006;39:S123–7.
19. Pickham D, Helfenbein E, Shinn JA, Chan G, Funk M, Drew BJ. How many patients need QT interval monitoring in critical care units? Preliminary report of the QT in practice Study. J Electrocardiol 2010;43:572–6.
20. Molnar J, Zhang F, Weiss J, Ehlert FA, Rosenthal JE. Diurnal pattern of Qtc interval: how long is prolonged? Possible relation to circadian triggers of cardiovascular events. J Am Coll Cardiol 1996;27:76–83.
21. Al Khatib SM, LaPointe NM, Kramer JM, Califf RM. What clinicians should know about the QT interval. JAMA 2003;289:2120–7.
22. Berger RD. QT variability. J Electrocardiol 2003(36 Suppl.):83–7.
23. Bilchick K, Viitasalo M, Oikarinen L, Fetics B, Tomaselli G, Swan H, et al. Temporal repolarization lability differences among genotyped patients with the long QT syndrome. Am J Cardiol 2004;94:1312–6.
24. Tereshchenko LG, Cygankiewicz I, McNitt S, Vazquez R, Bayes-Genis A, Han L, et al. Predictive values of beat-to-beat QT variability index across the continuum of left ventricular dysfunction: competing risks of noncardiac or cardiovascular death and sudden or nonsudden cardiac death. Circ Arrhythm Electrophysiol 2012;5:719–27.

CHAPTER 7
Validation of continuous QTc measurement in critically ill patients

ABSTRACT

Introduction

Prolongation of the corrected QT interval (QTc) can lead to torsades de pointes. This study is designed to determine the validity of the continuous QTc measurement in critically ill patients.

Materials and methods

In a retrospective cohort study, QTc analysis was performed with manual measurements on a single selected lead from a 12-lead ECG and continuous QTc measurement obtained at the same time. In addition, automated QTc measurement from the 12-lead ECG were also included in the study. Validation was performed by calculating Intraclass correlation coefficient (ICC), Pearson's correlation and Bland-Altman plot.

Results

119 patients with QRS <120 msec were included with a mean continuous QTc of 468 msec (standard deviation (SD) 37) and mean manually measured QTc of 449 msec (SD 41) (p= 0.001). Pearson's correlation was 0.65 (p= 0.01), ICC was 0.65 (95% CI: 0.53-0.74). Bland-Altman plot shows a mean difference of 19.5 msec (limits of agreement (LOA) -44.6 - 83.7). For continuous QTc compared to automated QTc from the 12-lead ECG the Intraclass correlation coefficient was 0.77 (95% CI: 0.68 - 0.83, p= 0.001) and the Bland-Altman plot shows a mean difference of 7.8 msec (LOA -40.2 - 55.8)

Conclusions

Continuous QTc measurement in critically ill patients with a QRS duration shorter than 120 msec shows an acceptable accuracy to be used in routine care.

INTRODUCTION

Prolongation of the corrected QT interval (QTc) can lead to torsades de pointes (TdP), a malignant ventricular arrhythmia. If TdP sustains, it can lead to ventricular fibrillation and sudden cardiac death [1,2,3].

Prolonged QTc can be hereditary or acquired. The most important causes for acquired prolonged QTc in the ambulatory population include older age, female sex, low body mass index (anorexia nervosa), bradycardia, heart disease (especially ischemia and left ventricular hypertrophy), electrolyte abnormalities (hypokalaemia, hypomagnesaemia, hypocalcaemia), liver- and kidney impairment, subarachnoid haemorrhage and use of QT prolonging medication [4,5,6]. Medication that prolongs QT include the widely used amiodarone, sotalol, erythromycin and haloperidol. In critically ill patients cumulative incidences of prolonged QTc of 24 - 52% are described [7,10]. This high incidence of QTc prolongation in critically ill patients is probably due to the appearance of several risk factors simultaneously and the administration of medication intravenously [7,8,9]. In addition, severity of disease by itself appears to be a risk factor for QTc prolongation [10].

We previously described that a significant QTc variability exists during the course of stay at the ICU. This is even more pronounced for patients with prolonged QTc (longer than 500 milliseconds (msec)) [10]. A diurnal variation of QTc is also described [13].

Because of the high cumulative incidence of QTc prolongation, the variability and possible detrimental consequence of prolonged QTc, a continuous measurement of QTc may be useful in critically ill patients. It may help the clinician to decide on withholding QT prolonging medication or addressing electrolyte disturbances in an early stage.

Since 2008, new validated software for patient monitors is available to measure the QT and QTc intervals continuously [11]. The algorithm computes an average ECG from 3,5,6 or 10 electrode ECG leads. The monitor then determines the QT interval every five minutes. The corrected QT interval is then calculated by the default formula of Bazett, but it is also possible to choose the Fridericia formula [11,12].

However, to our knowledge this continuous QTc measurement has not been validated yet in critically ill patients.

This study is designed to validate the continuous QTc measurement on a patient monitor against 1) manually measured QT and RR interval and calculated QTc on standard 12-lead ECG in critically ill patients and 2) automatically measured QTc from the 12-lead ECG.

MATERIALS AND METHODS

The study was conducted in the Intensive Care Unit of the Onze Lieve Vrouwe Gasthuis (OLVG) hospital, a 22 bedded, mixed surgical/medical ICU in a teaching hospital.

The study is a single-center retrospective cohort analysis of all consecutive patients admitted in June and July 2013. Patients aged eighteen years or over were included when on the first morning of admission a 12-lead ECG was performed and at the same time a valid continuous QTc measurement on the patient monitor (Philips Healthcare IntelliVue MP70) was obtained.

The study was approved by the medical ethical committee and a waiver for informed consent was given according to Dutch and European legislation because of the retrospective and observational design of the study.
Primary endpoint is accuracy of continuous QTc measurement with the patient monitor compared to manually measured QT and RR interval and calculated QTc on a single lead selected from a 12-lead ECG with a QRS duration shorter than 120 msec.
Secondary endpoint is accuracy of continuous QTc measurement with the patient monitor compared to automatically derived QTc on a 12-lead ECG with a QRS duration shorter than 120 msec.
In addition, the same accuracy comparisons are also performed with a QRS duration longer than 120 msec.

Data collection

We extracted demographic and clinical features from the patient clinical information system (CIS) (iMD-Soft: Metavison®, Tel Aviv, Israel), including the Acute Physiology and Chronic Health Evaluation IV predicted mortality (APACHE IV PM) score [14] and first day Sepsis-related Organ Failure Assessment (SOFA) score [15].

Continuous QTc measurement

All patients were monitored with a patient monitor by a 5-electrode ECG able to obtain 7 leads (I, II, III, aVR, aVL, aVF, V1). The novel continuous QT/QTc measurement software was developed by Philips Healthcare (Philips Healthcare, Andover, MA, USA, software revision G.01.78) and approved for clinical use in 2008 [7,12]. This software determined the QT interval per minute by constructing a root-mean-squared waveform, meaning a combined beat from all seven leads. Normal or atrial paced beats and beats with a similar morphology are averaged to form a representative waveform for further analysis. Onsets and offsets are derived from this representative waveform. Every five minutes, a mean of these QT measurements was calculated and displayed. The Bazett formula was used to calculate QTc: $QTc = QT/\sqrt{(RR)}$ [4]. RR is the interval between two R waves in the ECG. All the QTc interval data were stored in a clinical information system (MetaVision; iMDsoft, Tel Aviv, Israel).

QTc measurement on 12-lead ECG

On a 12-lead ECG, QT, QRS and RR interval were measured manually on a single selected lead by one of the investigators, blinded for both the results of the continuous measurements and automatically derived QTc on the 12-lead ECG. For the manual measurements, preferably lead II was used. If the QT measurement was hampered because of an uninterpretable T wave, another lead was used with the best interpretable T wave. In case of atrial fibrillation three consecutive RR intervals were measured and divided by three. QTc was calculated using Bazett's formula.

Furthermore, automatically derived QTc as described in the GE Healthcare physician's guide on the 12-lead ECG (General Electric MAC 5500 HD, 25 mm/sec and 10 mm/mV) were noted [16]. On the General Electric ECG apparatus heart rate is calculated by dividing all QRS complexes minus one by the time difference between the first and the last beats. For each of the 12 leads a representative (median) QRS complex is generated, after which the earliest onset of the QRS complex of any lead is taken as start of the QRS complex. Also, for each of the 12 leads a representative (median) T wave is generated. The latest offset of the T wave of any lead is taken as the end of the T wave. Onsets are defined as the earliest deflection in any lead, and offsets are defined as the latest deflection in any lead. Bazett's formula was used to calculate QTc. QTc measurement by hand or computer assisted manual measurement is usually recommended as being most reliable, however, manual measurements have some inherent weaknesses [17].

Statistical methods

A sample size calculation showed that 117 patients with an observation of each measurement method achieves 80% power for an Intraclass correlation of 0.80 at the alternative hypothesis when Intraclass correlation at zero hypothesis is 0.70, using a F-test with 0.05 significance.

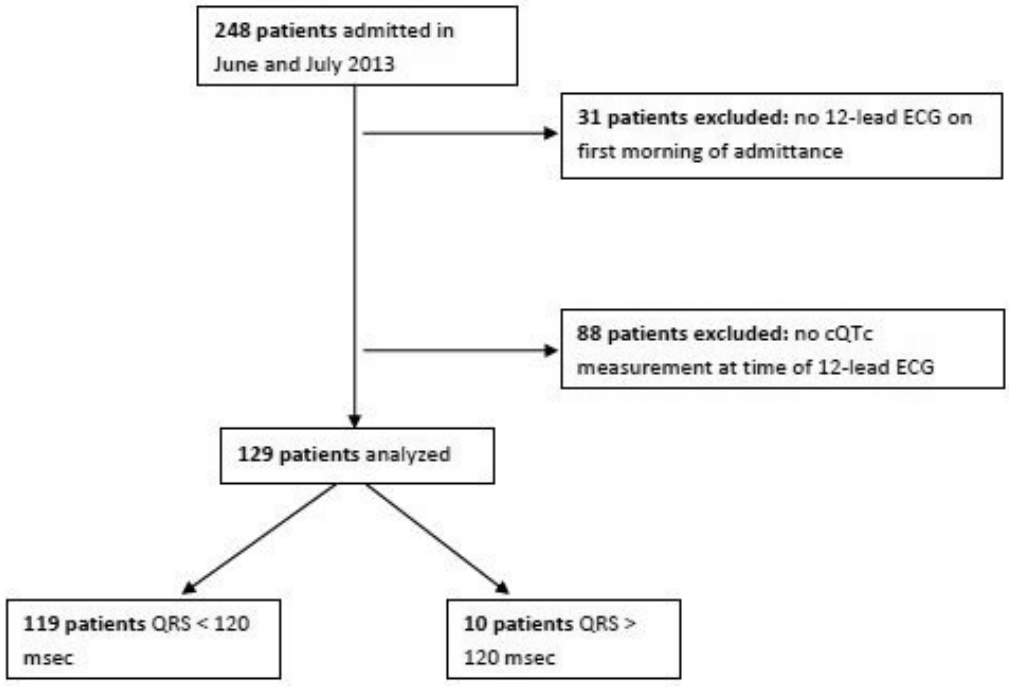
Baseline characteristics were expressed with a mean and standard deviation (SD) for continuous normally distributed variables, median and interquartile range (IQR) for continuous non-normally distributed or ordinal variables and percentages for categorical variables.

Pearson's correlation coefficient was used to evaluate the relationship between the measurements of normal distributed QTc measurements.

Inter-rater reliability of the continuous QTc measurement and QTc on the 12-lead ECG was tested by the calculation of Intraclass correlation coefficients (ICC) for all assessments (two way random consistency single measures) [18]. Values between 0.70 and 0.80 are considered as acceptable, values more than 0.80 as good [19].

All p-values were two-tailed and a p <0.05 was considered to be statistical significant. Data were analyzed with SPSS version 18.0 (SPSS, Inc. Chicago IL, USA).

Agreement between continuous QTc measurement and manually measured and calculated QTc on the 12-lead ECG was assessed using the Bland-Altman method with calculations of the limits of agreement (i.e. the mean difference ± 1.96 SD). The mean difference indicates any possible bias of one method over another, and the limits of agreement indicates the variability in the differences [20].



cQTc= continuous QTc measurement; msec= milliseconds

Figure 1. Patient inclusion

RESULTS

Two hundred and forty-eight patients were admitted to the OLVG hospital ICU in the study period (Figure 1). One hundred and nineteen patients were excluded. Patient baseline characteristics are described in Table 1.

Manually measured and calculated QTc vs. patient monitor QTc, QRS <120 msec

For the 119 patients with a QRS duration shorter than 120 msec mean continuous QTc for the patient monitor was 468 msec (SD 37 msec) and for manually measured QT and RR interval and calculated QTc on the 12-lead ECG was 449 msec (SD 41 msec) (Table 2).

Pearson's correlation between continuous QTc measurement with the patient monitor and the manually measured and calculated QTc on the 12-lead ECG was 0.65 (p= 0.01). The intraclass correlation coefficient was 0.65 (95% CI: 0.53 - 0.74, p= 0.001) as shown in Table 3. The Bland-Altman plot (Figure 2) shows a mean difference of 19.5 msec (limits of agreement (LOA) -44.6 - 83.7) with seven outliers (Table 4). The five negative outliers had all low T wave voltages on the 12-lead ECG in all leads.

Automatically derived QTc vs. patient monitor QTc, QRS <120 msec

For the 119 patients with a QRS duration shorter than 120 msec mean QTc for the automatically derived QTc on the 12-lead ECG was 461 msec (SD 35 msec) (Table 2).
Pearson's correlation between continuous QTc measurement with the patient monitor and the automatically derived QTc on the 12-lead ECG was 0.77 (p= 0.01). The Intraclass correlation coefficient was 0.77 (95% CI: 0.68-0.83, p= 0.001) (Table 3). The Bland-Altman plot (Figure 3) shows a mean difference of 7.8 msec (LOA -40.2 - 55.8) with five outliers (Table 4).
Bland-Altman plot for 12-lead ECG vs. manually measured and calculated QTc is shown in Figure 4.

Manually measured and calculated QTc vs. patient monitor QTc, QRS >120 msec

For the 10 patients with a QRS duration longer than 120 msec mean QTc for the patient monitor and manually measured and calculated QTc on the 12-lead ECG was 492 (SD 53) and 482 (SD 48) respectively (Table 2). Pearson's correlation was 0.65 (p= 0.023). The Intraclass correlation coefficient was 0.61 (95% CI: 0.11 - 0.89) (Table 3).

Automatically derived QTc vs. patient monitor QTc, QRS>120 msec

For the 10 patients with a QRS duration longer than 120 msec mean QTc for the automatically derived QTc on the 12-lead ECG was 489 msec (SD 64) (Table 2). Pearson's correlation was 0.84 (p= 0.002). The Intraclass correlation coefficient was 0.86 (95% CI: 0.53 - 0.96) (Table 3).

Table 1. Patient baseline characteristics

Variables	Patients n= 129 (100%)
Sex	
Male	95 (73.6)
Female	34 (26.4)
Age	67 [57.5 – 75]
APACHE IV pm [1]	3.2 [0.86 – 25.7]
SOFA day 1 [2]	6.0 [4.0 – 7.0]
Specialty of admittance	
Thoracic surgery	79 (61.2)
Internal medicine	13 (10.1)
Pulmonary medicine	13 (10.1)
General surgery	12 (9.3)
Cardiology	7 (5.4)
GE[3] and liver diseases	1 (0.8)
Neurology	1 (0.8)

Oncology	1 (0.8)
Orthopaedic surgery	1 (0.8)
Urology	1 (0.8)
Broadened QRS	
LBBB[4]	4 (3.1)
RBBB[5]	6 (4.7)
Ventricular pacemaker	0 (0)
Heart rhythm	
Sinus rhythm	119 (92.2)
Atrial fibrillation	`5 (3.9)
Junctional rhythm	4 (3.1)
Atrial pacemaker	1 (0.8)
Heart rate	85 (14.7)

Data are shown as number (%) or median [inter quartile range] or mean (Standard Deviation)
[1] APACHE IV p.m. = Acute Physiology and Chronic Health Evaluation IV predicted mortality;
[2] SOFA = Sepsis-related Organ Failure Assessment score;
[3] GE = gastroenterology;
[4] LBBB = left bundle branch block;
[5] RBBB = right bundle branch block.

Table 2. Mean QTc measurements

	Mean manually measured and calculated QTc 12-lead ECG in msec	Mean cQTc patient monitor in msec	Mean automatically derived QTc 12-lead ECG in msec
QRS<120 msec (n= 119)	449 (41)	468 (37)	461 (35)
QRS>120 msec (n= 10)	476 (46)	492 (53)	489 (64)

Data are shown as mean with (SD); Msec = milliseconds; SD = standard deviation

QRS < 120 msec

QTc manually measured vs. QTc patient monitor:	*p*= 0.000	95% Confidence Interval [13.6 - 24.5] *
QTc manually measured vs. QTc automatically 12-lead:	*p*= 0.000	95% Confidence Interval [5.9 - 17.6] *
QTc patient monitor vs. QTc automatically 12-lead:	*p*= 0.001	95% Confidence Interval [-12.2 - -3.4] *

* *Paired-Samples T-test*

QRS > 120 msec

QTc manually measured vs. QTc patient monitor:	*p*= 0.203	Z-1.274 **
QTc manually measured vs. QTc automatically 12-lead:	*p*= 0.203	Z-1.274 **
QTc patient monitor vs. QTc automatically 12-lead:	*p*= 0.859	Z-0.178 **

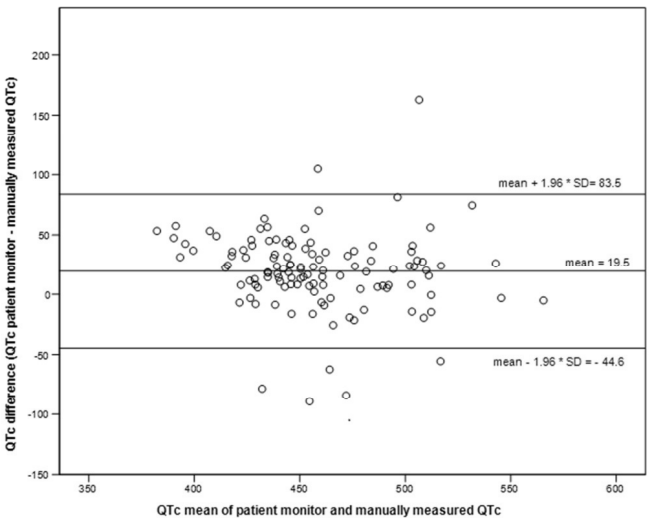
** *Wilcoxon Signed Ranks Test*

Table 3. Intraclass correlation coefficients

Measurement	n	ICC [95% CI]#	*p*-value	Correlation*	*p*-value
QRS <120 msec					
Manually measured and calculated QTc vs. patient monitor QTc	119	0.65 [0.53 - 0.74]	0.000	0.65	0.000
Automatically derived QTc vs. patient monitor QTc	119	0.77 [0.68 - 0.83]	0.000	0.7	0.000
Manually measured and calculated QTc vs. automatically derived QTc 12-lead	119	0.64 [0.52 - 0.74]	0.000	0.65	0.000
QRS >120 msec					
Manually measured and calculated QTc vs. patient monitor QTc	10	0.59 [-0.02 - 0.88]	0.023	0.61	0.06
Automatically derived QTc vs. patient monitor QTc	10	0.86 [0.53 - 0.96]	0.000	0.88	0.001
Manually measured and calculated QTc vs. automatically derived QTc 12-lead	10	0.28 [-0.39 - 0.75]	0.204	0.24	0.511

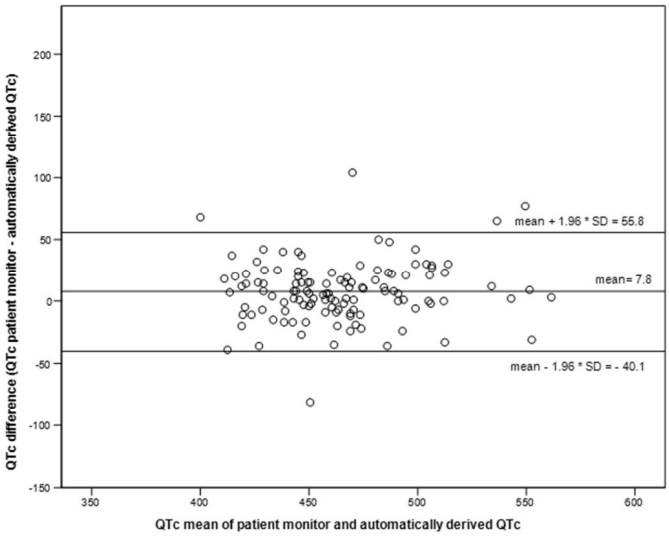
Intraclass correlation coefficient two-way random consistency single measures with 95% confidence interval

* Pearson correlation



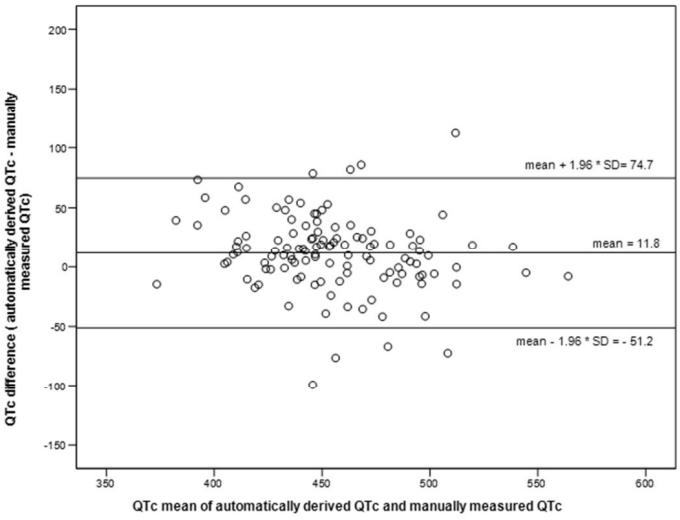
Mean difference = 19.5 msec. Limit of agreement (LOA) = -44.6 - 83.7 SD = standard deviation

Figure 2. Bland-Altman plots showing mean QTc difference between patient monitor continuous QTc measurements and manually measured and calculated QTc's on 12-lead ECG (QRS <120 msec)



Mean difference = 7.8 msec. Limits of agreement (LOA) = -40,1 - 55.8 SD = standard deviation

Figure 3. Bland-Altman plot showing mean QTc difference between patient monitor continuous QTc measurements and automatically derived QTc on 12-lead ECG (QRS <120 msec)



Mean difference = 11.8 msec. Limits of agreement (LOA) = -11.3 - 34.8 SD = standard deviation

Figure 4. Bland-Altman plot showing mean QTc difference between automatically derived QTc on 12-lead ECG and manually measured and calculated QTc (QRS <120 msec)

DISCUSSION

In this study we compared the continuous QTc measurement on a patient monitor against manually measured QRS, QT and RR interval and calculated QTc on standard 12-lead ECG in critically ill patients. We have shown that a moderate to strong significant correlation between the manually measured and calculated QTc on the 12-lead ECG and continuous QTc on the patient monitor for QRS duration shorter than 120 msec exists. In addition, we found that the correlation between automatically derived QTc on a 12-lead ECG and the continuous QTc on the patient monitor is even better. The QTc derived from manually measured QT and RR interval, or computer assisted hand measurement, is recommended as most reliable, but has been shown to have limited intra- and interobserver reliability [17, 23, 24]. This could explain the better correlation we found between both automatic QTc measurements.

The continuous registration and automatically derived QTc on the 12-lead ECG appear to have a higher mean QTc compared to the manual measured and calculated QTc on the 12-lead ECG recording. The comparison between automatically derived QTc's (patient monitor vs. automatically derived QTc) shows a mean 7.8 msec longer QTc on the monitor than the automatically derived QTc. This small and clinical irrelevant difference is hard to understand. The outliers that caused this difference probably have had an inaccurate determination of the T wave and as a consequence an overestimation QTc measurement by the patient monitor. The comparison between manually measured and calculated QTc and continuous QTc on the patient monitor has a bias of 20 msec. Such a difference is in concordance with previous research and can be explained by the use of more leads by the patient monitor for analysis instead of one lead with the manual measurement to calculate QTc [23]. As such, the continuous measurement is slightly more sensitive in finding prolonged QTc. For clinical practice, it is important and reassuring to realise that the automatically and continuously measured QTc is often longer than the QTc as manually measured in an individual lead on which normal QTc values are based [23].

Determination of the end of the T wave can be very difficult by hand measurement, but possibly also for the continuous registration by the patient monitor and 12-lead ECG, and is sometimes impossible [23]. This is shown by the five negative outliers on the Bland-Altman plot (Figure 2) which all had low voltage T waves. Also, it is possible that a substantial amount of excluded patients, who had no continuous QTc measurements at the time of performing the 12-lead ECG, were excluded due to impossible wave analysis by the patient monitor. Another reason for absence of the continuous QTc measurement could be that the patient monitor was not set to calculating continuous QTc at patient admittance. Due to the retrospective character of this study the reasons for absent continuous QTc cannot be differentiated.

Limits of agreement as shown by both Bland-Altman plots are fairly large. This can be explained by a known large QTc difference in different leads. Some regard differences of up to 50 msec in QT intervals measured in the various leads in normal subjects as being normal [24], others have suggested that differences of up to 65 msec were still within the limit of normal [23, 25].

Limitations

This study has several strengths and weaknesses. It is the first study to validate in a critical care setting the continuous measurement of QTc. We included enough patients according to the sample size calculation to draw conclusions about the patients with a QRS duration shorter than 120 msec. However, for QRS duration longer than 120 msec unfortunately the number of patients was too low to draw reliable conclusions and further study in this patient group is needed. We chose one time point per patient but due to the variability over the course of illness it is unclear whether our findings are consistent over time for an individual patient. We did cover a wide range of QTc values and showed consistent findings in the Bland-Altman plot over this range. There is a relatively large group of cardiothoracic surgery patients, which might influence the validation. In addition, patients with atrial fibrillation have a beat by beat change in QTc which may be difficult to measure by hand. In our analysis, however, only one outlier in the comparison of continuous QTc and manual measurement had atrial fibrillation. No outliers in the 12 lead ECG comparison with manual measurement were patients with atrial fibrillation. We conclude that patients with atrial fibrillation did not influence our results. Next to that, we chose for one observer for measuring the 12-lead ECG's. There might be benefits in manual measurements by more observers or more measurements by one observer. For clinicians on the intensive care this study is reassuring that the continuous QTc registration is reasonably accurate and is safe to use in clinical practice.

CONCLUSIONS

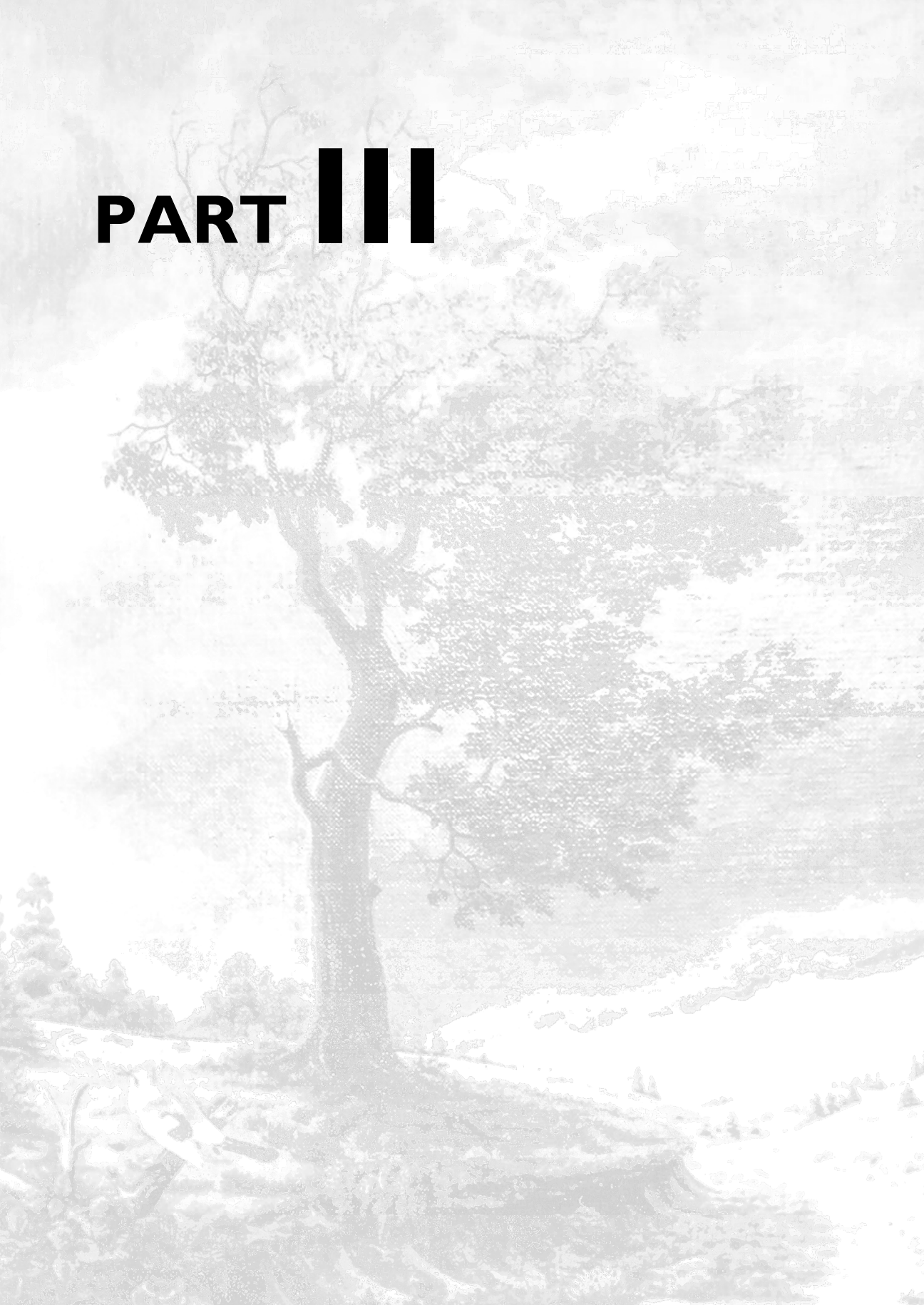
Continuous QTc measurement in critically ill patients with a QRS duration shorter than 120 msec shows an acceptable accuracy to be used in routine care. However, additional research in subgroups of patients is necessary.

REFERENCES

1. El Sherif N, Turitto G. Torsade de pointes. Curr Opin Cardiol 2003;18:6-13.
2. Yap YG, Camm AJ. Drug induced QT prolongation and torsades de pointes. Heart 2003;89:1363-1372.
3. Zeltser D, Justo D, Halkin A, Prokhorov V, Heller K, Viskin S. Torsade de pointes due to noncardiac drugs: most patients have easily identifiable risk factors. Medicine (Baltimore) 2003;82:282-290.
4. Drew BJ, Califf RM, Funk M Et al. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. Circulation 2004;110:2721-2746.
5. Benoit SR, Mendelsohn AB, Nourjah P, Staffa JA, Graham DJ. Risk factors for prolonged QTc among US adults: Third National Health and Nutrition Examination Survey. Eur J Cardiovasc Prev Rehabil 2005;12:363-368.
6. Tan HL, Hou CJ, Lauer MR, Sung RJ. Electrophysiologic mechanisms of the long QT interval syndromes and torsade de pointes. Ann Intern Med 1995;122:701-714.
7. Pickham D, Helfenbein E, Shinn JA et al. High prevalence of corrected QT interval prolongation in acutely ill patients is associated with mortality: results of the QT in Practice (QTIP) Study. Crit Care Med 2012;40:394-399.
8. Tisdale JE, Jaynes HA, Kingery JR, Mourad NA, Trujillo TN, Overholser BR, Kovacs RJ. Development and validation of a risk score to predict QT interval prolongation in hospitalized patients. Circ Cardiovasc Qual Outcomes. 2013 Jul;6(4):479-87. doi: 10.1161/CIRCOUTCOMES.113.000152.
9. Drew BJ, Ackerman MJ, Funk M et al. Prevention of torsade de pointes in hospital settings: a scientific statement from the American Heart Association and the American College of Cardiology Foundation. J Am Coll Cardiol 2010;55:934-947.
10. Hoogstraaten E, Rijkenberg S, van der Voort PH. Corrected QT-interval prolongation and variability in intensive care patients. J Crit Care. 2014 Oct;29(5):835-9. doi: 10.1016/j.jcrc.2014.05.005.
11. Helfenbein ED, Zhou SH, Lindauer JM et al. An algorithm for continuous real-time QT interval monitoring. J Electrocardiol 2006;39:S123-S127.
12. Philips Healthcare. Philips Healthcare information: QT/QTc Interval Monitoring, ST/AR Algorithm. IntelliVue Patient monitor and Information Center, Application Note. Internet communication 2013 May 1.
13. Craig P. Dobson, Maria Teresa La Rovere, Cara Olsen, Marino Berardinangeli, Marco Veniani, Paolo Midi, Luigi Tavazzi, Mark Haigney, on behalf of the GISSI-HF Investigators. 24-Hour QT variability in heart failure. Journal of Electrocardiology 2009;42:500–504. doi:10.1016/j.jelectrocard.2009.06.021
14. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006;34:1297-1310.
15. Vincent JL, Moreno R, Takala J, et al: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22:707-710.
16. GE Healthcare. Marquette™ 12SL™ ECG Analysis Program Physician's Guide 2036070-006 Revision A 30 August 2010:(3-13)-(3-15)
17. Darpo B, Nebout T, Sager PT. Clinical evaluation of QT/QTc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 guideline. J Clin Pharmacol. 2006 May;46(5):498-507.
18. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-428.
19. Streiner D.L, Norman G.R. Health measurement scales a practical guide to their development and use. third edition edn. Hamilton: Oxford university press; 2003.
20. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-160.

21. Charbit B, Samain E, Merckx P, Funck-Brentano C. QT interval measurement: evaluation of automatic QTc measurement and new simple method to calculate and interpret corrected QT interval. Anesthesiology. 2006 Feb;104(2):255-60.
22. Salerno S.M., Alguire P.C., Waxman H.S.; Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. Ann Intern Med. 2003;138:751-760.
23. Rautaharju PM, Surawicz B, Gettes LS, Bailey JJ, Childers R, Deal BJ et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. Circulation 2009;119: e241-e250.
24. Statters D.J., Malik M., Ward D.E., Camm A.J. QT dispersion: problems of methodology and clinical significance. J Cardiovasc Electrophysiol. 1997;5:672-685.
25. Surawicz B. Will QT dispersion play a role in clinical decision making? J Cardiovasc Electrophysiol. 1996;7:777-784.

PART III



Subcutaneous continuous glucose monitoring in critically ill patients

CHAPTER 8

Insulin treatment guided by subcutaneous continuous glucose monitoring compared to frequent point-of-care measurement in critically ill patients: a randomized controlled trial

D.T. Boom†, M.K. Sechterberger†, **S.Rijkenberg**, S. Kreder, R.J. Bosman, J.P.J. Wester, I. van Stijn, J.H. DeVries, P.H.J. van der Voort

†Equal contributors

Boom et al. Critical Care 2014,18:453

CHAPTER 9

The Clinical Benefits and Accuracy of Continuous Glucose Monitoring Systems in Critically Ill Patients -A Systematic Scoping Review

S.C.J. van Steen, **S. Rijkenberg**, J. Limpens, P.H.J. van der Voort, J. Hermanides, J.H. DeVries

Van Steen et al. Sensors (Basel, 2017,17(1):146

CHAPTER 10

Accuracy and reliability of a continuous glucose measurement device in critically ill patients

S.Rijkenberg, S.C. J. van Steen, J.H. DeVries, P.H. J. van der Voort

Accepted for publication in Journal of Clinical Monitoring and Computing, November 2017

The authors' own department funded this study

CHAPTER 8
Insulin treatment guided by subcutaneous continuous glucose monitoring compared to frequent point-of-care measurement in critically ill patients: a randomized controlled trial

ABSTRACT

Introduction

Glucose measurement in intensive care medicine is performed intermittently with the risk of undetected hypoglycemia. The workload for the ICU nursing staff is substantial. Subcutaneous continuous glucose monitoring (CGM) systems are available and may be able to solve some of these issues in critically ill patients.

Materials and Methods

In a randomized controlled design in a mixed ICU in a teaching hospital we compared the use of subcutaneous CGM with frequent point of care (POC) to guide insulin treatment. Adult critically ill patients with an expected stay of more than 24 hours and in need of insulin therapy were included. All patients received subcutaneous CGM. CGM data were blinded in the control group, whereas in the intervention group these data were used to feed a computerized glucose regulation algorithm. The same algorithm was used in the control group fed by intermittent POC glucose measurements. Safety was assessed with the incidence of severe hypoglycemia (<2.2 mmol/L), efficacy with the percentage time in target range (5.0 to 9.0 mmol/L). In addition, we assessed nursing workload and costs.

Results

In this study, 87 patients were randomized to the intervention and 90 to the control group. CGM device failure resulted in 78 and 78 patients for analysis. The incidence of severe glycemia and percentage of time within target range was similar in both groups. A significant reduction in daily nursing workload for glucose control was found in the intervention group (17 versus 36 minutes; p= 0.001). Mean daily costs per patient were significantly reduced with EUR 12 (95% CI -32 to -18, p= 0.02) in the intervention group.

Conclusions

Subcutaneous CGM to guide insulin treatment in critically ill patients is as safe and effective as intermittent point-of-care measurements and reduces nursing workload and daily costs. A new algorithm designed for frequent measurements may lead to improved performance and should precede clinical implementation.

INTRODUCTION

Stress-induced hyperglycemia is common and relates to adverse outcomes in critically ill patients [1,2]. The outcomes of two large intervention studies are in some way contradictory but the consensus is that hyperglycemia should be corrected, while avoiding hypoglycemia and high glucose variability [3-8]. On the basis of the available evidence, it seems preferable to maintain a blood glucose level around 8.0 mmol/L for the majority of critically ill patients [9,10].

Glucose regulation regimens require frequent monitoring of glucose, which leads to a considerable workload for the intensive care (IC) nurses. In addition, glucose regulation carries an inherent risk of insulin-induced hypoglycemia, which is associated with mortality [6]. Information about the glucose level is lacking for the period in between measurements with possible unnoticed hypoglycemic episodes. Continuous glucose monitoring (CGM) could be of value to facilitate or improve glycemic control. Previous studies have indicated an acceptable accuracy and reliability for subcutaneous CGM devices in critically ill patients [11-15]. The only prospective randomized controlled trial so far that assessed the role for CGM in glycemic control in critically ill patients showed that real-time CGM increased the safety of tight glycemic control in critically ill patients by significantly reducing severe hypoglycemic events [16]. However, an improvement of the mean glucose concentration by using real-time CGM was not found [16].

Thus, CGM may give us the ability to detect early (possible) hypo- and hyperglycemia as well as minimizing swings in glucose levels. Moreover, the use of CGM may facilitate the process of glycemic control and may reduce the number of blood samples and accompanying blood loss, nursing workload and costs. To date, there are few data available how CGM-driven glucose regulation compares to point-of-care (POC) -driven glucose regulation and no controlled studies specifically evaluated workload and cost of CGM. The aim of the present study was to assess the safety, efficacy, workload and costs of a subcutaneous CGM system-guided blood glucose regulation in comparison with frequent POC blood glucose-guided regulation in a mixed population of critically ill patients.

MATERIALS AND METHODS

Study design and participants

This was a randomized controlled open-label clinical trial, performed in a 20-bed mixed medical-surgical ICU of a teaching hospital (Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands). Patients were recruited over a period of 18 months from 2011 till late 2012.

Patients were eligible for inclusion within 24 hours after ICU admission if they were 18 years or older, in need of intravenous (i.v.) insulin treatment for glucose regulation and with an expected length of stay in the ICU of at least 24 hours. Patients could not be included if any of the following criteria was present: lack of informed consent, participation in another trial or previous participation in this trial or when a CGM system was currently not available. The study ended when patients were discharged from the ICU or because of technical failure of

the CGM device. The maximum study duration was set at five days for both treatment groups. The complete nursing staff was trained beforehand to handle all devices used in this study adequately. This study was approved by the ethics committee VCMO, Nieuwegein, The Netherlands and was in line with Dutch and European legislation. All patients or their legal representative provided written informed consent. This trial is registered with Clinical trials.gov, number NCT01526044.

Randomization

Patients who met the inclusion criteria were randomized in a 1:1 ratio with computerized block randomization to either the intervention group or the control group.

Study procedures

Algorithm

In all study participants, blood glucose regulation was performed by a sliding scale algorithm with a blood glucose target of 5.0 to 9.0 mmol/L, which was integrated into the patient data management system (PDMS, Meta- Vision; iMDsoft, Tel Aviv, Israel) [17]. Hypoglycemia was defined as a blood glucose level of <2.2 mmol/L in line with the Van den Berghe trial [3]. Below target was defined as a glucose level from 2.2 mmol/L till the lower target level of 5.0 mmol/L. Above target, all glucose levels were above 9.0 mmol/L. The algorithm instructed the insulin i.v. infusion rate (or glucose administration in case of hypoglycemia) after each glucose measurement. The time for the next glucose measurement was also defined from the algorithm and depended on the stability of the glucose level over time.

Glucose measurement

Study participants allocated to the intervention group received a subcutaneous CGM system (FreeStyle Navigator™, Abbott Diabetes Care, Alameda, CA, USA), which was used to guide blood glucose regulation. The nurses were trained to insert the subcutaneous glucose sensors on the patients' abdomen or upper arm. After insertion of the subcutaneous sensor, a transmitter was attached that connects through wireless communication to a receiver, which displays the real-time glucose readings every minute and stores glucose readings every 10th minute. The CGM system needed a one-hour stabilization period, in which glucose measurements were not performed. Calibration of the system using an arterial blood sample was performed five times in total, after 1, 2, 8 to 10, 24 to 32 and 72 to 80 hours, following manufacturer instructions. The CGM system sounded an alarm when additional calibrations were needed. On the times that the algorithm needed a new glucose measurement, the readings from the CGM system were entered in the computerized glucose regulation protocol that was embedded in the PDMS. Other CGM values were not used in the algorithm. The CGM system sounded an alarm when the glucose level was either <5.0 mmol/L or >9.0 mmol/L. When this occurred, the nurse entered this additional glucose level in the computerized protocol, which triggered the glucose algorithm to advise an insulin dosing

adjustment. The CGM repeated its alarm after 15 minutes when the glucose level was still out of target range. Again, this value was entered into the system and dose adjustments were made until target range was achieved. Every hypoglycemic event (<2.2 mmol/L) needed to be verified by an arterial blood glucose sample. In case of a discrepancy between the CGM value and the arterial blood glucose sample, the latter was leading in clinical decision-making.

Blood glucose regulation in the study participants allocated to the control group was performed by use of frequent point-of-care (POC) measurements using Accu-Chek™ (Roche/Hitachi, Basel, Switzerland). All blood samples were obtained from an indwelling arterial catheter. The displayed glucose levels were automatically stored in the PDMS. Participants in the control group also received a subcutaneous Freestyle Navigator CGM system, however, these data were blinded and not used for blood glucose regulation. Calibrations were performed following manufacturer instructions and no alarms were set.

In both groups arterial reference blood glucose samples were drawn six times daily at standardized times and analyzed by the ABL Flex automated blood gas analyzer (BGA) (Radiometer, Copenhagen, Denmark). These values were automatically stored into the PDMS but were blinded to both nurses and physicians.

Study endpoints

The primary safety outcome was the incidence of severe hypoglycemia (<2.2 mmol/L) during the intervention. Efficacy outcomes were the percentage of time that glucose levels were within the target range (5.0 to 9.0 mmol/L), below target range (2.2 to 5.0 mmol/L), and in the hyperglycemic range (>9.0 mmol/L). In addition, mean blood and sensor glucose levels and glucose variability defined as the mean absolute glucose (MAG) change (ΔGlucose/ ΔTime) were endpoints as well [8]. The accuracy of the CGM and the POC device was assessed by calculating the median relative absolute deviation (RAD) between reference glucose and CGM or POC glucose.

Nursing workload for glucose control per day was determined by the number of POC measurements or measurements from the sensor, which were entered in the computerized glucose regulation protocol and the amount of calibrations of the CGM sensor (in the intervention group only). A time-in-motion design was used to estimate the time that it took to execute targeted glucose control and insulin treatment per group. The following subtasks were observed: (1) POC measurement (this included the initiation, blood sampling, blood testing and processing), (2) sensor placement, (3) sensor calibration and (4) time needed to determine a CGM value and entering the value in the decision support module. The tenfold-recorded elapsed times per subtask were averaged and then multiplied by the 24-hour blood sample average collected from the clinical trial.

Cost analysis was performed from a health-care payer perspective with a one-day (24 hours) time horizon. The outcome measure in the economic evaluation was the costs per patient for glycemic control in 24 hours. Cost parameters included nursing personnel costs, device costs, materials needed for glucose monitoring and laboratory costs. Cost estimates for the parameters were derived from the hospital and laboratory ledger, devices manufacturers' data

and the Dutch guide for health economic research [18]. Costs are expressed in euros and are based on the year 2013. Because of the short time horizon of this analysis (24 hours), the costs were not discounted.

Data collection

Clinical and laboratory baseline data were extracted from the PDMS after randomization: demographic data, body mass index (BMI), reason for ICU admission, history of diabetes, history of renal failure, severity of disease scores (the sequential organ failure assessment (SOFA) score and acute physiology and chronic health evaluation (APACHE IV) score at admission), blood glucose levels at admission and the use of mechanical ventilation. Blood glucose data, that is reference arterial blood glucose samples and glucose values that were entered in the decision support module (CGM measurements in the intervention group, POC measurements in the control group) were also extracted from the PDMS. Continuous glucose data from the CGM device were uploaded to a computer using CoPilot™ Health Management System for FreeStyle Navigator (Abbott Diabetes Care, Alameda, CA, USA) and entered in the study database. All reference glucose measurements were linked by time with the concomitant CGM measurements and Accu-Chek measurements.

Statistical analysis

A sample size of 160 (80 participants in each group) conferred 80% power, with two-sided p= 0.05, to detect an absolute difference of 10% in the incidence of severe hypo- or hyperglycemia between the intervention and the control group. A total sample size of 178 patients (89 patients per group) is needed to correct for an expected 10% drop out. Results are expressed as percentages for categorical variables, mean and standard deviation (SD) for continuous normally distributed variables, and median and interquartile range (IQR) for continuous non-normally distributed variables. Groups were compared by using Fisher's exact test, Student's t test or Mann-Whitney rank-sum test as appropriate. Median RAD was calculated instead of mean because of its skewed distribution. Costs were calculated as the summed product of factors and resources used and their respective unit costs and were averaged per patient per day. Because of skewed (cost) distributions, we assessed group contrasts by calculating 95% confidence intervals for the mean differences following bias-corrected and accelerated nonparametric bootstrapping, that is drawing 1,000 samples of the same size as the original sample separately for each group. All statistical analyses were performed in SPSS 20.0 (IBM Corp, Armonk, NY, USA).

RESULTS

A total of 178 patients were randomized to either the intervention or the control group (Figure 1). Most of the patients who were not eligible were postoperative cardiac surgery patients with an expected length of stay (LOS) <24 hours. One patient was incorrectly randomized and did not receive a CGM device. Nine patients in the intervention group and

twelve patients in the control group were excluded from analysis due to lack of CGM data because of technical failure of the device, misplacement of the sensor (n= 3) and problems with extraction of the data (n= 18).

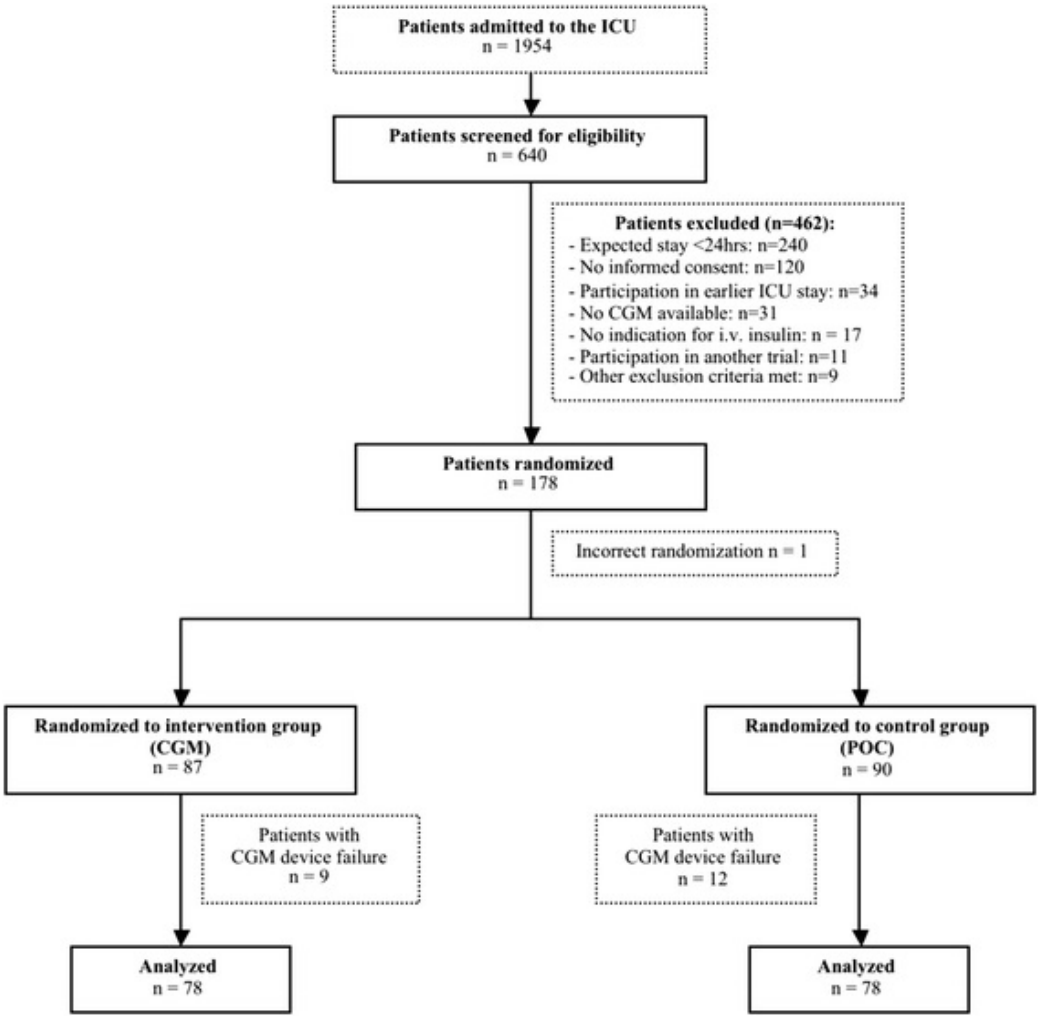


Figure 1. Flow chart of study participants: assessment, randomization and analysis

We performed a per protocol analysis from the data of 78 patients in each group. Table 1 shows the two groups, which were well matched with respect to all baseline characteristics.
During the intervention, a total of 37,570 (intervention group) and 32,957 (control group) CGM measurements were collected. The number of reference arterial blood gas glucose measurements was 1,599 in the intervention group and 1,325 in the control group. The median number of additional calibrations needed for the CGM was 1.9 per 24 hours (IQR 1.2 to 3.3). The number of glucose values entered in the PDMS (CGM measurements in the

intervention group and POC measurements in the control group) was 3,919 and 2,489 respectively.

Table 2 summarizes the outcome measures of the study. The incidence of hypoglycemia (<2.2 mmol/L), the primary safety endpoint, was similar in both the intervention and the control group. None of the severe hypoglycemic episodes detected by the CGM in the intervention group was verified by arterial blood sampling. In the control group, all severe hypoglycemic episodes detected by the CGM, occurred in between two POC glucose measurements and were not detected by the nurses. In total, there were 14 patients (3 patients in the control group and 11 patients in the intervention group) who experienced 19 'true' hypoglycemic events (<3.9 mmol/L) detected by ABL. Twenty-five percent (n= 4) of the 'true' hypoglycemic events in the CGM group and 67% (n= 2) in the control group were also identified by CGM or POC (difference in glucose ≤10%). All other endpoints such as percentage time in target range, below target range, mean reference and sensor glucose, glucose variability, hospital LOS, ICU and hospital, mortality were nonsignificantly different between the study groups. Moderate hyperglycemia (9.0 to 11.1 mmol/L) was significantly different in favor of the intervention group (p= 0.03). A total of 355 time-linked reference glucose CGM samples and 85 time-linked reference glucose POC samples were used to assess accuracy of the devices. Median (IQR) RAD of the POC device was 7.1% (3 to 12) whereas the median RAD of the CGM device was 13.7% (8 to 23) (p= 0.001). Bland-Altman plots per glucose monitoring system are shown in an additional file (Figure S1 in Appendix).

Table 3 summarizes nursing workload data per 24 hours. The first column displays the average time burden per subtask of glucose control. The average total time burden for glucose control was significantly lower in the intervention group compared to the control group (17 minutes versus 36 minutes; p= 0.001). The mean reduction in total nursing workload was 19 minutes per 24 hours or 53% in favor of the intervention group. As in this study, an open blood drawing system was used, 5 mL blood per POC measurement or calibration was taken from the patient. Blood loss was therefore significantly reduced in the intervention group (15.3 mL versus 60 mL per day; p= 0.001).

The economic analysis of both groups is shown in Table 4. The intervention group generated an average total daily cost of EUR 41, whereas the total daily cost in the control group was EUR 53. The difference in costs was EUR -12 in favor of the intervention group (95% CI -32 to -18, p= 0.02). The extra costs of the CGM devices in the intervention group were neutralized by the diminished costs for nursing personnel, material and laboratory costs.

Table 1. Baseline characteristics of participants

	Intervention – CGM[6] (n= 87)	Control – POCM[7] (n= 90)
Age (years)	66.4 (14 - 0)	67.2 (11 - 4)
Women	45 (52%)	35 (39%)
BMI[1] (kg/m[2])	27.8 (7 - 0)	27.4 (5 - 8)
Weight (kg)	81.8 (21 -7)	83.2 (21 - 5)
History of diabetes*	18 (21%)	21 (23%)
History of renal failure**	10 (12%)	5 (6%)
Reason for ICU[2] admission		
Surgical		
Elective	19 (22%)	16 (18%)
Emergency	12 (14%)	13 (14%)
Medical	56 (64%)	61 (68%)
Admission diagnosis		
Post cardiac surgery	12 (14%)	11 (12%)
Severe sepsis/septic shock	23 (26%)	18 (20%)
Pneumonia	12 (14%)	11 (12%)
Cardiac failure	10 (12%)	9 (10%)
COPD[3]	3 (3%)	8 (9%)
Hemorrhagic shock	7 (8%)	10 (11%)
Cardiac Arrest/resuscitation	10 (12%)	14(16%)
Other	10 (12%)	9 (10%)
APACHE IV[4] predicted mortality (%)	32 (10 – 70)	31 (20 – 60)
SOFA[5] score on admission	8 (6 – 10)	7 (6 – 10)
Blood glucose level on admission (mmol/L)	9.0 (2 – 6)	9. 2 (2 – 5)
Mechanical ventilation	80 (92%)	83 (92%)

Data are mean (SD), median (IQR) or n (%)

[1] BMI: body mass index

[2] ICU: intensive care unit

[3] COPD: chronic obstructive pulmonary disease

[4] APACHE: acute physiology and chronic health evaluation

[5] SOFA: sepsis-related organ failure assessment

[6] CGM: continuous glucose monitoring

[7] POCM: point-of-care measurement

* Diabetes was defined as present when this diagnosis was mentioned in the medical history.

** Renal failure was present when the pre-admission serum creatinin was above 177umol/L.

Table 2. Safety, efficacy and clinical study outcomes

	Intervention- CGM[1] (n= 78)	Control – POCM[2] (n= 78)	*p*-value
Study period (days)	3.2 (2 – 5)	2.8 (1 – 5)	0.18
Incidence severe hypoglycemia (<2.2 mmol/L) #	None	None	
Detected by CGM[1]			
Number of subjects	3 (4%)	4 (5%)	1.0
Episodes < 2.2 mmol/L	3	4	
% of time for the reference glucose level (SD)[3]			
In target range (5.0 to 9.0 mmol/L)	69 (26)	66 (26)	0.47
Below target range (2.2 to 5.0 mmol/L)	5 (7)	3 (5)	0.21
Mild moderate hypoglycemia (2.2 to 3.9)	1 (3)	0 (1)	0.03
Above target range (>9.0 mmol/L)	28 (26)	34 (27)	0.06
Mild moderate hyperglycemia (9.0 to 11.1)	17 (16)	26 (23)	0.01
Hyperglycemia (>11.1)	11(19)	7(14)	0.19
% of time for the sensor glucose levels (SD)[3] ##			
In target range (5.0 to 9.0 mmol/L)	75 (18)	71 (20)	0.18
Below target range (2.2 to 5.0 mmol/L)	11 (13)	9 (12)	0.44
Mild moderate hypoglycemia (2.2 to 3.9)	2 (7)	1 (2)	0.14
Above target range (>9.0 mmol/L)	15 (16)	20 (21)	0.06
Mild moderate hyperglycemia (9.0 to 11.1)	12 (11)	16 (16)	0.03
Hyperglycemia (>11.1)	3 (7)	4 (9)	0.35
Mean reference blood glucose (mmol/L)	8.2 (1.6)	8.3 (1.3)	0.53
Mean sensor glucose (mmol/L)	7.1 (1.1)	7.5 (1.3)	0.07
MAG[4] change (mmol/L/h) ###	0.33 (0.2 - 0.5)	0.32 (0.2 - 0.4)	0.31
LOS[5] ICU[6] (hours)	137 (71 – 250)	95 (51 – 157)	0.04
LOS hospital (days)	15 (8 – 270)	14 (8 – 31)	0.91
Mortality ICU	15 (19%)	12 (15%)	0.67
Mortality hospital	22 (28%)	17 (22%)	0.46

Data shown are mean (SD), median (IQR), or n *(%)*

[1] CGM: continuous glucose monitoring

[2] POCM: point-of-care measurement

[3] SD: standard deviation

[4] MAG: mean absolute glucose change

[5] LOS: length of stay

[6] ICU: intensive care unit.

Patients who experienced at least one severe hypo- or hyperglycemic episode, verified by blood gas analysis.

Percentages do not add up to 100 due to rounding off.

When at least three reference glucose measurements were available (intervention n= 73, control n= 71).

Table 3 Nursing workload per day (24 hours)

	Time per action (min)	Nr[4] of actions control group	Nursing time control group (min)	Nr of actions intervention group	Nursing time intervention group (min)
POC[1] measurement	3	12 (8)	36 (24)	0.06 (0.4)	0.2 (0.4)
Sensor CGM[2] placement	3 ·5	-	-	1	3.5
Sensor CGM calibration	2 ·5	-	-	1.9 (1.2 - 3.3)	8 (11)
Sensor CGM data to enter in PDMS[3]	0 ·3	-	-	18 (10)	5.3 (3)
Total time (min)			**36 (24)**		**17 (12)***

Data are expressed as mean (SD), or median (IQR) * p= *0.001 in comparison with control group*

[1] POC: point-of-care

[2] CGM: continuous glucose monitoring

[3] PDMS: Patient Data Management System

[4] Number

Table 4. Cost analysis

	Costs per unit	Factor control group	Costs control group	Factor intervention group	Costs in intervention group	Difference in cost (95%C.I.)[3]
Nursing time	€38/hr	36 min	€22.98	17 min	€10.87	€-12.11 (–16 - –9)
CGM receiver	€1009.59	-	-	€1.38 per day[4]	€1.38	€1.38
CGM Sensor	€61.00	-	-	€24.40 per day[5]	€24.40	€24.40
CGM calibration[6]	€1.19	-	-	3.3	€3.95	€3.95 (3,5)
Accu-Chek Inform II device	€892.37	€1.22 per day[2]	€1.22	-	-	€-1.22
Material POC measurement[7]	€0.70	12.2	€8.51	0.06	€0.04	€-8.47 (–10 - -7)
Laboratory[8]	€1.66	12.2	€20.18	0.06	€0.10	€-20.08 (–23 - -18)
Total costs			**€52.89**		**€40.74**	**€-12.42 (–22 - –5)**

Factors and costs are expressed as means per patient per day (24-hour)

[1] POC: point-of-care measurement [2] CGM: continuous glucose monitoring

[3] 95% confidence interval based on 1000 stratified bootstrap samples [4] Assuming a lifetime of two years [5] Assuming a manufacturers' sensor lifetime of two and a half days

[6] Calibration strip CGM [7] Includes syringes, non-sterile gloves, gauzes, alcohol, cap (used for blood sampling) and testing strip POC

[8] Costs for a single point-of-care glucose measurement

DISCUSSION

The present study showed that a subcutaneous CGM system to guide blood glucose regulation was equally effective and safe in glycemic control compared to frequent POC-guided blood glucose regulation. However, CGM significantly reduces nursing workload, blood loss and the daily costs for glucose control.

Comparison with other studies

This is the second but largest randomized controlled trial in which CGM is used to guide glycemic control in critically ill patients. In contrast to our findings, Holzinger and colleagues did find less severe hypoglycemia in the CGM group [16]. This may be caused by the very low incidence of severe hypoglycemia in the present study, which was true for both the intervention and the control group. This may be related to a change of policy after the publication of the NICE-SUGAR trial [4], which was a reason for our and most other ICUs to increase their blood glucose target range. The increased target range may have reduced the incidence of hypoglycemic events [19,20]. Indeed, the blood glucose target used in the current study (5.0 to 9.0 mmol/L) was higher than in the Holzinger trial [16] (4.4 to 6.1 mmol/L) and this is reflected in the achieved mean blood glucose levels (8.1 vs. 6.3 mmol/L). Moreover, the use of a fully computerized algorithm for glucose control and the high familiarity of the protocol among our IC nurses may have contributed to the low incidence of severe hypoglycemia. The available studies to date on tight glucose control showed an increase in nursing workload [21-23]. The potential benefits of CGM in the reduction of blood samples, blood loss and nursing workload was assumed in previous studies, but was not systematically assessed before. We now observed that CGM significantly reduced the amount of blood samples and the daily nursing workload for glucose control up to 53%. This finding seems clinically relevant, especially in a busy clinical IC environment. Two studies focused on the cumulative nursing workload accompanied with tight glucose control protocols [21,22]. Gartemann et al. estimated that nurses devoted approximately 42 minutes during a 12-hour shift of their time to administering a tight glycemic control (TGC) protocol, whereas Aragon et al. even reported that up to 2 hours might be required for tight glycemic control for a single patient in a 24-hour period. In our POC control group, the mean nursing workload estimate was less (36 minutes per 24 hours) than the published estimates reported by other groups. This might partly be explained by the use of a fully computerized algorithm for glucose control in our ICU. In addition, the familiarity of the protocol is very high among our ICU nurses.

Effectiveness and costs

The use of CGM did not achieve improved glycemic control in our study. We found similar percentages of time-in-target and below-target range between the study groups. The not-significantly lower percentage of time in the hyperglycemic range in the intervention group could be explained by the fact that CGM measurements were more frequently entered in the glucose protocol than POC measurements in the control group. This probably resulted in

more adjustments in the insulin treatment with lower blood glucose levels as a consequence. The significantly increased ICU LOS, which was observed in the intervention group, may be a coincidence or reflect unmeasured case-mix factors but is, in our view, unrelated to the glucose measurement strategy.

In contrast to our expectations, the cost analysis shows that the use of CGM systems for glucose control in an ICU setting is not a priori an expense. However, we should be cautious in interpreting these results due to the rather short time horizon (24 hours) in the analysis of costs determination and the single-center study design. Also, cost savings cannot immediately be monetized due to the short time horizon used in this cost analysis.

Accuracy of the subcutaneous measurements

The subcutaneous Freestyle Navigator CGM device that we used in the present study showed a median RAD of 13.7%, which is higher than the 10.6 and 11.6% that was found in previous validation studies of this device in critically ill patients, suggesting an accuracy acceptable for clinical use [11,14]. The lag time that may be needed for the subcutaneous compartment to adapt to the intravenous compartment appeared not to be clinically relevant [11]. However, the accuracy as assessed in the current study seems to indicate a need for improvement, because the accuracy was less than the accuracy of the Accu-Chek and because a substantial number (75% in the CGM group and 33% in the control group) of hypoglycemic events was not detected. Of note, Leelarathna et al. [24] recently investigated whether there was a difference in accuracy of the Freestyle Navigator in a critical care setting using two methods of calibration: (1) calibration according to the manufacturer's instructions (1, 2, 10, and 24 h) or (2) calibration at variable intervals of 1 to 6 hours using ABG. Using enhanced calibration, at a median (interquartile range) every 169 (122 to 213) minutes, the absolute relative deviation was lower (7.0% (3.5 - 13.0) vs. 12.8% (6.3 - 21.8), p= 0.001). So, further significant improvements in accuracy may be obtained by frequent calibrations with ABG measurements. In the current study forced calibration was not possible, calibration was only performed when the CGM device indicated the need for calibration by itself.

In addition, technical problems with the subcutaneous CGM device were observed during the study and led to a 12% dropout. The most important reason was the temporary loss of sensor signal from several minutes to hours that resulted in a loss of data. Difficulties in the calibration process were also identified as the CGM could only be calibrated if the system indicated a calibration by itself, which occurred for median 1.9 times per 24 hours. Most of the technical difficulties, however, may have been due to lack of experience working with the CGM device despite the training of all ICU nurses. We expect such problems to be easily resolved with additional training and with the improved next generation Freestyle Navigator II, which has recently been introduced and showed good utility and sensor performance in critically ill patients [25]. This study aimed to define safety, efficacy and costs and therefore we neglected the system dropout at this moment. It is true, however, that this device can only become part of routine care when the dropout percentage diminishes.

Strengths and limitations

The strengths of our study include the relatively large sample size, the randomized controlled study design and the wide variety in case mix. However, some limitations of the present study merit further consideration. First, the study was performed in a single Dutch intensive care unit, which limits the generalizability of the study. Second, the study was designed to blind the values of the CGM in the control group. However, the CGM needed to be calibrated several times during the study period, which made it impossible to blind it completely. Third, the nursing staff did not verify the severe hypoglycemia that was indicated by CGM in two of the three patients despite specific instructions to do so. One of these two patients had evolved into a 'withholding care policy', which was the reason to accept the severe hypoglycemia. We assume that in the other patient priority was given to other important nursing tasks. Thus, the available data are insufficient to define the accuracy of the CGM in the hypoglycemic range. In our previous studies this was not identified as a clinical problem [11,14]. Also, with an adapted algorithm, the CGM should be able to detect a decreasing glucose level before hypoglycemia is present and give a timely alert. Fourth, the computerized algorithm was designed for intermittent POC measurements and not for (semi-) continuous data. As such, the patients did not fully benefit from the frequent glucose measurements by CGM. An algorithm based on 10-minute glucose input might have led to other results. We did identify this issue beforehand but we decided to keep the algorithm for both groups the same to be able to investigate the contribution of CGM per se. It can be expected that an adapted algorithm will further improve the performance of CGM in the guidance of glycemic control.

CONCLUSION

Subcutaneous CGM to guide blood glucose regulation in critically ill patients was shown to be safe in terms of hypoglycemia incidence. With an identical insulin treatment algorithm, the CGM was equally effective as POC measurement. A new algorithm designed for frequent measurements may further improve the results and should precede clinical implementation. CGM significantly reduced nursing workload, blood loss and the daily costs for glucose control.

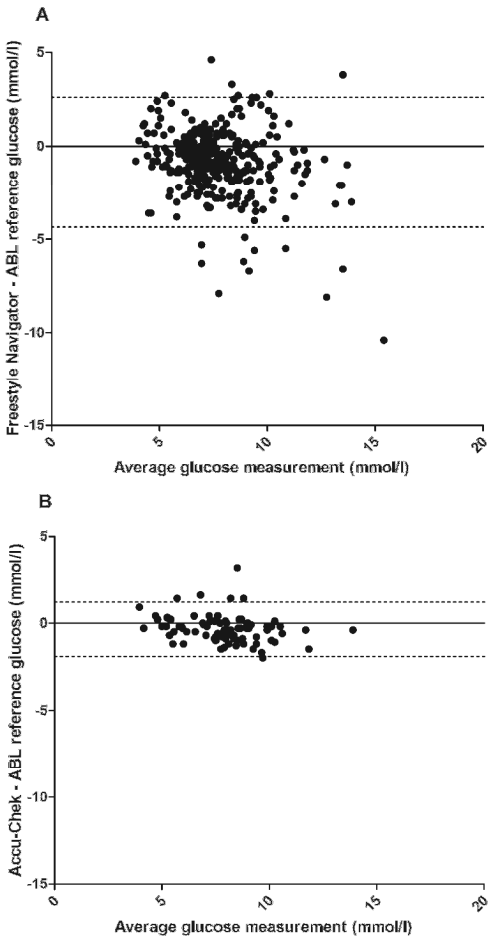
Key messages

- Insulin treatment based on continuous subcutaneous glucose monitoring (CGM) revealed the same number of hypoglycemic events compared to point of care (POC
- Subcutaneous CGM was equally effective as POC measured as glucose time in target range
- Total costs were lower when using subcutaneous CGM than frequent POC
- Nursing workload with glucose regulation was reduced by subcutaneous CGM compared to frequent POC
- A new algorithm designed for continuous measurement should be developed before CGM can be implemented clinically

Acknowledgments

We express our gratitude to the medical, nursing and information technology staff from the OLVG intensive care unit. The authors' own department funded this study. Abbott did not fund this study in any way.

APPENDIX



Bland-Altman plots per glucose monitoring system

(A) CGM system (Freestyle Navigator) (B) Point of care measurement (Accu-Chek). The x-axis represents the average of sensor or device and reference glucose values in mmol/L. The y-axis represents the absolute difference between sensor or device and reference glucose values in mmol/L. The dotted lines represent the 5th and 95th percentile.

REFERENCES

1. Bagshaw SM, Egi M, George C, Bellomo R: Early blood glucose control and mortality in critically ill patients in Australia. Crit Care Med 2009, 37:463–470.
2. Dungan KM, Braithwaite SS, Preiser JC: Stress hyperglycaemia. Lancet 2009, 373:1798–1807.
3. van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, Vlasselaers D, Ferdinande P, Lauwers P, Bouillon R: Intensive insulin therapy in critically ill patients. N Engl J Med 2001, 345:1359–1367.
4. The NICE-SUGAR Study Investigators, Finfer S, Chittock DR, Su SY, Blair D, Foster D, Dhingra V, Bellomo R, Cook D, Dodek P, Henderson WR, Hébert PC, Heritier S, Heyland DK, McArthur C, McDonald E, Mitchell I, Myburgh JA, Norton R, Potter J, Robinson BG, Ronco JJ: Intensive versus conventional glucose control in critically ill patients. N Engl J Med 2009, 360:1283–1297.
5. Marik PE, Preiser JC: Toward understanding tight glycemic control in the ICU: a systematic review and metaanalysis. Chest 2010, 137:544–551.
6. Hermanides J, Bosman RJ, Vriesendorp TM, Dotch R, Rosendaal FR, Zandstra DF, Hoekstra JBL, De Vries JH: Hypoglycemia is associated with intensive care unit mortality. Crit Care Med 2010, 38:1430–1434.
7. The NICE-SUGAR Study Investigators, Finfer S, Liu B, Chittock DR, Norton R, Myburgh JA, McArthur C, Mitchell I, Foster D, Dhingra V, Henderson WR, Ronco JJ, Bellomo R, Cook D, McDonald E, Dodek P, Hébert PC, Heyland DK, Robinson BG: Hypoglycemia and risk of death in critically ill patients. N Engl J Med 2012, 367:1108–1118.
8. Hermanides J, Vriesendorp TM, Bosman RJ, Zandstra DF, Hoekstra JB, DeVries JH: Glucose variability is associated with intensive care unit mortality. Crit Care Med 2010, 38:838–842.
9. Mesotten D, van den Berghe G: Glycemic targets and approaches to management of the patient with critical illness. Curr Diab Rep 2012, 12:101–107.
10. Moghissi ES, Korytkowski MT, DiNardo M, Einhorn D, Hellman R, Hirsch IB, Inzucchi SE, Ismail-Beigi F, Kirkman MS, Umpierrez GE: American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control. Diabetes Care 2009, 32:1119–1131.
11. Westhoff D, Bosman RJ, Oudemans-van Straaten HM, DeVries JH, Wester JPJ, van point-of-care AccuChek® in critically ill patients; a pilot study. Neth J Crit Care 2010, 14:381–387.
12. Brunner R, Kitzberger R, Miehsler W, Herkner H, Ma dL C, Holzinger U: Accuracy and reliability of a subcutaneous continuous glucose monitoring system in critically ill patients. Crit Care Med 2011, 39:659–664.
13. Holzinger U, Warszawska J, Kitzberger R, Herkner H, Metnitz PG, Ma dL C: Impact of shock requiring norepinephrine on the accuracy and reliability of subcutaneous continuous glucose monitoring. Intensive Care Med 2009, 35:1383–1389.
14. Siegelaar SE, Barwari T, Hermanides J, Stooker W, van der Voort PH, DeVries JH: Accuracy and reliability of continuous glucose monitoring in the intensive care unit: a head-to-head comparison of two subcutaneous glucose sensors in cardiac surgery patients. Diabetes Care 2011, 34:e31.
15. Rice MJ, Coursin DB: Continuous measurement of glucose. Facts and challenges. Anesthesiology 2012, 116:199–204.
16. Holzinger U, Warszawska J, Kitzberger R, Wewalka M, Miehsler W, Herkner H, Ma dL C: Real-time continuous glucose monitoring in critically ill patients: a prospective randomized trial. Diabetes Care 2010, 33:467–472.
17. Rood E, Bosman RJ, van der Spoel JI, Taylor P, Zandstra DF: Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation. J Am Med Inform Assoc 2005, 12:172–180.
18. Hakkaart-van Roijen L, Tan SS, Bouwmans CAM: Guidelines for cost calculations, methods and recommended prices for economic evaluations in health care (in Dutch). Amstelveen: College voor Zorgverzekeringen; 2010:2004.
19. Kaukonen KM, Bailey M, Pilcher D, Orford N, Finfer S, Bellomo R: Glycaemic control in Australia and New Zealand before and after the NICE-SUGAR trial: a translational study. Crit Care 2013, 17:R215.

20. Okawa M, Kunimoto F, Kanamoto M, Narahara H, Hinohara H, Tobe M, Yanagisawa A, Saito S: Effect of different blood glucose target levels on the incidence of hypoglycemia during insulin therapy in the intensive care unit. J Diabetes 2013, 5:51–56.
21. Aragon D: Evaluation of nursing work effort and perceptions about blood glucose testing in tight glycemic control. Am J Crit Care 2006, 15:370–377.
22. Gartemann J, Caffrey E, Hadker N, Crean S, Creed GM, Rausch C: Nurse workload in implementing a tight glycaemic control protocol in a UK hospital: a pilot time-in-motion study. Nurs Crit Care 2012, 17:279–284.
23. Malesker MA, Foral PA, McPhillips AC, Christensen KJ, Chang JA, Hilleman DE: An efficiency evaluation of protocols for tight glycemic control in intensive care units. Am J Crit Care 2007, 16:589–598.
24. Leelarathna L, English SW, Thabit H, Caldwell K, Allen JM, Kumareswaran K, Wilinska ME, Nodale M, Haidar A, Evans ML, Burnstein R, Hovorka R: Accuracy of subcutaneous continuous glucose monitoring in critically ill adults: improved sensor performance with enhanced calibrations. Diabetes Technol Ther 2014, 16:97–101.
25. Leelarathna L, English SW, Thabit H, Caldwell K, Allen JM, Kumareswaran K, Wilinska ME, Nodale M, Mangat J, Evans ML, Burnstein R, Hovorka R. Feasibility of fully automated closed-loop glucose control utilizing continuous subcutaneous glucose measurements in critical illness: a randomised controlled trial. Crit Care 2013, 17:R159.

CHAPTER 9
The clinical benefits and accuracy of continuous glucose monitoring systems in critically ill patients – a systematic scoping review

ABSTRACT

Continuous Glucose Monitoring (CGM) systems could improve glycemic control in critically ill patients. We aimed to identify the evidence on the clinical benefits and accuracy of CGM systems in these patients. For this, we performed a systematic search in Ovid ME DLINE, from inception to July 26th, 2016. Outcomes were efficacy, accuracy, safety, workload and costs. Our search retrieved 356 articles, of which 37 were included. Randomized controlled trials on efficacy were scarce (n= 5) and show methodological limitations. CGM with automated insulin infusion improved time in target and mean glucose in one trial and two trials showed a decrease in hypoglycemic episodes and time in hypoglycemia. Thirty-two articles assessed accuracy, which was overall moderate to good, the latter mainly with intravascular devices. Accuracy in critically ill children seemed lower than in adults. Adverse events were rare. One study investigated the effect on workload and cost, and showed a significant reduction in both. In conclusion, studies on the efficacy and accuracy were heterogeneous and difficult to compare. There was no consistent clinical benefit in the small number of studies available. Overall accuracy was moderate to good with some intravascular devices. CGM systems seemed however safe, and might positively affect workload and costs.

INTRODUCTION

Stress-induced hyperglycemia occurs in over 90% of the patients admitted to an Intensive Care Unit (ICU), irrespective of a previous diagnosis of diabetes [1]. This hypermetabolic state is a response to severe illness and results from increased gluconeogenesis, enhanced peripheral insulin resistance and beta-cell secretory defects, due to a complex interaction between excessive counter regulatory hormones and cytokines [2]. This phenomenon was regarded as physiological, although several studies showed an association between hyperglycemia and mortality in critically ill patients [3-6]. Likewise, hypoglycemia and glycemic variability were shown to relate to adverse outcomes [7]. Although the mechanism by which dysglycemia results in adverse clinical outcomes is not fully understood, these findings have highlighted the importance of glucose control. In 2001, van den Berghe and colleagues were able to show a substantial mortality benefit in a surgical ICU when intensive insulin therapy (IIT) was used to treat hyperglycemia to target a glucose between 80 and 110 mg/dL [1]. Several trials on IIT have been conducted since then, but the initial beneficial effects could not be confirmed [8-13]. Also, there was an increased risk for hypoglycemia associated with IIT in these trials, with 5.1% to 18.7% of patients experiencing one or more episodes of severe hypoglycemia [1, 8, 12], which is accompanied by a ± 2 fold increased mortality risk [14, 15]. Due to differences in study populations, pursued target ranges and measurement devices, it is difficult to compare results and draw definite conclusions. Nevertheless, recent meta-analyses did not show a beneficial effect of tight glycemic control on mortality [16-18]. Nowadays, consensus states that hyperglycemia in critically ill patients should be monitored and treated, with guidelines recommending glucose levels between 100 and 150 mg/dL (Society of Critical Care Medicine [19]) or between 140 and 180 mg/dL (American Diabetes Association [20]). However, tighter ranges might be feasible when the average ICU has the ability to safely control glucoses in such a range [21].

Currently, glucose control in the ICU is mostly based on intermittent measurements with handheld meters for point-of-care glucose testing. These periodic measurements are used to guide intravenous insulin administration based on a (local) algorithm. Handheld glucose meters are not designed for ICU use, and their accuracy is questionable and marke dLy inferior to central laboratory or blood gas analysis, especially in patients with anemia, hypoxia or when exposed to certain drugs [22, 23]. Moreover, the intermittent character makes it impossible to observe important glucose fluctuations. In critically ill patients, between 4% and 15% of hypoglycemic events are undetected [24] and hypoglycemic episodes occur more frequently when there is a longer time interval between glucose measurements [25]. Continuous Glucose Monitoring (CGM) systems provide (near-) continuous information about glucose levels, thereby creating the possibility to detect acute changes and real-time trend data, and improve the quality and efficiency of glucose control. Moreover, they could decrease the time spent on achieving glycemic control, since tight glycemic control is burdensome and reported to take almost 1.5 to 2 hours of a 24 hours single patient nursing period [26, 27]. Considering the impact of frequent blood glucose monitoring in the critically ill, continuous systems could be

advantageous by improving glycemic control and reduce the burden on (nursing) staff. There are different CGM devices available, employing various measurement techniques (glucose oxidase, mid-infrared spectroscopy or fluorescence) with positioning in either the interstitial (minimal invasive) or intravascular (invasive) space [28]. One noninvasive transdermal device (Symphony, Echo Therapeutics), is claimed to be under investigation. Devices can be labeled as 'continuous' when they have a measurement or sampling frequency of at least once every 15 minutes or more frequently [29]. The use of subcutaneous systems is already accepted in the outpatient setting, where a high accuracy is shown as compared to reference blood glucose measurement [30]. CGM systems have been evaluated over almost 10 years in the ICU. Although the use of a CGM system, especially when combined with an appropriate insulin dosing protocol, has the potential to improve efficient and safe glucose control, it is still not common practice in the ICU. Moreover, accepted standardization of metrics to evaluate the benefit of CGM systems was lacking, until an ICU expert consensus statement came out [31]. With this systematic scoping review, we aimed to assess the evidence regarding the clinical benefits and accuracy of CGM systems in critically ill patients.

METHODS

Eligibility criteria

For this review, we included articles that reported original empirical data on the use of a CGM system (located subcutaneous or intravascular) in critically ill patients, admitted to an ICU. Outcome measures of interest were efficacy, accuracy, adverse events, workload and costs. To estimate efficacy we included only randomized controlled trials (RCTs). Assessment of efficacy in these trials had to at least cover one metric of the average glucose level or time spent in different glucose ranges. The CGM system had to be compared to standard of care or head-to-head with another CGM system. Articles were excluded when they did not, or not explicitly, report the outcomes of interest, when the CGM system had an sampling interval over 15 minutes (since this is considered as non-continuous [29]), or when the patients studied were considered to be not representative for the general ICU population (e.g. highly specific patient groups, such as (premature) neonates and patients undergoing pancreatic surgery). Studies with the STG closed-loop system by Nikkiso were excluded since this system is not a CGM system, but an artificial pancreas system that is only available in Japan [32]. To assess the accuracy, we included both RCTs and observational studies. Accuracy studies were considered eligible when the investigational CGM system output was compared to an arterial or venous reference sample (since capillary reference measurements are considered inaccurate) [22]. To be eligible, at least two of the following accuracy outcome measures had to be reported (for overview see appendix B): Mean (or median) Absolute Relative Difference (MARD), Clarke Error Grid (CEG [33]), (modified) Bland-Altman plot [34] or agreement with the accuracy standards of the International Organization for Standardization (ISO) [35, 36].

Search Methods

A medical information specialist [JL] performed an electronic search in Ovid ME DLINE (including Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)) from inception to July 26th, 2016. The search strategy consisted of MesH terms and text words for the concepts CGM (including continuous or real-time glucose, glucose sensor, glucose monitor and specific CGM devices) and critically ill patients (including (pediatric) intensive care, burn center, coronary care). No methodological search filter and language or date restrictions were applied. Animal studies were safely excluded by double negation (not (animals/not humans/)) (see appendix A for entire ME DLINE search strategy). Additionally, we applied forward and backward snowballing of identified relevant articles.

Study selection

Two independent authors [SCJS and SR] screened title and/or abstract of the articles against the in- and exclusion criteria and included appropriate articles based on their full text. In case of doubt or disagreement about the inclusion, a third author [JHD] was involved, and consensus was reached by discussion.

Data extraction and handling

Data extraction was independently performed by two authors [SCJS and SR] based on a predefined form. For each eligible article, the following data items were extracted: publication year, study design, type and number of included patients, in- and exclusion criteria, intervention and control, reference method, baseline characteristics, local glucose protocol and target ranges, outcome measures and authors conclusions. Discrepancies were resolved through consensus. We described the main results of the study qualitatively and in evidence tables.

RESULTS

Results of the search

The MEDLINE search retrieved 356 unique citations. Snowballing did not yield additional publications. We excluded 274 citations on the basis of title and/or abstract and screened 82 articles full text. Of these, 37 articles met our in- and exclusion criteria and were included in this review. The study selection process and reasons for exclusion are presented as a flow diagram (Figure 1).

Study descriptives

Table 1 provides an overview of the 37 articles included in this review, sorted by their main outcome (efficacy versus accuracy). Five RCTs primarily assessed the efficacy of subcutaneous CGM systems [37-41]. We found no RCTs on intravascular devices. Accuracy was the main outcome of the other 32 articles, with a minority focusing on intravascular devices. Almost half

of the articles studied a mixed ICU population, but 9 (28%) of 32 accuracy studies were restricted to cardiac surgery patients. Four observational studies concerned children [42-45]. CGM systems were studied for a median of 72 hours, with a maximum of 7 days. Overall, the number of analyzed patients varied between 8 and 174, with a median of 24 patients. Most studies used arterial or both arterial and venous samples as reference; four studies used only venous reference samples.

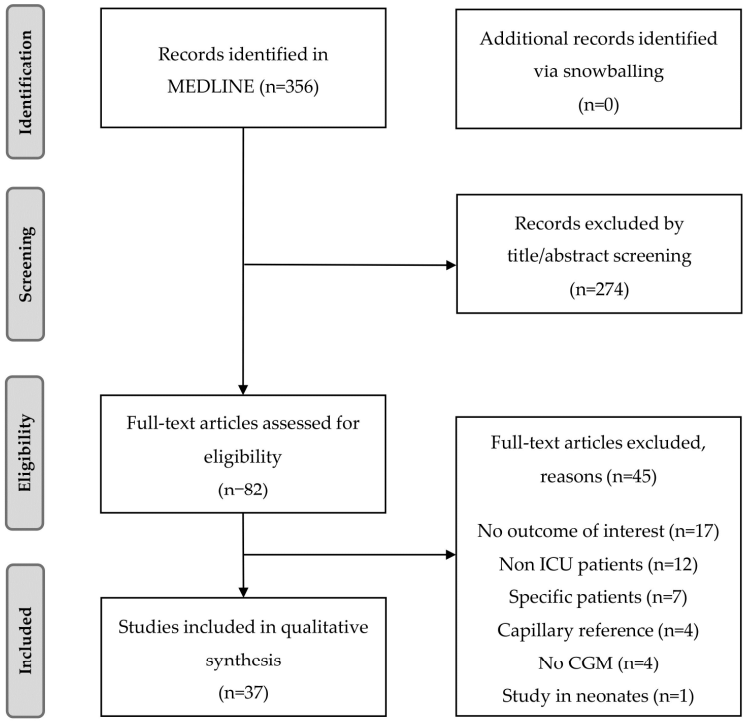


Figure 1. Flow diagram of study selection

Efficacy

We identified 5 RCTs, conducted between 2010 and 2015, that investigated the efficacy of subcutaneous CGM systems on glycemic control (Table 2). One trial studied a microdialysis based system (GlucoDay, A. Menarini Diagnostics), the other devices were based on an electrochemical electrode technique (FreeStyle Navigator, Abbott and Guardian REAL-Time, Medtronic). The sample size ranged from 24 to 156 patients. Study duration varied between 24 hours and 5 days. One study included only cardiac surgery patients [40], the other trials included ICU patients with various medical or surgical conditions. All trials included both patients with and without preexisting diabetes (percentage diabetes patients 25-40%). The CGM systems used different approaches to guide treatment. De Block and colleagues used

high rates of glucose change (>25 mg/dL per 30 minutes) to prompt the nursing staff to take extra arterial reference samples in a group of patients with a high APACHE score (mean 28 ± 7) [37]. In the control patients (n= 19) CGM readings were blinded. These reference samples were used to adjust intravenous insulin dose based on a modified Yale protocol. The CGM device in the intervention group (n= 16) was calibrated 6 times in 48 hours as compared to 2 times in 48 hours in the control group (the latter following manufacturer's instructions). The use of this CGM system did not improve mean glucose (intervention 119 ± 17 mg/dL versus control 122 ± 11 mg/dL, not significant (NS) (actual *p*-value not reported), time in target (intervention 37 ± 12% versus control 34 ± 10%, NS), or glycemic variability (NS). Although not significant, the time in hypoglycemia (intervention 9 ± 23 minutes per 24 hours versus control 35 ± 62 minutes per 24 hours, NS) as well as the number of patients with hypoglycemic events (intervention 3 versus control 9, NS) was considerably lower in the intervention group. Likewise, adjustment of the insulin infusion rate by regularly inserting CGM readings into an insulin advising algorithm (every 15 minutes by Kopecky [40], and every 2 hours by Holzinger [41]), did not improve mean glucose or time in range, defined as 80 - 110 mg/dL by Kopecky and <110 mg/dL by Holzinger. Both studies did show lower (severe) hypoglycemia rates in the CGM intervention group, in the latter with a hypoglycemia rate of 1.6% in the intervention and 11.5% in the control group, giving a 9.9% absolute risk reduction (95% CI 1.2 - 18.6, *p*= 0.031). Boom et al. used the alarms of the CGM system, set at glucose <90 mg/ dL or >162 mg/dL, to enter additional glucose values into an insulin advising algorithm (designed for intermittent measurements) in two groups of 78 patients [38]. As compared to usual care with blinded CGM this did not increase time in target (intervention 69 ± 26% versus control 66 ± 26%, *p*= 0.47). When combining the CGM system with automated closed-loop insulin therapy in mainly neurosurgical patients, as done by Leelarathna et al., mean glucose was significantly reduced and time in target improved with over 35% (intervention 54.3% [44.1 - 72.8] versus control 18.5% [0.1-39.9], *p*= 0.001), especially within the first 24 hours, and this effect was persistent when ranges were widened [39]. There were no hypoglycemic events (<72 mg/dL) and there was no significant between group difference in the amount of insulin administered. However, in the intervention and control group different target ranges were used (intervention 108 - 144 mg/ dL versus control 126 - 180 mg/ dL) and the intervention group had a higher frequency of calibration. Thus, with this closed-loop system a lower and smaller target glucose range could be achieved, without inducing the risk for hypoglycemia. We did not perform a meta-analysis on these trials given the (clinical) heterogeneity introduced by different devices, study populations and glucose targets.

Accuracy

The accuracy of CGM systems was assessed in all included articles, with 26 articles investigating subcutaneous systems (Table 3) and 11 investigating intravascular systems (Table 4). The relevant accuracy metrics are explained in appendix B. Four studies used a blood gas

analyzer for reference measurements, but did not specify the sample location [43, 45-47]. In all other studies the reference was specified as arterial and/or venous.

MARD

The Mean Absolute Relative Difference (MARD) quantifies the deviation from the reference measurement. A lower MARD corresponds with better accuracy. MARD varied widely among studies. With regard to the subcutaneous devices, the highest reported MARDs were 30.5% with the FreeStyle Libre (Abbott Diabetes) [48], and 23.2 - 23.7% with the Guardian REAL-Time, (Medtronic) depending on location (thigh or abdomen) [49]. Excluding these outliers, MARD ranged from 7% to 15.6% (both with the FreeStyle Navigator [50]). The needle-free, transdermal device, the Symphony, showed a MARD of 12.3% [51]. Intravascular devices showed overall lower MARD values, ranging between 5.1% (GlucoClear, Edwards Lifesciences [52]) and 14.2% (GluCath, GluMetrics [53]). When comparing arterial (63 sensors) with venous (9 sensors) positioning, as done by Strasma et al., arterial resulted in lower MARD values (arterial 9.6% versus venous 14.2%) [53]. It has to be noticed that intravascular devices can be placed either in the central venous or peripheral (venous or arterial) circulation. Peripherally placed devices suffer more from hypothermia, movements and vasospasm, which can impair accuracy of the device.

ISO

The ISO guideline described the accuracy requirements for intermittent self-monitoring devices in order to achieve regulatory approval but is also used to assess CGM accuracy. Most studies used the ISO 2003 criteria, accepting a 20% bias from the reference measurement for glucose levels >75 mg/dL (and 15 mg/dL bias when glucose level <75 mg/ dL) [35]. The stricter guideline from 2013 accepts a 15% bias when glucose levels are >100 mg/dL, and 15 mg/dL when glucose is <100 mg/dL [36]. The proportion of subcutaneous CGM readings within 20% of the reference value was between 68.1% and 94.0%, with exception of the FreeStyle Libre [48], which showed only 7.0% of the CGM readings within 20% of the reference. Intravascular devices had overall higher ISO agreement, with even up to 100% with the Eirus system (Maquet Critical Care) [54]. Stricter than the ISO requirement, the ICU expert consensus states that it is desirable to have 98% of readings within 12.5% of the reference standard [31]. This was reported in two studies [55, 56], and with 60.3% and 58.0% within the 12.5% zone these devices did not meet this criterion.

Clarke Error Grid (CEG

CEG analysis indicates the clinical accuracy of the CGM system by connecting the imprecision of the device to the therapy implications [33]. A CEG analysis was performed in 29 (78%) out of 37 studies, and all used grid glucose target between 70 mg/dL and 180 mg/dL, as originally described by Clarke. Two subcutaneous devices had 100% of the paired samples in the acceptable zones A and B (DGMS, San Meditech [57] and CGMS System Gold, Medtronic [58]). In these studies only a few points were in the hypoglycemic range (exact numbers not

reported). All other studies had a minority of samples in possibly dangerous zones. All but one of the studies that reported a CEG of intravascular devices showed 100% to be in zone A and B [46, 52, 54, 59, 60].

Bland-Altman

Bland-Altman plots show the mean bias and limits of agreement (1.96 x standard deviation) between CGM reading and reference measurement, indicating the systematic and random errors. Both subcutaneous and intravascular devices showed generally low mean bias, but there were outliers, with bias in intravenous devices ranging from -10.8 mg/dL to 4.1 mg/dL, and in subcutaneous devices from -43.2 mg/dL to 14.9 mg/dL. Limits of agreement were overall high, as can be seen from table 3 and 4.

Intravascular versus subcutaneous

One study made a head-to-head comparison between the Eirus intravascular CGM system and the subcutaneous FreeStyle Libre, showing the latter to be by far inferior in terms of accuracy (MARD of 30.5% versus 6.5%) [48]. Another study, comparing the intravascular GluCath with the subcutaneous FreeStyle Navigator in 8 patients reported similar accuracy between the two devices in terms of MARD, ISO agreement and Bland-Altman analysis [61].

Effect of calibration

Leelarathna et al. investigated the effect of calibration frequency on accuracy outcomes in a subcutaneous device (FreeStyle Navigator) [50]. The intervention group calibrated at variable intervals of 1 to 6 hours, on average 9.5 times in the first 24 hours, and 7 times in the second 24 hours. They were compared to a calibration frequency of 4 times in the first 24 hours and no calibrations in the second 24 hours (following manufacturer's instructions). Enhanced calibration resulted in a significant lower MARD (intervention 7.0% [3.5 - 13.0] versus control 12.8% [6.3 - 21.8], p <0.001), more points in zone A of the CEG (intervention 87.8% versus control 70.2%) and higher agreement with the ISO criteria (70.2% versus 87.8%). Three other studies investigated the effect of calibration frequency on accuracy. Van Hooijdonk et al. used routinely obtained blood glucose measurement as additional calibrations, resulting in a mean increase of 6 times in contrast to the requested calibrations [56]. They showed that the number of calibrations had a positive effect on accuracy. With each additional calibration, the absolute difference between CGM reading and reference decreased with 1.4%. In the study by Yue and colleagues, there was significant improvement in MARD when comparing calibration within 6 hours with calibration between 6 and 12 hours (8.8 ± 7.2% versus 20.1 ± 13.5%, p <0.0001) [57]. The same was seen for data points in zone A of the CEG (92.4% versus 57.1%, p <0.0001). De Block reported that the data points in zone A and B increased from 95% to 97%, when calibration frequency of the GlucoDay went from 2 to 6 times per day, with fewer points in zone C (4.5% versus 1.6%) [62].

Factors influencing accuracy

Five studies tried to identify factors that could influence accuracy. Accuracy was not influenced by use of norepinephrine [63], or dependent on reason of admission (medical versus surgical) [64]. In the study by Kosiborod et al., MARD was equal between patients with a high and low cardiac surgery risk score [65]. It did however deteriorate in diabetes patients [56, 66] and with use of vasopressors, higher SOFA scores, glycemic variability and in the hyperglycemic range [55]. Septic status seemed to improve accuracy [64]. There was no evident difference between various sensor locations (abdomen, thigh, shoulder) [49]. Microcirculation, measured using a microvascular flow index, had no effect on accuracy [66]. None of the studies discussed the use of acetaminophens, although its influence on CGM accuracy is well known from studies in the outpatient setting [67].

Safety

In 14 (40%) of 35 studies there was no description on the occurrence of adverse events. In the studies that did report adverse events, no serious adverse events occurred. Complications of the subcutaneous device were minor bleeding after insertion in 4 (20%) of 20 patients [55], bruises in 13 (13%) of 102 sensors and redness in 13 (13%) of 102 sensors [56]. One study reported a thrombus rate (on ultrasound) of 21% in arterial devices and even of 66% in venous devices, with two complete venous occlusions that required treatment, one of which turned out to be device related thrombosis [53]. Macken et al. and Crane et al. reported both 2 cases of thrombus formation on ultrasound with two intra-arterial catheters (GluCath, Glumetrics and GlySure, GlySure), but no treatment was required [60, 68].

Workload and costs

Boom et al. were the only one to investigate the effect of a CGM, the FreeStyle Navigator I, on costs in a 24 hours timeframe [38]. Their analysis showed a mean 12-euro benefit in favor of the CGM system (95% CI 5 to 22 euro, p= 0.02). This profit came mainly from the reduction in nursing time and the decline in cost due to laboratory and POC measurements. They also calculated the effect on nursing workload and showed a mean reduction of 19 minutes per 24 hours (intervention 17 minutes, control 36 minutes, p= 0.001). Wollersheim and colleagues used questionnaires to assess the nurse feedback [55]. During the study they were assigned with observing the trend line and performing additional BG measurements, and not with inserting the sensor or performing calibrations. With a 1/3 response rate, almost 80% of the nurses rated the subcutaneous Sentrino CGM system (Medtronic) as not beneficial and more than half of the nurses described disadvantages of the system, mainly inadequate alarms performance (mentioned in 23.3% of the replies). This is in accordance with findings by Kosiborod et al, who showed high false alarm rates with the Sentrino for both hypoglycemia (70.2%) and hyperglycemia (53.5%) [65]. In contrast, the latter reported better nurse acceptance with the same device, with a 100% positive opinion on performance after using it in two patients.

Children

Accuracy in critically ill children was assessed in 4 observational studies [42-45]. Overall, accuracy seemed lower in children than in adults. MARD ranged from 15.3% [43] to 23.0% [44]. Like in the above studies, most data points were in zone A and B of the Clarke error grid, with exception of the trial by Prabhudesai et al., which had 7.2% of data in zone D [45]. This study also showed high mean bias in the Bland-Altman plot. There were no adverse events reported.

DISCUSSION

This systematic scoping review evaluated 37 articles that included both RCTs and observational studies. The majority of the studies were single center studies with modest sample sizes. Study duration was in general short, with an average of 72 hours. In addition, the number of RCTs was small (n= 6). Although the number of studies increased over the years, the heterogeneity in study populations, interventions and reported outcomes impeded us to draw general conclusions. Moreover, the fact that studies were performed in settings with different local standards of care probably had an important effect on outcomes.

Overall, in terms of efficacy, the use of subcutaneous CGM systems does not seem to improve the glycemic control of critically ill patients convincingly in a clinically significant manner. With regard to this conclusion, it has to be taken into account that RCTs were scarce (n= 5) and that the included trials showed methodological shortcomings to a greater or lesser extent. Even for the two largest RCTs it could be argued that they lacked the appropriate sample size to conclude on their secondary glycemic outcomes. In the trials by Boom and Holzinger the difference in mean sensor glucose was on the border of statistical significance favoring the group receiving a CGM system, with *p*-values of respectively 0.07 and 0.076. Although the 95% confidence intervals should have been reported, the differences in mean glucose of 0.3 - 0.4 mmol/L were small, with unclear clinical significance.

Moreover, the RCTs used the readings of the CGM system in different manners. Two studies used the CGM as a prompt to obtain additional arterial samples [37, 38]. This approach will detect important changes between intermittent measurements, but does not fully use the advantages of the continuous measurements. In the study by de Block, the nurses had to notice high rates of change, without help from an alarm function, since the GlucoDay is not equipped with an alarm function. The *p*-values in this study were denoted as non-significant, but the actual p-values were not reported. The trial by Kopecky suffers from the same methodological flaw. Two other studies assigned their nurses to regularly use CGM reading as additional input to the local algorithm [40, 41]. This might have led to bias as alarm functions were not used and the input was therefore dependent on nurse adherence to the protocol. In addition, this compliance to the protocol was not measured or reported in these trials. Moreover, this approach might even increase workload, instead of reducing it. The only study that evidently showed improved glycemic control used a fully automated closed-loop system and an adapted glucose algorithm [39]. In the included trials, hypoglycemic events occurred in

11.5% of patients [41] to 2.4% of time [37]. In three of the five RCTs there was an overall reduction in the number of hypoglycemic episodes, as well as the time in hypoglycemia in the CGM group [37, 40, 41], but this did not reach statistical significance. Hypoglycemia rate is highly dependent on local glucose targets, which varied in the different trials. When the treatment target range is increased, which was done after publication of the NICE-SUGAR results, the hypoglycemia rate diminishes substantially [69]. Comparability of results would benefit from greater consistency in reporting and consensus on outcome measures, as for example stated in a consensus report for artificial pancreas studies in outpatients [70]. To show an improvement in glycemic control with CGM systems is difficult, since most ICUs already successfully maintain adequate glucose control, with adequate mean glucose and time in range, and low rates of hypoglycemia with the current standard of care. To take full advantage of CGM devices, the glucose control algorithm probably needs to be adapted to the continuous measurements, which was usually not done in the included studies, except for the closed-loop system. Thus, the number of RCTs on this topic, as well as the sample sizes, are relatively small, which makes it difficult to draw firm conclusions. There are no RCTs conducted that investigated intravascular devices, so we are not able to conclude on the clinical benefits of intravascular systems in the ICU.

Accuracy is based on comparison between sensor and simultaneously obtained reference values. There are multiple assessment methods, but there is no consensus yet for determining and reporting CGM accuracy, as can be seen from the different methods in which accuracy was assessed and reported. The current expert consensus recommendations on reporting state that a MARD <14% is acceptable and that 98% of the readings >100 mg/dL should be within 12.5% and the remaining 2% within 20% of the reference measurement ([31], appendix B). The consensus statement does not differentiate between pooled MARD of all data points and individual MARDs. By requiring a pooled MARD to be below a certain cut-off, patients with substantially higher MARDs may go unnoticed. In our opinion, it would be advisable to also set a requirement for the dispersion around the average MARD. The third consensus recommendation is that all data pairs should ideally be in zone A of the CEG. None of the included studies meets all these recommendations. Moreover, not all articles describe the information required, such as nature of reference blood sampling and measurement technology. Some studies used reference measurements from multiple sources, which can be seen as a weakness in design. To assess accuracy, time-matched sensor-reference pairs are used by all included studies. These time-matched points are statistically matched, but might be sub optimally matched from a physiological point of view, because of the difference in glucose concentration between compartments depending upon the glucose rate of change (time lag), which is especially a problem with subcutaneous devices, and the unknown contribution of physiological and device-related delay [71].
Difference between CGM reading and reference measurement, as expressed by MARD, was in some subcutaneous devices as high as 30%, i.e. in the FreeStyle Libre [48]. This device, however, is developed for outpatient use. It applies to the definition of CGM since it measures

interstitial glucose every 15 minutes, but it has no alarm functions, and lacks the possibility to manually calibrate (since it is factory-calibrated). The MARD of the intravascular devices was overall lower than of subcutaneous devices. This might in part be due to the systematic error that is introduced since calibration of interstitial positioned devices is performed with blood glucose levels. Moreover, the MARD depends highly on the reference method, which has its own error, as well as the possible delay between actual blood samples and reported time of analysis. Meanwhile, even when MARD is higher than recommended, the provided trend information could still be beneficial in terms of early detection and treatment of hypo- and hyperglycemia. Not all studies gave a clear description of reference sample location and measurement device. It is known that the MARD is considerably lower in hypoglycemic ranges. However, included studies all reported a low number of hypoglycemic events, due to adequate glucose algorithms. Thereby we cannot conclude on accuracy in the hypoglycemic area.

Bland-Altman plots show the difference between the CGM reading and the reference, plotted against their mean [34]. Overall, the mean bias was low, indicating a low systematic error, but there are wide limits of agreement, indicating high random errors. These random errors could be due to both the CGM device as patient specific factors.

The ISO criteria are originally used to determine whether a device is accurate enough to be marketed commercially for outpatient use. The 2003 criteria state that 95% of the readings should be within 20% of the reference measurement. In 2013 the accuracy limits were revisited to at least 95% of the reading within 15% of the reference. Only some intravascular devices met this criteria [52, 54, 68, 72]. By requiring 95% of measured values to meet this criterion, still 5% of the measurements can differ from the reference by any amount. This might be acceptable in home-use, for which these meters were designed, but seems potentially dangerous in ICU setting.

The majority of the included studies used the CEG to assess the clinical accuracy of the CGM system. CEG categorizes pairs in terms of the consequence of treatment decisions. It was initially designed for evaluation of self-monitoring devices, for which 95% of the pairs should fall in zone A [33]. Rate and direction, two important features are not taken into account in the original CEG. For this, modified (continuous) error grids have been developed, but their value in the assessment is questionable [71]. In most studies the majority of data pairs were in the acceptable zones A and B, especially with the intravascular devices. However, there were quite some studies reporting a certain percentage of pairs in the dangerous D and E zones. The targets used in the original CEG are 70 to 180 mg/dL, but, as shown in the table, most ICUs use different targets, but do not adjust this in their CEG.

Therefore, overall the intravascular devices show better accuracy than the subcutaneous devices, possibly because the former are not sensitive for disturbances in microcirculation and they have no lag time to consider. Most of the subcutaneous devices were originally designed for home-use and thereby not equipped to deal with critical ill patients. However, 80% of the articles studying intravascular devices included only elective (cardio surgical) patients. Considering that these patients are not as ill as the general ICU population, external validity is

limited. In addition, most subcutaneous devices were studied in the general ICU population. Subcutaneous devices are less invasive, with a lower risk for disturbance by infused medication and glucose solutions. The Sentrino CGM system is the only subcutaneous device primarily designed for use in critically ill patients. It was studied in three included studies, of which in two it did not perform with satisfactory accuracy [55, 56]. It is not clear what the effect of rapid glucose changes, severity of illness and interference of medication (e.g. vasopressin, acetaminophen) is on the accuracy of both types of CGM devices, since this was not directly investigated. Not surprisingly, the accuracy of subcutaneous systems markedly improves when performing more frequent calibration than advised by the manufacturer for outpatient use [37, 50, 56, 57]. In the study by Leelarathna et al., this meant one calibration per 2.5 to 3.5 hours. Thus, some of one of the advantage of continuous glucose monitoring, reducing the amount of required blood samples, and reduced time spent on blood glucose control, is partly lost when calibrating more frequently. It is recommended that CGM systems should ideally not need more than 3 calibrations every 24 hours at the ICU [31]. As long as a CGM system needs calibration with blood glucose measurements, it will be difficult to truly replace this. Optiscanner is the only (intravascular) device that does not need calibration.

Adverse events of subcutaneous devices seemed rare, and consisted in the worst case of minor bleeding [55], bruises or redness [56]. With the intravascular devices, thrombosis was described, especially with the intravenous devices [53], leading to one case of device related thrombosis. Overall, the adverse event rate in all studies was low, and no serious adverse events were reported. Thereby, adverse effects will not limit the use of CGM systems in the intensive care unit.

Only one study by Boom et al. investigated the effect of a subcutaneous CGM system on workload and costs, and showed that using these systems significantly improved both these outcomes [38]. However, the timeframe was relatively short (24 hours), and the cost benefit was small, so this requires further investigation. It did show that the use of CGM systems, which are quite expensive, is not an a priori expense. CGM did improve nursing workload. The advantage of using a continuous monitor is mainly determined by the reduced number of point-of-care measurements, which are time intensive. However, CGM systems do require regular calibrations to achieve a certain accuracy, which carry their own workload. Theoretically, a totally automated closed-loop system that does not need to be calibrated will be able to bring the workload of glucose control to a minimum.

In general, CGM has potential benefits in the ICU, such as improvement of time in target, reduction of glycemic variability and less staff workload. However, the current evidence on the use of CGM systems in critically ill patients is not sufficient. Large randomized trials have not yet been performed, especially not with use of an adapted glucose algorithm for continuous data. Thus, we lack sufficient data to draw conclusions on the clinical benefit, although the available evidence seems to point in a slightly favorable direction, mainly because of less

hypoglycemic events and especially when combined with an adequate glucose algorithm. CGM accuracy seems moderate to good with intravascular devices, but the majority of studies was limited to cardiac surgery patients. The accuracy of both subcutaneous and intravascular devices might be adequate enough to guide alarms, but when CGM is used to guide therapy, it might require improvement. Accuracy metrics lack standardization and do not seem tailored for the assessment of CGM in critically ill patients, despite the recently made consensus statement. We emphasize the importance of standardization of the assessment methods and future research on accuracy and efficacy in these patients. Ultimately, this may lead the development of a true closed-loop glucose control system.

CONCLUSION

There is sparse evidence for the effect of CGM systems in critically ill patients compared to standard blood glucose measurement, with only five RCTs assessing the impact of subcutaneous devices on glycemic control. Overall, CGM systems do not seem to clearly improve glycemic control in a clinically significant manner. Only when incorporated into a fully automated closed loop the mean glucose and time in the target range did improve in a single trial. In two trials hypoglycemia decreased with the use of CGM, and one trial showed decreased nursing workload. The accuracy of intravascular devices seems better than subcutaneous devices, at the cost of some risk for adverse events like thrombus formation. Intravascular devices are however assessed in a relative small number of studies, and mainly in cardiac surgery patients. The reported accuracy metrics of both subcutaneous and intravascular devices differ widely among studies and a clear definition of assessment methods is limited to an expert consensus statement. Safety in terms of local complications is good, but there is a potential danger as a consequence of inaccurate measurements, making improvements desirable. However, theoretically CGM systems still have the potential to improve glycemic control, especially when technically improved or combined with an appropriate glucose algorithm adapted to continuous measurements. More robust data is needed, preferably from larger multi-center studies with head-to-head comparisons, to demonstrate beneficial effect on both outcome (glycemic control, length of stay, mortality) and costs.

Author Contributions

S.C.J.v.S. and S.R. screened and selected the articles, retrieved relevant information and interpreted the data. J.L. performed the systematic electronic search. S.C.J.v.S. and J.H.D. wrote the manuscript.J.H. and P.H.J.v.d.V. reviewed the manuscript for important intellectual content.

Table 1. Descriptives of the included articles by their main outcome (n= 37)

Main outcome	Accuracy (n= 32)	Effectivity (n =5)
Year of publication (range)	2006-2016	2010-2015
Study design (n, %)		
RCTs	1 (3.1%)	5 (100%)
Observational trial	30 (93.8%)	0
Pooled analysis of two RCTs	1 (3.1%)	0
Type of patients (n,%)		
Mixed ICU patients	12 (37.5%)	3 (60.0%)
Medical patients	0	0
General surgical patients	5 (15.6%)	0
Cardiac surgery patients	9 (28.1%)	1 (20.0%)
Neurosurgical patients	2 (6.3%)	1 (20.0%)
Children	4 (12.5%)	0
Maximum study duration (hours)		
Median [IQR]	72 [48 - 72]	72 [36 - 108]
Range	24 - 168	24 - 120
Not reported	4 (12.5%)	0
Number of analyzed patients		
Median [IQR]	23 [19 - 48]	35 [24 - 140]
Range	8 - 174	24 - 156
Type of CGM device studied (n, %)		
Subcutaneous	19 (59.4%)	5 (100%)
Intravascular	10 (31.3%)	0
Transdermal	1 (3.1%)	0
Subcutaneous and intravascular	2 (6.3%)	0
Reference measurement (n, %)		
Arterial	21 (65.6%)	5 (100%)
Venous	4 (12.5%)	0
Arterial and venous	4 (12.5%)	0
Not described	3 (9.4%)	0
Number of paired samples (n, %)		
Median [IQR]	672 [346 - 1028]	440 [277 - 603]
Range	34 - 2045	277 - 635
Not reported	2 (6.3%)	2 (40%)

Percentages are based on the total amount of 37 articles. Due to rounding percentages might not sum up to 100%.
Abbreviations: ICU, Intensive Care Unit; IQR, interquartile range; n, number; RCT, randomized controlled trial

Table 2. Reported efficacy outcomes of randomized controlled trials assessing subcutaneous CGM systems (n= 5)

Author *Year*	CGM system	N[a]	Study population		Average glucose[b] (mg/dL)	Time in range[c] (%)	Time in hypoglycemia[c] (%)		Time in hyperglycemia[c] (%)		Target glucose range[d] (mg/dL)
De Block et al. *2015*	GlucoDay, A. Menarini Diagnostics	35	Mixed ICU population	Intervention	119.0 ± 17.0	37.0 ± 12.0	0.6	± 1.6	4.0	± 5.0	
				Control	122.0 ± 11.0	34.0 ± 10.0	2.4	± 4.3	2.0	± 3.0	80 - 120
Boom et al. *2014*	FreeStyle Navigator I, Abbott	156	Mixed ICU population	Intervention	127.9 ± 19.8	75.0 ± 18.0		3 episodes in 3 patients	3.0	± 7.0	
				Control	135.1 ± 23.4	71.0 ± 20.0		4 episodes in 4 patients	4.0	± 9.0	90 - 160
Leelarathna et al. *2013*	FreeStyle Navigator I, Abbott	24	Neurosurgical patients	Intervention	142.3 [133.3 - 147.7]	54.3* [44.1-72.8]	0.0			1 episode in 1 patient	
				Control	164.0 [149.6 - 234.2]	18.5* [0.1 – 39.9]	0.0			11 episodes in 5 patients	110 - 140
Kopecky et al. *2013*	Guardian REAL-Time, Medtronic	24	Cardiosurgical patients	Intervention	111.7 ± 1.8	46.3 ± 5.5		0 episodes			80 - 110
				Control	109.9 ± 10.8	46.2 ± 6.5		2 episodes			
Holzinger et al. *2010*	Guardian REAL-Time, Medtronic	124	Mixed ICU patients	Intervention	105.8 ± 18.1	59.0 ± 20.4		1.6% of patients*			
				Control	110.6 ± 10.4	55.0 ± 18.0		11.5% of patients*			80 - 110

Values are displayed as mean ± SD or median [IQR]. [a] Number of analyzed patients. [b] Average glucose levels are based on the sensor measurement of the CGM system. [c] The time in different ranges is dependent of the predefined ranges of the different studies and could thereby differ among studies. [d] Ranges are when necessary converted into mg/dL, and rounded to dozens. *Indicates a statistically significant difference between the intervention and control group on the reported outcome. Abbreviations: CGM, continuous glucose monitoring; N, number (of patients)

Table 3. Reported accuracy outcomes of the included studies that assessed subcutaneous continuous glucose monitoring (CGM) systems (n= 26)

Author *Year* ADULTS	CGM system	N[a]		MARD[b] (%)	ISO[c] (%)	Clarke error grid[d] (% in zones A-E) A	B	C	D	E	Bland-Altman[e] (mg/dL)	Target range ICU[f] (mg/dL)	Reference sample
Wollersheim et al. *2016*	Sentrino, Medtronic	532 (20)		15.3 (13.5-17.0)	76.9	76.9	21.6	0.2	0.9	0.4	0.5 (-63.5 - 64.6)	80 - 149	Arterial *or* venous
Schierenbeck et al. *2016*	FreeStyle Libre, Abbott***	578 (26)		30.5 ± 12.4	7.0	18.9	80.2	0.9			-43.2 (-82 - -4.5)	90 - 180	Arterial
v. Hooijdonk et al. *2015*	Unspecified, Medtronic	929 (50)		14.8	75.8	75.3	23.5	0.3	0.9	0.0	-0.6 (-57.2 - 56.0)	90 - 144	Arterial
Song et al. *2015*	Guardian REAL-Time, Medtronic	331 (22)	Thigh	23.7 ± 30.2		60.1	34.4	3.3	1.5	0.6	6.6. (-109.7 - 122.9)		Arterial
		270 (22)	Abdomen	23.2 ± 19.5		57.0	36.7	3.0	3.3	0.0	14.9 (-108.2 - 138.1)		
Sechterberg et al. *2015*	FreeStyle Navigator I, Abbott	183 (8)		11.1 ± 8.3	84.2						-8.0 (-49.7 - 33.8)	90 - 160	Arterial
De Block et al. *2015*	GlucoDay, A. Menarini Diagnostics	635 (35)		11.2	87.0	87.1	11.5	0.4	1.0	0.0		80 - 120	Arterial
Saur et al. *2014*	Symphony, Echo Therapeutics*	570 (15)		12.3		81.7	18.3	0.0			7.8 (-31.5 - 47.2)	100 - 180	Arterial

Leelarathna et al. *2014*	FreeStyle Navigator I, Abbott	516 (12)	Enhanced calibration	9.6 ± 8.9	87.8	87.8	12.2	0.0			-1.8 (-12.6 - 7.2)		Arterial
		544 (12)	Normal calibration	15.6 ± 12.0	70.2	70.2	29.0	0.0	0.8	0.0	-19.8 (-41.4 - 1.8)		
Kosiborod et al. *2014*	Sentrino, Medtronic	870 (21)		12.8 (11.9 - 13.6)		83.0	16.0	0.8	0.0		2.5 (-43.7 - 48.7)	<140	Venous
Boom et al. *2014*	FreeStyle Navigator, Abbott	440 (177)		13.7* [8.0 - 23.0]								90-160	Arterial
Aust et al. *2014*	CGMS System Gold, Medtronic	342 (10)				86.3	12.9	0.0	0.9	0.0	0 (limits not reported)	80-150	Arterial
Yue et al. *2013*	DGMS, San MediTech	314 (18)		14.4 ± 12.2		74.8	25.2	0.0			1.8 (-59.5 - 63.1)	140-200	Venous
Siegelaar et al. *2013*	Guardian REAL-Time, Medtronic	(60)		14.0 [11.0 - 18.0]		73.2	25.2	1.3				90-144	Arterial
	FreeStyle Navigator I, Abbott			11.0 [8.0 - 16.0]		81.8	17.7	0.5		0.0			
Leelarathna et al. *2013*	FreeStyle Navigator I, Abbott	(27)		7.0* [3.5 - 13.0]	87.8							110 - 180	Arterial
Kopecky et al. *2013*	Guardian REAL-Time, Medtronic	277 (24)				66.4	31.1	0.0	2.5	0.0		80 - 110	Arterial
Lorencio et al. *2012*	Unspecified, Medtronic	956 (41)		13.5 (6.0 - 24.1)	68.1						6.4 (-53.1 - 65.8)	120 - 160	Arterial

Siegelaar et al. *2011*	Guardian REAL-Time, Medtronic	1017 (60)		14.0 [11.0 - 17.0]								Arterial
	FreeStyle Navigator I, Abbott			10.0 [8.0 - 16.0]								
Brunner et al. *2011***	Unspecified, Medtronic	2045 (177)		7.3 (6.8 - 7.8)	92.9	99.1		0.5	0.4	0.0	2.0 (-21.0 - 25.0)		Arterial
Rabiee et al. *2009*	Unspecified, Dexcom	84 (19)				75.0	25.0	0.0				90 - 120	Unknown
Holzinger et al. *2009*	CGMS System Gold, Medtronic	736 (50)			94.0	98.6		0.0	0.7	0.7	0.7 (-1.4 - 2.9)		Arterial
De Block et al. *2006*	GlucoDay, A. Menarini Diagnostics	820 (50)	2-pt calibration			72.5	22.2	4.5	0.7	0.1		110 - 140	Arterial
		555 (50)	6-pt calibration			80.5	16.2	1.6	1.4	0.2			
Corstjens et al. *2006*	CGMS System Gold, Medtronic	165 (19)				87.3	12.7	0.0			1.8 (-41.4 - 36.9)	110 - 140	Arterial
CHILDREN													
Piper et al. *2006*	CGMS System Gold, Medtronic	246 (20)		17.6		66.3	32.5	0.0	1.2	0.0			Arterial
Branco et al. *2010*	CGMS System Gold, Medtronic	34 (14)		23.0		53.0	47.0	0.0					Arterial

Bridges et al. *2010*	Guardian REAL-Time, Medtronic	1555 (47)	15.3	74.6	23.3	2.1		0.0	-1.5 (-59.5 - 56.5)		Unknown
Phrabhudesai et al.	Guardian REAL-Time, Medtronic (Enlite sensor)	235 (19)	17.3*	66.0	28.5	0.0	7.2	0.0	-5.1 (-76.8 - 66.6)		Unknown

[a] Total number of paired samples, in parenthesis the number of included patients.

[b] MARD is reported with its corresponding 95% confidence interval, SD (±), or [IQR]**.** Reported MARD is of the entire glycemic range

[c] Percentage of measurements >75 mg/dL that are within 20% of the reference measurement (ISO15197:2003)

[d] Clarke error grid reports the percentage of measurements in zones A to E

[e] Bland-Altman analysis is reported as mean bias (limits of agreement)

[f] Ranges are when necessary converted into mg/dL, and rounded to dozens

* Indicates a median ARD instead of a MARD

** Combined analyses of Holzinger, 2009 and Holzinger, 2010

*** Patients received both the subcutaneous FreeStyle Libre and the intravascular Eirus System (results on the latter are in table 4).

Abbreviations:

CGM, continuous glucose monitoring;

CI, confidence interval;

ISO, international organization for standardization;

MARD, mean absolute relative difference;

N, number (of patients)

Table 4. Reported accuracy outcomes of the included studies that assessed intravascular CGM systems (n= 11)

Author *Year*	CGM system	N[a]		MARD[b] (%)	ISO[c] (%)	Clark error grid[d] (% in zones A - E) A	B	C	D	E	Bland-Altman[e] (mg/ dL)	Target range ICU[f] (mg/dL)	Reference sample
Schieren-beck et al. *2016*	Eirus System, Maquet Critical Care*	514 (26)		6.5 ± 8.2	90.0	94.0	6.0		0.0		0.9 (-27.0 - 29.0)	80 - 149	Arterial or venous
Nohra et al. *2016*	Optiscanner 5000, Optiscan	347 (24)		8.0 (7.3 - 8.7)		94.8	5.2		0.0		- 5 (-28 - 18)		Unknown
Leopold et al. *2016*	Eirus System, Maquet Critical Care	594 (12)		7.5	93.6	93.6	6.4		0.0		4.1 (-20.5 - 28.6)	90 - 144	Arterial
Strasma et al. *2015*	Glucath, Medtronic	1799 (70)	Arterial sensor	9.6	89.4						-2.1 (-34.5 - 29.6)	100 - 180	Arterial or central venous
		1799 (70)	Venous sensor	14.2	72.2						-6.5 (-53.8 - 39.8)		
Macken et al. *2015*	GluCath, Medtronic	758 (20)		6.4	97.0						-10.8 (-466.2 - 446.4)		Arterial
Crane et al. *2015*	GlySure, GlySure	(33)	Cardiac surg. patients	9.9		88.2		11.8					Venous
		(14)	General patient	8.0		95.0		5.0					

Bochiccio et al. *2015*	IVBG System, Edwards Lifesciences	996 (100)	8.2 ± 10.5	93.3	93.2	5.8	0.2	0.8	0.0			Arterial or venous
Foubert et al. *2014*	GlucoClear, Edwards Lifesciences	1093 (10)	5.1	99.4	99.4	0.6	0.0			-3 (-15.6 - 9.6)	80-110	Venous
Flower et al. *2014*	GluCath, Medtronic	437 (21)	13.0	80.8						-5.8 (-54.5 - 42.9)		Arterial
Schieren-beck et al. *2013*	Eirus System, Maquet Critical Care	607 (30)	5.6	97.2	97.0		3.0	0.0		-2.2 (-14.8 - 10.5)		Arterial
Schieren-beck et al. *2012*	Eirus System, Maquet Critical Care	994 (50)	5.0	99.2	99.0		1.0	0.0		0.4 (-19.5 - 22.0)		Arterial and venous

[a] Total number of paired samples, in parenthesis the number of included patients
[b] MARD is reported with its corresponding 95% confidence interval, SD (±), or [IQR]. Reported MARD is of the entire glycemic range
[c] Percentage of measurements >75 mg/dL that are within 20% of the reference measurement (ISO15197:2003)
[d] Clarke error grid reports the percentage of measurements in zones A to E
[e] Bland-Altman analysis is reported as mean bias (limits of agreement)
[f] Ranges are when necessary converted into mg/dL, and rounded to dozens
* Patients received both the intravascular Eirus System and the subcutaneous FreeStyle Libre (results on the latter are in table 3).

Abbreviations:
CGM, continuous glucose monitoring;
CI, confidence interval;
ISO, international organization for standardization;
MARD, mean absolute relative difference;
N, number (of patients)

APPENDIX A
Ovid MEDLINE search

Database(s): **Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)** 1946 to Present
Search Strategy: **2016-07-26**

#	Searches	Results
1	((glucose or BG) adj1 (sensor* or biosensor* or continuous* or realtime or real time)).tw,kf.	5099
2	((glucose or BG) adj monitor*).tw,kf.	5227
3	((continuous* or real time or subcutan* or arter* or venous or intravasc*) adj monitor* adj12 glucos*).tw,kf.	325
4	(continuous adj3 (glucose measurem* or glucosemeter*)).tw,kf.	93
5	(CGM or CGMs or GCMS* or RTCGM* or BGM or BGMs or CIGM* or IVCGM* or scCGM* or sCGM* or tCGM* or MDCGM* or IACGM).tw,kf. and glucose.mp.	1543
6	(IVBG or (intraven* adj (blood glucose or BG) adj3 system*)).tw,kf.	4
7	(GlucoDay* or Freestyle or Libre or Navigator or Medtronic* or MiniMed* or Sentrino* or Enlite* or Optiscanner or Eirus* or Glucath* or Glysure* or Symphony or Glucoclear* or (DexCom adj1 STS) or (Guardian adj2 real time)).tw,kf. and glucose.mp.	552
8	clarke error grid.tw,kf.	184
9	**or/1-8 [CGM]**	**8143**
10	animals/ not humans/	4248414
11	**9 not 10 [human CGM]**	**7501**
12	critical care/ or exp life support care/ or subacute care/ or intensive care units/ or critical illness/ or burn units/ or coronary care units/ or recovery room/ or respiratory care units/ [MESH]	105986
13	((intensive or critical*) adj (care or ill*)).tw,kf.	145729
14	(severe* adj (burn* or ill)).tw,kf.	8464
15	((coronary or cardiac) adj2 care).tw,kf.	8404
16	(((acute care or respiratory care or acute stroke or burn*) adj unit*) or acute stroke care).tw,kf.	3140
17	(trauma adj2 (cent* or unit*)).tw,kf.	13030
18	recovery room*.tw,kf.	2924
19	emergency medicine.tw,kf.	11055
20	(intensivmed* or (intensive and care)).jw,ot.	25004
21	(ICU or ICUs or CCU or CCUs or MICU or MICUs or CVICU* or SICU or SICUs or BICU or BICUs).tw,kf.	43021
22	(ECMO or ECLS or (extracorporeal adj3 (circulation or circuit or bypass* or life support* or ventricular assist*))).tw,kf.	14712
23	**or/12-22 [ICU]**	**261652**
24	**11 and 23 [CGM + ICU]**	**370**
25	**remove duplicates from 24**	**356**

APPENDIX B Overview of important assessment tools to evaluate point accuracy of continuous glucose monitoring (CGM) systems

Tool	Definition	Strengths	Weaknesses	ICU recommendations
Mean Absolute Relative Difference (MARD)	Percentage difference between CGM sensor reading and a value measured at the same time using a reference method. Derived as ([sensor-reference]/reference) x 100%.	Easy to compute and interpret, can be computed in different glucose range.	No distinction between positive and negative or systematic and random errors. Affected by glucose values and study design. Often unclear whether MARD or median absolute relative difference is computed.	Acceptable when <14% [73], >18% indicates poor accuracy [31].
ISO 15197 guideline (2003)	Percentage CGM sensor readings within 15 mg/dL from the reference when the blood glucose is ≤75 mg/dL or within 20% from the reference when the blood glucose is >75 mg/dL.	Simple.	Does not take rate of glucose change and temporal order of the measurements into account. Resting 5% can differ by any amount.	≥ 98% within 12.5% of a reference standard (or within 10 mg/dL for reading <100 mg/dL), remaining 2% within 20% [31].
ISO 15197 guideline (2013)	Percentage CGM sensor readings within 15 mg/dL from the reference when the blood glucose is ≤100 mg/dL or within 15% from the reference when the blood glucose is >100 mg/dL.			
Clarke error grid	Pairs CGM sensor readings with reference measurements, and categorizes pairs in terms of the consequence of treatment decisions. **Zone A** Within 20% of the reference value, clinically accurate. **Zone B** > 20% difference from the reference value, benign errors since it will not lead to inappropriate clinical decisions. **Zone C** Overcorrection errors, unnecessary but harmless corrections. **Zone D** Dangerous failure to detect hypo- or hyperglycemia **Zone E** Erroneous treatment error (opposite of intended treatment)	Simple. Indicates clinical significance by showing the implication on therapy.	Developed for capillary blood glucose testing systems. Original grid was designed with an arbitrary target range of 70-180 mg/dL and assumes no change in treatment when readings lie within that range. No allowance for the rate at which blood glucose concentration is changing or the frequency with which the blood glucose concentration is being measured.	100% in zone A + B, favorably in zone A [74]
Bland-Altman plot	Plot of the reference measurement or average of the two (x-axis) against the difference between CGM system and reference measurement (y-axis).Reported as mean bias with upper and lower limits of agreement (mean bias ± 1.96 x SD). Represents the random variation around the mean bias.	Simple. Possibility to distinct between systematic and random error.	Does not allows for the effect of different ranges and trend.	No recommendations.

REFERENCES

1. van den Berghe, G., et al., Intensive insulin therapy in critically ill patients. N Engl J Med, 2001. 345(19): p. 1359-67.
2. Preiser, J.C., et al., Metabolic response to the stress of critical illness. Br J Anaesth, 2014. 113(6): p. 945-54.
3. Krinsley, J.S., Association between hyperglycemia and increased hospital mortality in a heterogeneous population of critically ill patients. Mayo Clin Proc, 2003. 78(12): p. 1471-8.
4. Gale, S.C., et al., Poor glycemic control is associated with increased mortality in critically ill trauma patients. Am Surg, 2007. 73(5): p. 454-60.
5. Whitcomb, B.W., et al., Impact of admission hyperglycemia on hospital mortality in various intensive care unit populations. Crit Care Med, 2005. 33(12): p. 2772-7.
6. Sung, J., et al., Admission hyperglycemia is predictive of outcome in critically ill trauma patients. J Trauma, 2005. 59(1): p. 80-3.
7. Krinsley, J.S., Glycemic control in the critically ill - 3 domains and diabetic status means one size does not fit all! Crit Care, 2013. 17(2): p. 131.
8. Van den Berghe, G., et al., Intensive insulin therapy in the medical ICU. N Engl J Med, 2006. 354(5): p. 449-61.
9. Devos, P. and J.C. Preiser, Current controversies around tight glucose control in critically ill patients. Curr Opin Clin Nutr Metab Care, 2007. 10(2): p. 206-9.
10. Brunkhorst, F.M., et al., Intensive insulin therapy and pentastarch resuscitation in severe sepsis. N Engl J Med, 2008. 358(2): p. 125-39.
11. Arabi, Y.M., et al., Intensive versus conventional insulin therapy: a randomized controlled trial in medical and surgical critically ill patients. Crit Care Med, 2008. 36(12): p. 3190-7.
12. Finfer, S., et al., Intensive versus conventional glucose control in critically ill patients. N Engl J Med, 2009. 360(13): p. 1283-97.
13. Preiser, J.C., et al., A prospective randomised multi-centre controlled trial on tight glucose control by intensive insulin therapy in adult intensive care units: the Glucontrol study. Intensive Care Med, 2009. 35(10): p. 1738-48.
14. Hermanides, J., et al., Hypoglycemia is associated with intensive care unit mortality. Crit Care Med, 2010. 38(6): p. 1430-4.
15. Finfer, S., et al., Hypoglycemia and risk of death in critically ill patients. N Engl J Med, 2012. 367(12): p. 1108-18.
16. Wiener, R.S., D.C. Wiener, and R.J. Larson, Benefits and risks of tight glucose control in critically ill adults: a meta-analysis. JAMA, 2008. 300(8): p. 933-44.
17. Griesdale, D.E., et al., Intensive insulin therapy and mortality among critically ill patients: a meta-analysis including NICE-SUGAR study data. Cmaj, 2009. 180(8): p. 821-7.
18. Marik, P.E. and J.C. Preiser, Toward understanding tight glycemic control in the ICU: a systematic review and metaanalysis. Chest, 2010. 137(3): p. 544-51.
19. Jacobi, J., et al., Guidelines for the use of an insulin infusion for the management of hyperglycemia in critically ill patients. Crit Care Med, 2012. 40(12): p. 3251-76.
20. American Diabetes Association. Standards of Medical Care in Diabetes Diabetes Care, 2015. 38 (Suppl. 1):S5-S7.
21. Haluzik, M., et al., Glucose control in the ICU: is there a time for more ambitious targets again? J Diabetes Sci Technol, 2014. 8(4): p. 652-7.
22. Kanji, S., et al., Reliability of point-of-care testing for glucose measurement in critically ill adults. Crit Care Med, 2005. 33(12): p. 2778-85.
23. Hoedemaekers, C.W., et al., Accuracy of bedside glucose measurement from three glucometers in critically ill patients. Crit Care Med, 2008. 36(11): p. 3062-6.
24. Boyd, J.C. and D.E. Bruns, Effects of measurement frequency on analytical quality required for glucose measurements in intensive care units: assessments by simulation models. Clin Chem, 2014. 60(4): p. 644-50.
25. Juneja, R., et al., Computerized intensive insulin dosing can mitigate hypoglycemia and achieve tight glycemic control when glucose measurement is performed frequently and on time. Crit Care, 2009. 13(5): p. R163.
26. Aragon, D., Evaluation of nursing work effort and perceptions about blood glucose testing in tight glycemic control. Am J Crit Care, 2006. 15(4): p. 370-7.
27. Gartemann, J., et al., Nurse workload in implementing a tight glycaemic control protocol in a UK hospital: a pilot time-in-motion study. Nurs Crit Care, 2012. 17(6): p. 279-84.
28. Preiser, J.C., et al., Glucose Control in the ICU: A Continuing Story. J Diabetes Sci Technol, 2016.

29. Klonoff, D.C., et al., Consensus report of the coalition for clinical research-self-monitoring of blood glucose. J Diabetes Sci Technol, 2008. 2(6): p. 1030-53.
30. Kropff, J., et al., Accuracy of two continuous glucose monitoring systems: a head-to-head comparison under clinical research centre and daily life conditions. Diabetes Obes Metab, 2015. 17(4): p. 343-9.
31. Finfer, S., et al., Clinical review: Consensus recommendations on measurement of blood glucose and reporting glycemic control in critically ill adults. Crit Care, 2013. 17(3): p. 229.
32. Introduction on Blood Glucose Controller | Medical Equipment | NIKKISO. 2016: Nikkiso.com.
33. Clarke, W.L., et al., Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care, 1987. 10(5): p. 622-8.
34. Bland, J.M. and D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986. 1(8476): p. 307-10.
35. International Organization for Standardization. In vitro diagnostic test systems—requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus. EN ISO 15197:2003.
36. International Organization for Standardization. In vitro diagnostic test systems—requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus. ISO 15197:2013.
37. De Block, C.E., et al., Randomized Evaluation of Glycemic Control in the Medical Intensive Care Unit Using Real-Time Continuous Glucose Monitoring (REGIMEN Trial). Diabetes Technol Ther, 2015. 17(12): p. 889-98.
38. Boom, D.T., et al., Insulin treatment guided by subcutaneous continuous glucose monitoring compared to frequent point-of-care measurement in critically ill patients: a randomized controlled trial. Crit Care, 2014. 18(4): p. 453.
39. Leelarathna, L., et al., Feasibility of fully automated closed-loop glucose control using continuous subcutaneous glucose measurements in critical illness: a randomized controlled trial. Crit Care, 2013. 17(4): p. R159.
40. Kopecky, P., et al., The use of continuous glucose monitoring combined with computer-based eMPC algorithm for tight glucose control in cardiosurgical ICU. Biomed Res Int, 2013. 2013: p. 186439.
41. Holzinger, U., et al., Real-time continuous glucose monitoring in critically ill patients: a prospective randomized trial. Diabetes Care, 2010. 33(3): p. 467-72.
42. Piper, H.G., et al., Real-time continuous glucose monitoring in pediatric patients during and after cardiac surgery. Pediatrics, 2006. 118(3): p. 1176-84.
43. Bridges, B.C., et al., Continuous glucose monitors prove highly accurate in critically ill children. Crit Care, 2010. 14(5): p. R176.
44. Branco, R.G., A. Chavan, and R.C. Tasker, Pilot evaluation of continuous subcutaneous glucose monitoring in children with multiple organ dysfunction syndrome. Pediatr Crit Care Med, 2010. 11(3): p. 415-9.
45. Prabhudesai, S., et al., Accuracy of a real-time continuous glucose monitoring system in children with septic shock: A pilot study. Indian J Crit Care Med, 2015. 19(11): p. 642-7.
46. Nohra, E., et al., Results of a near continuous glucose monitoring Technology in Surgical Intensive Care and Trauma. Contemp Clin Trials, 2016.
47. Rabiee, A., et al., Numerical and clinical accuracy of a continuous glucose monitoring system during intravenous insulin therapy in the surgical and burn intensive care units. J Diabetes Sci Technol, 2009. 3(4): p. 951-9.
48. Schierenbeck, F., A. Franco-Cereceda, and J. Liska, Accuracy of 2 Different Continuous Glucose Monitoring Systems in Patients Undergoing Cardiac Surgery: Intravascular Microdialysis Versus Subcutaneous Tissue Monitoring. J Diabetes Sci Technol, 2016.
49. Song, I.K., et al., Continuous glucose monitoring system in the operating room and intensive care unit: any difference according to measurement sites? J Clin Monit Comput, 2015.
50. Leelarathna, L., et al., Accuracy of subcutaneous continuous glucose monitoring in critically ill adults: improved sensor performance with enhanced calibrations. Diabetes Technol Ther, 2014. 16(2): p. 97-101.
51. Saur, N.M., et al., Accuracy of a novel noninvasive transdermal continuous glucose monitor in critically ill patients. J Diabetes Sci Technol, 2014. 8(5): p. 945-50.
52. Foubert, L.A., et al., Accuracy of a feasibility version of an intravenous continuous glucose monitor in volunteers with diabetes and hospitalized patients. Diabetes Technol Ther, 2014. 16(12): p. 858-66.
53. Strasma, P.J., et al., Use of an Intravascular Fluorescent Continuous Glucose Sensor in ICU Patients. J Diabetes Sci Technol, 2015. 9(4): p. 762-70.
54. Schierenbeck, F., A. Franco-Cereceda, and J. Liska, Evaluation of a continuous blood glucose monitoring system using central venous microdialysis. J Diabetes Sci Technol, 2012. 6(6): p. 1365-71.

55. Wollersheim, T., et al., Accuracy, reliability, feasibility and nurse acceptance of a subcutaneous continuous glucose management system in critically ill patients: a prospective clinical trial. Ann Intensive Care, 2016. 6(1): p. 70.
56. van Hooijdonk, R.T., et al., Point accuracy and reliability of an interstitial continuous glucose-monitoring device in critically ill patients: a prospective study. Crit Care, 2015. 19: p. 34.
57. Yue, X.Y., et al., Real-time continuous glucose monitoring shows high accuracy within 6 hours after sensor calibration: a prospective study. PLoS One, 2013. 8(3): p. e60070.
58. Corstjens, A.M., et al., Accuracy and feasibility of point-of-care and continuous blood glucose analysis in critically ill ICU patients. Crit Care, 2006. 10(5): p. R135.
59. Leopold, J.H., et al., Point and trend accuracy of a continuous intravenous microdialysis-based glucose-monitoring device in critically ill patients: a prospective study. Ann Intensive Care, 2016. 6(1): p. 68.
60. Crane, B.C., et al., The Development of a Continuous Intravascular Glucose Monitoring Sensor. J Diabetes Sci Technol, 2015. 9(4): p. 751-61.
61. Sechterberger, M.K., et al., Accuracy of Intra-arterial and Subcutaneous Continuous Glucose Monitoring in Postoperative Cardiac Surgery Patients in the ICU. J Diabetes Sci Technol, 2015. 9(3): p. 663-7.
62. De Block, C., et al., Intensive insulin therapy in the intensive care unit: assessment by continuous glucose monitoring. Diabetes Care, 2006. 29(8): p. 1750-6.
63. Holzinger, U., et al., Impact of shock requiring norepinephrine on the accuracy and reliability of subcutaneous continuous glucose monitoring. Intensive Care Med, 2009. 35(8): p. 1383-9.
64. Lorencio, C., et al., Real-time continuous glucose monitoring in an intensive care unit: better accuracy in patients with septic shock. Diabetes Technol Ther, 2012. 14(7): p. 568-75.
65. Kosiborod, M., et al., Performance of the Medtronic Sentrino continuous glucose management (CGM) system in the cardiac intensive care unit. BMJ Open Diabetes Res Care, 2014. 2(1): p. e000037.
66. Siegelaar, S.E., et al., Microcirculation and its relation to continuous subcutaneous glucose sensor accuracy in cardiac surgery patients in the intensive care unit. J Thorac Cardiovasc Surg, 2013. 146(5): p. 1283-9.
67. Maahs, D.M., et al., Effect of acetaminophen on CGM glucose in an outpatient setting. Diabetes Care, 2015. 38(10): p. e158-9.
68. Macken, L., et al., Continuous intra-arterial blood glucose monitoring using quenched fluorescence sensing in intensive care patients after cardiac surgery: phase II of a product development study. Crit Care Resusc, 2015. 17(3): p. 190-6.
69. Orford, N.R., et al., Glycaemic control and long-term outcomes following transition from modified intensive insulin therapy to conventional glycaemic control. Anaesth Intensive Care, 2014. 42(2): p. 239-47.
70. Maahs, D.M., et al., Outcome Measures for Artificial Pancreas Clinical Trials: A Consensus Report. Diabetes Care, 2016. 39(7): p. 1175-9.
71. Wentholt, I.M., J.B. Hoekstra, and J.H. Devries, A critical appraisal of the continuous glucose-error grid analysis. Diabetes Care, 2006. 29(8): p. 1805-11.
72. Schierenbeck, F., et al., Evaluation of a continuous blood glucose monitoring system using a central venous catheter with an integrated microdialysis function. Diabetes Technol Ther, 2013. 15(1): p. 26-31.
73. Wernerman, J., et al., Continuous glucose control in the ICU: report of a 2013 round table meeting. Crit Care, 2014. 18(3): p. 226.
74. Clarke, W. and B. Kovatchev, Statistical tools to analyze continuous glucose monitor data. Diabetes Technol Ther, 2009. 11 Suppl 1: p. S45-54.

CHAPTER 10

Accuracy and reliability of a subcutaneous continuous glucose monitoring device in critically ill patients

ABSTRACT

Introduction

Subcutaneous continuous glucose monitoring (CGM) may have benefits in achieving glycemic control in critically ill patients. The aim of this study was to assess the accuracy and reliability of the FreeStyle Navigator I in critically ill patients and to assess patient related factors influencing the accuracy and reliability.

Materials and Methods

This study is a retrospective analysis of data from a randomized controlled trial (RCT) conducted in a 20-bed mixed intensive care unit (ICU). Analytical accuracy, clinical accuracy and reliability were assessed against arterial blood glucose samples as reference. Assessment was according to recent consensus recommendations with median absolute relative difference (median ARD), Bland-Altman plots, the ISO system accuracy standards (ISO 15197:2013) and Clarke error grid analysis (CEG).

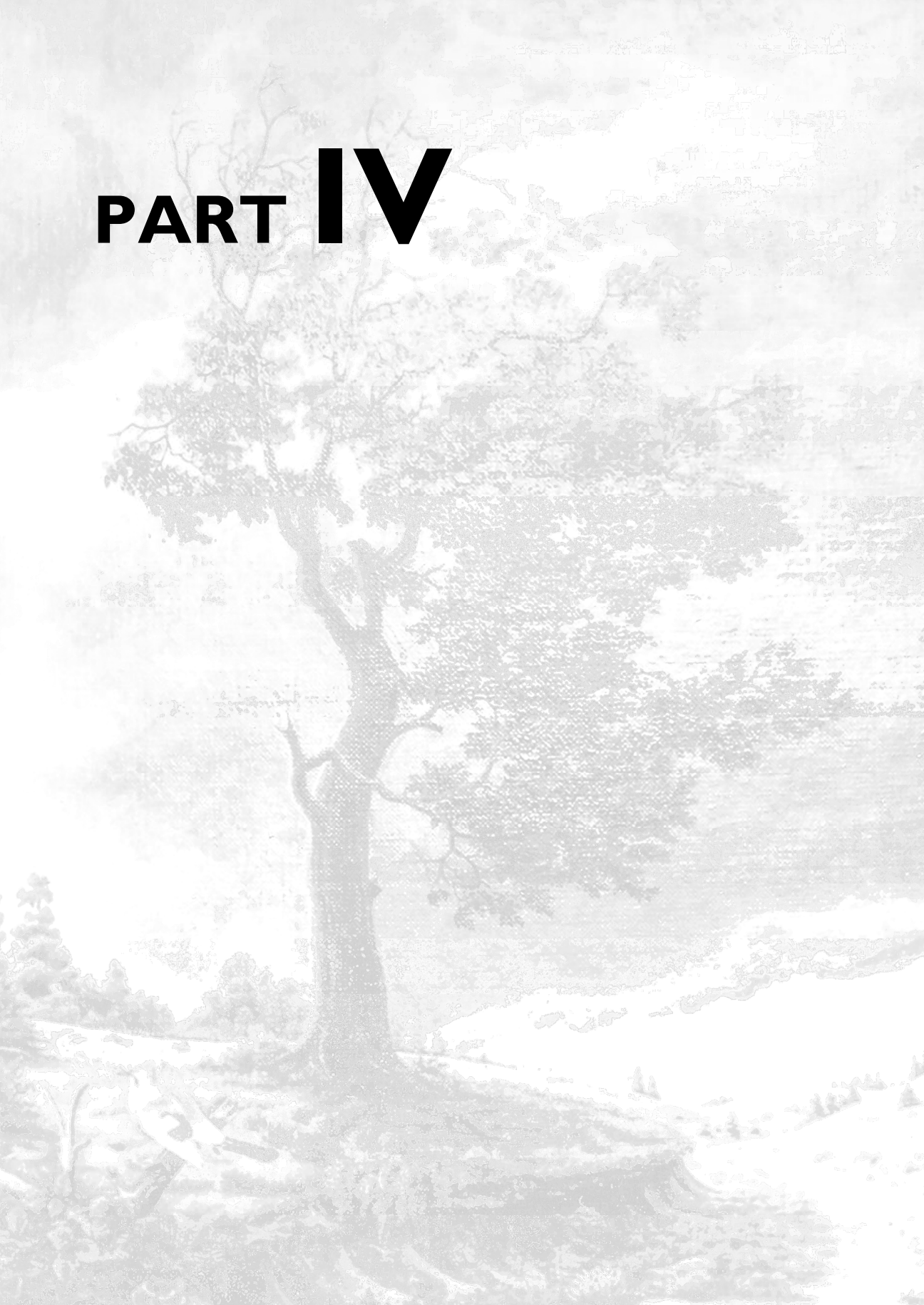
Results

We analyzed 2,840 paired measurements from 155 critically ill patients. The median ARD of all paired values was 13.3 [6.9 - 22.1]%. The median ARD was significantly higher in both the hypoglycemic and the hyperglycemic range (32.4 [12.1 - 53.4]% and 18.7 [10.7 - 28.3]% respectively, p= 0.001). The Bland-Altman analysis showed a mean bias of -0.82 mmol/L with a lower limit of agreement (LOA) of -3.88 mmol/L and an upper LOA of 2.24 mmol/L. A total of 1,626 (57.3%) values met the ISO-2013, standards and 1,334 (47%) CGM values were within 12.5% from the reference value. CEG: 71.0% zone A, 25.8% zone B, 0.5% zone C, 2.5% zone D, 0.3% zone E. The median overall real-time data display time was 94.0 ± 14.9% and in 23% of the patients, the sensor measured <95% of the time. Additionally, data gaps longer than 30 minutes were found in 48% of the patients.

Conclusion

The analytical accuracy of the FreeStyle Navigator I in critically ill patients was suboptimal. Furthermore, the clinical accuracy, did not meet the required standards. The reliability was satisfactory, however, in almost a quarter of the patients the real-time data display was <95%. The accuracy was considerably and significantly lower in hyper- and hypoglycemic ranges.

PART IV



Evidence-based practice in intensive care medicine

CHAPTER 11

Psychometric properties of the adapted McColl questionnaire: attitudes and knowledge towards evidence-based practice among nurses

S. Rijkenberg, C. van Oostveen, R. Lindeboom, J. Sturm, H. Vermeulen, A.M. Eskes

CHAPTER 12

GENERAL DISCUSSION & SUMMARY

S. Rijkenberg

GENERAL DISCUSSION AND SUMMARY

General background

The execution of critical care medicine depends on measurements. As a consequence, the quality of the applied measurement instruments is very important. Numerous clinical measurement tools are available in critical care, though several of them have not been validated sufficiently. As a consequence, it can be challenging to find the most appropriate instrument for a specific purpose. Critical care physicians and nurses require knowledge of the methods of clinical epidemiology and clinimetrics to determine whether a measurement tool or device is applicable, valid (accurate) and reliable for their clinical practice.

Aims of this thesis

The aims of this thesis were fourfold:

1. To compare the clinimetric characteristics of clinical pain assessment tools for adult mechanically ventilated patients who cannot communicate
    - To translate the most appropriate clinical pain assessment tool into Dutch and perform a cross cultural validation
2. To assess the diagnostic accuracy of a continuous QTc measurement algorithm on a patient monitor in critically ill patients
    - To determine the frequency and variability of QTc interval prolongation in critically ill patients
3. To assess the accuracy and reliability of subcutaneous continuous glucose monitoring devices in critically ill patients
    - To assess the efficacy of a subcutaneous CGM system-guided blood glucose regulation in comparison with frequent Point of Care (POC) blood glucose-guided regulation in critically ill patients.
4. To assess the psychometric properties of the Dutch version of the McColl questionnaire concerning EBP among nurses

KEY FINDINGS OF THE THESIS PART I

Pain measurement in critically ill mechanically ventilated patients

A prospective observational cohort study in 68 mechanically ventilated ICU patients who were unable to self-report pain, showed the results of a comparison of the psychometric properties of the Behavioral Pain Scale (BPS) and Critical-Care Pain Observation Tool (CPOT) (*chapter 2*). The main results were:

1. The discriminant validation of the BPS and CPOT was demonstrated by a significant increase in scores during a painful procedure compared to a non-painful procedure;
2. The CPOT score remained unchanged when comparing a non-painful procedure with rest, whereas the BPS score significantly increased during a non-painful procedure;
3. The BPS and CPOT showed fair to good overall inter-rater reliability across the four procedures and a moderate inter-rater reliability during the painful procedure; and

4. The BPS and the CPOT showed acceptable internal consistency reliability during the painful procedure.

Patients after elective cardiac surgery were not included in this study, since these patients most likely differ from general ICU patients. These patients are post-anesthesia with opioids and underwent specific surgical procedures using thoracotomy, which can lead to different behaviors and pain experiences. Therefore, in *chapter 3* a separate analysis is provided concerning mechanically ventilated critically ill patients post-cardiac surgery who were unable to communicate.

The results of the discriminant validation in the post-cardiac surgery patients were in line with the results of the general intensive care patients. Both tools, the BPS and CPOT, showed significant increases in scores during painful procedures. In addition, the BPS showed more significant increases during the non-painful procedure than the CPOT. Another similar finding was the poor discriminative performance of both tools in a subgroup of agitated patients. It appears to be difficult to discriminate between behaviors as a result of agitation and from behaviors resulting from pain. Due to the low number of agitated patients, no definite conclusions on this issue can be drawn. A new finding in the cohort of patients post-cardiac surgery (*chapter 3)* were the significant increases of the BPS and CPOT in a subgroup of sedated patients. Due to residual analgesia and sedatives, patients post-anesthesia may show different behavioral reactions to pain in comparison to general ICU patients. In addition, nurses may have more difficulties in interpreting pain related behaviors when patients are more sedated.

Both the medical ICU patients and the post-cardiac surgery patients showed that the BPS and CPOT had a substantial inter-rater reliability of the overall assessments. A striking difference between these groups of patients was the considerable lower intraclass correlation coefficients (ICC) during rest in the post-cardiac surgery patients. The ICC values are assumed to be susceptible to between subject variability and a lack of subject variability can result in low ICC values [1;2]. In this case: the proportion of the measurement error is large compared to the variability in pain scores during rest, which hampers the ability to discriminate between the patients' states. In other words, the amount of measurement error will affect the discrimination of patients with and without pain because the variation in pain scores is low during rest [3]. A lack of variability in pain scores during rest could be the result of post-surgical pain. Patients may have been afraid to move, which could have resulted in fewer movements and as a consequence lower pain scores during rest. The opposite could also have been the case: due to residual analgesia, patients could have been sedated enough to not experience pain during rest [4]. The moment of the second assessment during rest (RestII) was longer after the surgery and the patients were, as expected, less sedated as confirmed by the Richmond Agitation-Sedation Scale (RASS) score. The BPS interquartile ranges (IQRs) of the assessments of both nurses during restII were wider, which suggest patients had a higher variability in BPS scores. The ICC of the BPS was almost twice as high during restII compared to restI. The standard deviation (SD) and the histogram of the CPOT suggested a higher

variability of the CPOT scores during restII (Appendix General discussion figure 2). The ICC of the CPOT increased from 0.34 (RestI) to 0.47 (RestII). Additional analyses have been performed to gain insight in the amount and nature of measurement error (Appendix General discussion). The standard errors of measurement (SEMs) of the BPS and CPOT were 0.25 and 0.22 points higher, respectively at restII compared to restI (Appendix General discussion table 4). The amount of measurement error seems smaller compared to the variability of the pain scores at restII, despite the higher SEMs. Therefore, the higher ICCs of both scores at restII appear to be the result of a higher variability of the pain scores. The ICCs of both scores during turning were higher compared to rest (RestI & RestII), although the BPS had a higher ICC (substantial) than the CPOT (moderate). The variability of the BPS and CPOT was, as expected, higher during turning which was reflected in wider IQRs. The IQRs and SDs of both scores were similar, thus the difference in ICCs between the BPS and CPOT seem to be the result of a higher SEM of the CPOT (0.75 and 0.91 respectively). Bland Altman analyses of the agreement of both raters (all moments) showed no substantial systematic measurement error between the raters (Appendix General discussion figure 1). This suggest the measurement error is due to random error. Although, it is possible to correct for systematic error, but not for random error. Random error is inevitable and may be due to true differences between patients, differences between test conditions or differences between raters [5]. The patients were assessed at the same time, therefore a difference between test conditions can not be the source of the random error. The order of the assessments may have caused the difference between BPS (substantial ICC) and CPOT (moderate ICC). The BPS was always completed first which could have affected the quality of the second (CPOT) assessment.

The ICCs of the general critically ill patients (*chapter 2*) showed almost the reverse pattern; the ICCs during the painful procedure (turning) were lower than during the other assessments. Since IQRs and SDs of the BPS and CPOT were wider during turning, the lower ICCs during turning do not appear to be the consequence of a lower variability in pain scores. The SEMs of both the BPS and CPOT scores were more than 0.5 points higher during turning compared to rest, suggesting that the differences in ICCs are caused by a higher proportion of random measurement error in relation to the variability. In this case, the patients were likely to show different behaviors during turning than during rest and therefore, this may have been a source of random error. Although, the assessment during rest was performed a few minutes before the assessment during turning, the test circumstances were different. The raters not only had to assess the behaviors, but also had to perform the action of turning, which is complicated in critical ill patients due to the medical equipment, tubes, catheters etcetera. Finally, the raters were aware of which procedure was supposed to be painful. Therefore, they may have expected certain behaviors of the patient or interpreted behaviors as painful even though they were due to agitation or tube irritation (coughing). These expectations and interpretations can differ between nurses as a result of their level of experience.

Several previous studies generally found a higher overall inter-rater reliability [4;6-15], although it is problematic and not advisable to compare their results with our studies due to following reasons.

Firstly, numerous previous studies utilized a weighted kappa to assess the inter-rater reliability [16]. The value of kappa depends upon the prevalence of patients in each category, e.g. the proportion of patients with no pain vs. severe pain. As a consequence, it is misleading to compare kappa values from different studies when the prevalence of categories differ [17]. Other studies have used the ICC to assess the inter-rater reliability, but the same principle of different prevalences, applies to this parameter. Reliability parameters are only generalizable to populations with a similar variation in pain scores [3]. Previous study populations are potentially not comparable with our population due to our restricted sedation policy. This may have led to different experiences of pain and different behaviors of the patients in our studies. Several studies have provided limited information of the baseline characteristics or specific data concerning the amount and timing of used sedation and analgesics. Therefore, a comparison of the generalizability of their inter-rater reliability is hampered.

Secondly, none of the studies which assessed the interrater reliability with an ICC, described the model of the ICC used (Introduction table 2). In several studies the assessments were performed either by a small number of researchers and trained nurses or nurses and physicians, which may have led to systematic measurement errors between the raters [7;12;18]. Systematic errors may lead to different ICC values in the various models and therefore it is recommended to use the correct model [19;20].

Thirdly, several studies only assessed the overall inter-rater reliability or the inter-rater reliability of the domains rather than the painful versus non-painful procedures [6;7;9;19;21-23]. This could potentially lead to higher values because of a higher variability in pain scores as the painful and non-painful assessments were combined.

In addition to a potential difference in ICU-populations and variability in pain scores, the inter-rater reliability in *chapters 2 and 3* compared to previous studies may be lower due to a higher level of random measurement error. Our nurses had to assess the BPS and CPOT simultaneously, which is more demanding. A lack of time and busy shifts may have led to random errors. Furthermore, the large number (105) of individual nurses on our ICU, resulted in a relatively low exposure to the pain assessments. Nurses may have differed in level of experience and training in the pain scales which can be a source of random error [5]. Finally, in various previous studies the assessments were performed by researchers and/or a small sample of nurses. Perhaps these assessors were more motivated to perform the study procedures than our large nursing team. None of the studies presented agreement parameters like Bland-Altman analyses and SEMs, thus it is not possible to assess the amount and type of measurement error in these studies.

The internal consistency reliability in *chapter 2* and *3* was assessed during the painful procedure turning. The Cronbach's alpha values in general ICU patients during turning *(chapter 2)* for the BPS and CPOT were 0.70 (nurse1) and 0.76 (nurse1), respectively (Appendix discussion table 5a/b). The inter-item correlations ranged from 0.396 to 0.51 (BPS) and from 0.337 to 0.579 (CPOT). Inter-item correlations for items within one dimension should be between 0.2 and 0.5. A correlation above 0.7 indicates items are measuring the same thing and indicates one item could be deleted (24). All items of the BPS and CPOT had corrected-item

total values above 0.3, reflecting a sufficient degree to which an item is consistent with the other items in the subscale. Item values less than 0.3 indicates that the items are not contributing much to the discrimination of the patients [24].The internal consistency reliability in the post-cardiac surgery patient group (*chapter 3*) was lower than in general ICU patients. The values of Cronbach's alpha for both scores in *chapter 3* were <0.70 (insufficient), which indicates that the average correlation between the behavioral domains is insufficient to be considered reliable to measure a single construct (pain) and harm the discriminative ability of the pain scores [25]. These differences may be the result of differences in the pain score variability between general ICU-patients and patients post-cardiac surgery. The pain scores of the patients post-cardiac surgery have lower standard deviations than the general ICU patients in *chapter 2,* thus a lower variability. This is also supported by the Bland-Altman plots as well (Appendix General discussion figure 1). Even though, the coefficients in former studies are generally higher, these values should be interpreted with caution. Several studies did not calculate Cronbach's alpha for the sample size but used all paired assessments, which potentially leads to higher coefficient values due to a higher variability of the pain scores [7;9;26]. Another explanation for our lower internal consistency reliability compared to previous studies may be the difference between studied populations. The Cronbach's alpha is a characteristic of the pain scales used in an ICU population. Its value depends on the variation of pain scores in this population [20]. As mentioned previously, the patients in former studies may not be comparable to our patients, so the values for Cronbach's alpha may not be comparable either. Finally, it is important to stress that alpha in not an index of unidemensionality. A pain scale having a large Cronbach's alpha is not necessarily unidimensional [24;27] .

The results of *chapters 2* and *3* showed a reduced discriminant validity of the BPS, because it increased during a non-painful stimulus as well. We therefore prefer the use of the CPOT above the BPS for the assessment of pain in critically ill patients who are unable to self-report pain.

We performed a cross-cultural translation and validation of the CPOT into Dutch (CPOT-NL) for 108 critically ill medical (25%) and surgical (75%) patients (*chapter 4*). The ICCs during rest and turning in the group of mixed ICU patients in *chapter 4* were lower than in the patients post cardiac surgery (*chapter 3*). The results showed a similar pattern: a lower ICC during rest. The overall ICC value in *chapter 4* (moderate) was lower than the ICC in *chapter 3* (substantial). Bland Altman analyses of the mixed ICU population did not show substantial systematic error and the scatter plot showed less variability in pain scores than for the patients post-cardiac surgery and general ICU patients. The SEM's during rest and turning were almost equal to the values of the patients post-cardiac surgery, but the SEM of all scores was about 0.20 higher in the mixed ICU population compared to the patients post cardiac surgery (Appendix General discussion table 4). The combination of a substantial lower pain score variability with a slightly increased SEM, resulted in a decreased ability of the CPOT-NL to discriminate between patients.

The difference in overall ICCs between our studies is probably mainly the result of less variability in pain scores, thus the amount of measurement error in *chapter 4* is too large to distinguish differences in pain scores. The hypothesis that the simultaneous assessment of the BPS and CPOT may have caused random measurement error is not applicable to this study since only the CPOT was used. The raters are members of the same nursing team as *chapters 2* and *3,* and the validation of the CPOT-NL was the second study using the CPOT, so nurses should have had more experience with these pain assessments. Due to the large number of patients post-cardiac surgery in *chapter 4*, the reasons for ICU admission are generally comparable between *chapters 3* and *4*. However, there is an important difference between the patients that may be the source of random measurement error. All patients in chapter 4 had to be able to answer an "yes/no" question and were therefore less sedated. Nurses have possibly interpreted the behaviors differently because patients were able to communicate their pain. Despite the requirement of the research protocol that patients were first assessed with the CPOT and then asked to rate their pain, they could have communicated their pain level earlier during the day/shift. Therefore, the CPOT assessments could have been influenced by that knowledge. When patients did not communicate their pain, nurses could have thought their patients had no pain, while the opposite could was the case. This hypothesis may be supported by the results of the patient's pain assessments (yes/no question). Only 39% of self-reports (pain yes) measurements were assessed as painful with a CPOT score ≥2 (cut-off pain) by the nurses in charge of the patient. The second rater could have been any other available ICU nurse and, therefore, this rater was possibly not involved in the patient's care. A more independent rater is perhaps less biased by patients' previous expressions of pain. This hypothesis may be supported by the increased sensitivity of the CPOT scores of the second rater (nurse not in charge of the patient). The second rater assessed 54% of the patients' self-reports (pain yes) as painful (CPOT score ≥2).

The internal consistency reliability (Cronbach's alpha) of all measurements of the nurse in charge in this group of mixed medical surgical patients was 0.56.The Cronbach's alpha during rest of the nurse in charge was 0.55 and during turning 0.50. The Cronbach's alpha values of the second rater were 0.40 and 0.50 respectively. The inter-item correlations ranged from 0.066 to 0.414 and the corrected item-total correlations from 0.151 to 0.461, reflecting a low degree to which an item is consistent with the other items in the subscale. This indicates that the items are not measuring the intended underlying construct (pain) and harm the discriminative ability of the scale. The internal consistency reliability in this mixed ICU population was considerably lower than in the general ICU patients (*chapter 2*) and slightly lower than in the post-cardiac surgery patients (*chapter 3*). The mixed ICU population has a majority of patients post-cardiac surgery and therefore may be more or less comparable with patients post- cardiac surgery in *chapter 3*. This theory appears to be supported by the Bland Altman plots, which show a higher variation in CPOT scores in the general ICU patients and a slightly higher score variability in the patients post cardiac surgery (Appendix General discussion figure 1).

The criterion validity of the CPOT-NL (index test) was analyzed using a patients' self-reported pain (reference test). A total of 65/216 (30%) pain assessments (pain yes) in 47/108 patients (44%) were self-reported by the patient as painful. Nearly two thirds of the self-reported pain was during turning. The sensitivity of the CPOT-NL was low (39%) with a threshold score of 2. Several previous studies showed a similar threshold score, but with a higher sensitivity [13;28-30]. A low sensitivity results in an underestimation of experienced pain. As previously described, this may be due to the hypothesis that nurses had a biased perception of the patient's pain experience. The sensitivity during the painful procedure (turning) was higher (44%), while the sensitivity during rest was very low (10%). This is in line with the results of a previous study with post-cardiac surgery patients, although this study presented a substantial higher sensitivity [28]. One possible explanation for the very low sensitivity during rest could be that some patients may not be able to show their behaviors during rest even if they are in pain. Patients might have been afraid to move, which could have resulted in fewer movements and as a consequence lower pain scores. The specificity was high, thus the number of false-positive assessments is low. Consequently, the risk of unnecessary treatment with analgesics and unwanted side effects such as hypoventilation and gastrointestinal hypomotility would be avoided.

The combined findings of our studies suggest that more research is needed on the assessment of pain in restless, agitated or delirious patients because the discriminative performance of the CPOT in these patients was severely hampered. In *Chapter 5* we performed a critical appraisal of a study that strived to assess the psychometric properties of the CPOT in critically ill patients with a delirium [19]. Kanji et. al. concluded that their results suggest that the CPOT is a valid and reliable tool for the detection of pain in non-comatose, delirious adult ICU patients. There are a few methodological issues why this conclusion should be interpreted with caution. This study did not provide data about the severity of delirium, the subtype of delirium and the relation between the Richmond Agitation-Sedation Scale (RASS) and CPOT score. In addition, the assessment of the inter-rater reliability is debatable. *Kanji et al.* reported the inter-rater reliability of the four domains of the CPOT instead of the inter-rater reliability of the different procedures (painful versus non-painful or rest). A drawback of this method is that it does not comply with daily ICU practice since the CPOT is used as the sum of four domains during different occasions like tracheal suctioning or rest. However, the study of Kanji and coworkers is an important first step in the validation of the CPOT in critically ill patients with a delirium, it remains unclear whether the CPOT is an appropriate tool to assess pain in these patients.

STRENGTHS AND LIMITATIONS PART I

Pain measurement in critically ill mechanically ventilated patients

Validity

A strength of *chapters 2* and *3* (the comparisons of the psychometric properties of clinical pain assessment tools for adult mechanically ventilated patients) is the analysis of the discriminant validity in subgroups of sedated and agitated patients. Particularly in agitated patients, little is known about the discriminative performance of the BPS and CPOT [31]. Although the sample

size in these subgroups is too small to draw firm conclusions, our studies provide insight into the discriminative performance of behavioral pain tools in agitated and sedated patients. Furthermore, in contrast with previous studies, we included patients with a restricted sedation policy. Therefore, as a consequence, our patients received substantially less sedation and analgesics than patients in most previous studies. Since a restricted sedation policy is recommended in recent guidelines, it is to be expected that ill patients will receive less sedation going forward [32]. This may increase the generalizability of our results.

Our studies have a number of limitations. A limitation of *chapters 2* and *3* is the translation of the BPS and the CPOT. We did not officially translate and validate the translated pain scores before the start of the study. The study team, including a qualified English language translator, translated the short descriptions of both tools into Dutch in our clinical information system, thus language misinterpretations may have occurred.

A second limitation in these chapters is, since the assessors are trained ICU nurses, they were aware of which procedure was potentially painful. Therefore, they may have observed more behavioral changes during turning which could have resulted in higher scores and an inflation of the discriminant validation.

A possible limitation of *chapter 4* (the cross-cultural translation and validation of the CPOT into Dutch) is the lack of a confirmatory factor analysis (CFA). CFA is performed to confirm previous hypotheses, based on theory or previous analyses, about dimensions of the construct (pain). It tests whether the data fit the hypothesized factor structure of the construct to be measured [33]. CFA is a useful technique to compare two versions of an instrument (e.g. the original CPOT and the Dutch translation) [34]. Only one previous study with the CPOT performed a factor analysis [13]. This was a principal component factor analysis which is considered as a part of exploratory factor analysis (EFA). EFA is used when there is no clear idea about the number of dimensions in an instrument [35]. Since its development, the CPOT has been presented as the sum score of four behavioral domains and therefore EFA was possibly not the appropriate method. Additionally, the minimum sample size for factor analysis is a topic for debate [36]. Therefore, it is unclear whether the study of Li et al. (70 participants) [13] has a sufficient number of patients to perform a factor analysis. We considered the sample size in our study (*chapter 4*) to be insufficient to perform a factor analysis. An assessment of the reliability and validity of the CPOT by ICU nurses in a clinical setting requires a considerable amount of time from nurses, who are primarily responsible for the care of critical ill patients. Each pain assessment required two nurses simultaneously, and each patient was assessed twice. Therefore, we did not include a larger number of patients to enable an assessment of the dimensionality by factor analysis.

Finally, it is debatable whether the mixed ICU population, in which the cut-off value of the CPOT-NL has been established, is appropriate. Although the majority of this sample consists of patients post cardiac surgery, general ICU patients were included as well. The differences in psychometric properties between general ICU patients (*chapter 2*) and patients post cardiac surgery (*chapter 3*) suggest that the cut-off values may differ in these groups. This is also supported by the studies of Gelinas et al. who established different cut-off values in a sample of

55 conscious and unconscious mainly medical and trauma ICU patients (cut-off >3) and 99 patients post cardiac surgery (cut-off >2) (12;28). This issue concerns the assessment of the reliability as well. In retrospect, it might have been better to validate the CPOT-NL in general ICU-patients only or patients post-cardiac surgery.

Reliability

A strength of the clinical studies in part I, is the assessment of the inter-rater reliability in a large group of nurses. Several previous studies had a lower number of assessors. Furthermore, pairs of nurses differed in our study, only the bedside nurse was constant. In numerous previous studies, either one of the investigators or the physicians participated in the assessments. However, pain assessment by a large group of nurses is a reflection of real life intensive care.

A limitation with regard to the reliability of our clinical studies (*chapters 2 3,4*) is the lack of the assessment of the intra-rater reliability. Each assessor might apply different criteria from day to day, which potentially results in random measurement errors. The intra-rater reliability could have been tested by using videotapes of mechanically ventilated critically ill patients during potential painful procedures and rest [37].

Secondly, none of our studies assessed the amount and type of measurement error between the raters. Therefore, to address this limitation, additional analyses have been performed and these results have been processed in the discussion (Appendix General discussion). Finally, the incorrect ICC model was presented in chapters 2 and 3, even though we decided to use the one-way ANOVA (model 1). The model 2 ICC (two-way ANOVA random) is the incorrect model since the patients were not assessed by each rater but by a different set of randomly selected raters (nurses) from a larger population [2]. It is important to note that none of the models (1, 2 or 3) resulted in different coefficients, however, the correct one should have been presented since application of an incorrect model can lead to different coefficients. This could have resulted in an incorrectly inflated ICC, since the one-way ANOVA is in general the most conservative model (Introduction of the thesis table 2).

IMPLICATIONS AND FUTURE RESEARCH PART I

Pain measurement in critically ill mechanically ventilated patients

The inter-rater reliability in this thesis was generally lower compared to the inter-rater reliability in previous studies. This may be the result of a lower number of assessors in these studies, which may imply that these assessors are better trained, have more experience with using the pain scales and be more motivated to conduct the assessments. In addition, in numerous previous studies, either one of the investigators or the physicians participated in the assessments, which is not a reflection of daily clinical practice. However, pain assessments performed by a large group of intensive care nurses is a reflection of daily clinical practice and the generally lower inter-rater reliability in our studies could therefore be a real consequence of this. Furthermore, reliability parameters are only generalizable to populations with a similar

variation in pain scores [3]. Previous study populations are potentially not comparable with our population; therefore, a comparison of the inter-rater reliability is hampered.

The inter-rater reliability of the CPOT-NL was fair during rest and moderate during turning. The variability in pain scores was low, thus the amount of measurement error will affect the ability of the CPOT-NL to discriminate between patients with or without pain.

The internal consistency reliability was insufficient in patients post cardiac surgery (during turning) and in the mixed ICU population (overall, turning and rest), which suggests that the items are not measuring the intended underlying construct (pain) and the discriminative ability of the CPOT-NL is hampered. Our aim was to validate the complete original CPOT into Dutch. Therefore, we chose not to delete the domain compliance ventilator, although removal of this domain would have increased the Cronbach's alpha up to 0.60 which is still insufficient. When the internal consistency reliability is low (< 0.70), additional revisions and improvements to the scale should be considered. An assessment of the internal consistency reliability of the CPOT-NL in a cohort with only general ICU patients, with and without the domain compliance ventilator, may result in a higher Cronbach's alpha. Furthermore, additional behavioral domains may improve the reliability of the CPOT. Domains of the COMFORT scale (assessment of distress, sedation and pain in nonverbal pediatric patients), such as alertness and calmness/agitation are possible options, since the reliability of this score seems adequate [38]. Although adding more items into a scale can increase the Cronbach's alpha, the addition of items that measure the same thing as the current items leads to a redundancy that is inefficient. As a consequence, the pain score takes longer to assess, which hampers the use in daily clinical practice [27]. This topic requires further research and revision of the CPOT-NL seems necessary in order to improve its psychometric properties.

Patients post cardiac surgery have generally a limited period in which the assessment of pain with a behavioral pain scale is appropriate. The ICU length of stay of this population is usually less than one day and patients are relatively soon after ICU admission sufficiently conscious to be able to rate their pain with a NRS. Despite the poor internal consistency reliability of the CPOT-NL, the fair inter-rater reliability during rest and the moderate inter-rater reliability during turning, it is strongly advisable to use this tool because there is currently no superior alternative [39]. Clinicians should, however, be aware of the limitations when translating their findings to clinical decision making. Furthermore, the reliability of the CPOT-NL has yet not been assessed in the appropriate population: patients who are unable to rate their pain. Future studies should assess the inter-rater reliability and internal consistency reliability of the CPOT-NL in patients who are unable to communicate their pain as well. In addition, the psychometric properties of the CPOT-NL should be assessed separately in patients post-cardiac surgery and general ICU patients. The results of chapter 2 suggest that the reliability and validity may be better in general ICU patients compared with patients post cardiac surgery or in a mixed ICU population.

Delirium is a common complication in ICU patients and the incidence of delirium after cardiac surgery varies between 3 - 55% [40]. The overall incidence in critically ill patients is on average 30 - 50% [41]. Self-report of pain in this vulnerable group of patients is complicated because of

the limited communication, the variable level of consciousness and a potential different presentation of pain. Pain and agitation may interfere in delirious patients resulting in a higher CPOT score due to agitation instead of pain. The findings of this thesis suggest that future research is needed on the assessment of pain in restless, agitated and delirious patients. To date, both the BPS and CPOT seem not applicable in critically ill patients with a delirium.

The Dutch version of the CPOT showed a low sensitivity of 39% and a specificity of 85% at a CPOT threshold score of 2. For use in daily clinical practice, a patients' self-report of pain is always superior. The CPOT should only be used when patients are not able to self-report pain and nurses must remain vigilant for other factors that could increase the CPOT score. However, more importantly, nurses need to be aware of the risks of procedural pain in critically ill patients and therefore frequently assess pain. Furthermore, they must assess the level of sedation and be aware of other signals which may indicate pain, in order to prevent underestimation of pain in these vulnerable patients. Due to the high specificity, the risk of unnecessary treatment with analgesics is low.

The design of the pain scales; three and four domains with the assignment of 1 to 4 and 1 to 2 points, respectively, suggests an ordinal scale. Nevertheless, the scales have never been validated for the application in a typical ordinal manner. To date both the BPS and CPOT do not contain categories like mild, average or severe pain. Therefore, the CPOT and BPS are not able to discriminate between mild, moderate and severe pain. The CPOT can only detect the presence versus absence of pain [42] and the threshold scores for pain vary in previous studies (Appendix *chapter 3*). Since the cut-off score is uncertain and the discrimination of the CPOT is often hampered when pain score variability is low, critical care nurses should combine the cut-off score with their clinical judgment when assessing pain and initiating pain reducing interventions in non-verbal critically ill patients [43].

The pain scores should be used in combination of an assessment of the level of sedation with, for example, a RASS score. This involves interpreting an increase or decrease of a score, rather than interpreting a specific score. This requires structural and frequent pain assessments during rest and during potential painful procedures and after administering analgesics. By repeating the pain assessments and taking the average of the pain scores, the random measurement error may be decreased [5].

Therefore, critical care nurses need to be frequently educated in the assessment and consequences of pain in critically ill patients. Nurses must also be motivated to assess pain in their patients. An assessment of the intra-rater reliability by using videotapes of critically ill mechanically ventilated patients could be part of this training. To establish a valid cut-off score for the CPOT and increase the criterion validity, future research should include advanced technology such as pain assessment monitors to investigate physiological reactions involved in nociception (44). Furthermore, clinicians must be aware that the psychometric properties of a measurement tool or device must be assessed in the (patient) population to which it will be applied.

In conclusion, our studies in mechanically ventilated ICU patients who were unable to self-report pain, suggest that the overall inter-rater reliability of the BPS and CPOT is substantial. However, during rest, the inter-rater reliability of both scores was fair to moderate in patients post-cardiac surgery. In general ICU patients the ICC of the BPS was moderate during turning. The amount of measurement error in our studies affected the ability of the pain scores to discriminate between patients with or without pain when the variability in pain scores was low.

The internal consistency reliability in general ICU patients was acceptable which is in contrast to the internal consistency reliability in patients post cardiac surgery. The Cronbach's alpha of both the BPS and CPOT was insufficient, which suggests that sum scores should not be calculated in these patients.

The BPS had a slightly reduced discriminant validity compared to the CPOT. Therefore, we advise to use the CPOT in critically ill patients. The cross-cultural translation of the CPOT resulted in a Dutch version of the CPOT with a fair inter-reliability during rest and a moderate inter-rater reliability during a painful procedure. The internal consistency reliability (Cronbach's alpha) was <0.70 (insufficient). Consequently, the impeded reliability of the CPOT-NL in our study indicates that the amount of measurement error will affect discrimination of the patients when the variability in pain scores is low. The sensitivity of the CPOT-NL was 39%, with a specificity of 85% at a CPOT threshold score of 2. The result of the low sensitivity of the CPOT-NL is an underestimation of pain. To date it is unclear whether the CPOT or the BPS are appropriate tools to assess pain in critically ill patients with a delirium.

Pain assessment in critically ill patients who are unable to adequately communicate is complex. Despite its psychometric limitations, the CPOT is considered as the most appropriate tool to assess pain in critically ill patients who are not able to self-report their pain [39].Currently, we do not have a superior alternative, however not assessing pain in this vulnerable population is not an option. Future research should focus on assessing and improving the psychometric properties of the CPOT in different populations of critically ill patients.

KEY FINDINGS OF THE THESIS PART II

Continuous QTc measurement in critically ill patients

Chapter 6 presents the results of a single center prospective observational cohort study of 50 critically ill patients with a QRS duration <120 milliseconds. The primary aim of this study was to determine the frequency of QTc interval prolongation measured by continuous QTc measurement with the patients' monitor (Philips Healthcare IntelliVue MP70, Amsterdam, the Netherlands). The secondary aim was to determine the variability of the QT and QTc interval over time. A prolonged QTc interval was defined as more than 500 milliseconds during at least 15 consecutive minutes, equal for men and women. The variability of the QT interval over time was assessed by using the QT variance and the QT variability index (QTVI) [45]. Patients were monitored by a 5-electrode ECG, allowing to obtain 7 leads (I, II, III, aVR, aVL, aVF, V1).

All patients were followed up until their discharge from the ICU or until the subjectively chosen maximum follow-up time of 14 days.
A total of 221,405 minutes of continuous QT-interval monitoring data were obtained, with a median of 2,892 minutes per patient, ranging from 709 to 19,614 minutes. The mean Bazett QTc interval was 457 ±33 milliseconds, ranging from 301 to 669 milliseconds. Corrected QT-interval prolongation occurred in 26 patients (52%). The proportion of QTc values that were prolonged in a single patient varied from 0.2% to 91.3%, with a median of 13% and interquartile range (IQR) from 2% to 45%. The mean monitoring time (number of values) in the prolonged QTc and nonprolonged QTc groups was not different (respectively 2884 IQR [1359 to 7176] and 2892 IQR [1610 to 4854], p= 0.88). Univariate analysis showed that the Acute Physiology and Chronic Health Evaluation IV (APACHE) predicted mortality score, and the sequential organ failure assessment score (SOFA) were significantly higher in the patients (n= 26) with a prolonged QTc interval compared to the patients without QTc interval prolongation (n= 24). The incidence of cardiac arrhythmias was higher in the prolonged QTc group but did not reach a level of significance. The incidence of TdP was 1 (3.8%) of 26 in patients with prolonged QTc interval.
The variance of both the QT and the QTc over the monitoring time was significantly prolonged in the patients with a prolonged QTc, which indicates more electrical instability of the ventricular conductance.
Although continuous QTc measurement has been available since 2008, it had not yet been validated for use in critically ill patients. Therefore, the aim of the study in *chapter 7* was to validate the continuous QTc measurement on a patient monitor against the golden standard (manually measured QT and RR interval and calculated QTc on standard 12-lead ECG) in critically ill patients. This retrospective single-centre cohort presents the analysis of 119 consecutive patients with a QRS <120 msec, which was a sufficient number according to the sample size calculation. Patients were included when, on the first morning of admission, a 12-lead ECG was performed and at the same time a valid continuous QTc measurement on the patient monitor was obtained. On the 12-lead ECG – QT, QRS and RR interval were measured manually on a single selected lead by a cardiologist, blinded for the results of the continuous measurements.
QTc Pearson's correlation coefficient and the ICCs w ere substantial (46). The Bland-Altman analysis showed a mean bias of 20 msec, and thus continuous QTc (index test) measures a structurally higher QTc duration than manual measurement on standard 12-lead ECG (reference test). This is in line with previous research and can be explained by the use of more leads by the patient monitor for analysis instead of one lead with the manual measurement to calculate QTc [47]. Limits of agreement were relatively large, indicating a large random error. This may be the consequence of a known large QTc difference in different leads. Some consider differences of up to 50 msec in QT intervals measured in the different leads in normal subjects as being normal [48], others have suggested that differences of up to 65 msec were still within the average limits [47;49].

STRENGTHS AND LIMITATIONS PART II

Continuous QTc measurement in critically ill patients

Chapter 6 is the second study using continuous QTc measurements to evaluate the prolongation of the QTc interval in critically ill patients. The study of Pickham et al. [50] concluded that QTc prolongation in critically ill patients was common (24%) and increased the mortality risk almost three times. Our study suggests that QTc prolongation in the ICU might be even more common (52%). Another finding was the association between severity of illness (Apache IV pm) and prolonged QTc. In addition, the present study measured QTc variability and QT variance, which appeared to be associated with a prolonged QTc interval as well.

The most important limitation of *chapter 6* was the limited sample size of 50 patients and the relatively short follow up time. Therefore, a comparison of mortality in patients with QTc prolongation versus patients without QTc prolongation was not achievable.

Another limitation is that 15 of the analyzed patients were already present in the ICU at the start of the study and were consequently not assessed from the start of their admission. However, 8 (53%) of them had a prolonged QTc, which corresponds with the 52% of the total sample. Finally, the automated continuous measurements of the QTc in *chapter 6* were not compared with the golden standard: manual measurements on a single selected lead from a 12-lead ECG. As a consequence, measurement error could have occurred.

Chapter 7 is the first study to validate continuous QTC measurements in critically ill patients. The analysis covered a broad range of QTc values and presented consistent findings in the Bland-Altman plot over this range.

A limitation of this study is the analysis of one time point per patient. Consequently, due to the variability over the course of illness, it is uncertain whether our results are consistent over time for an individual patient. Another important point is the lack of consensus in the literature about what QTc interval threshold should be considered clinically important. Accordingly, we did not report the sensitivity (true positive) and specificity (true negative) of the continuous QTc . Additionally, the onset of QTc prolongation, as well as the total duration of the prolonged QTc, varies. The manual measurements should have been repeated frequently to determine whether QTc prolongation was present for a longer period of time (e.g. 15 minutes). This would have been difficult to achieve, since only one 12-lead ECG device was available. Furthermore, it would have been a burden for both nurses and patients.

Finally, we did not determine the inter-rater reliability of the manual measurements in the 12-lead ECG. A second rater might have resulted in differences in the QTc duration of the manual measurements, since measurement error may have occurred.

CLINICAL IMPLICATIONS AND FUTURE RESEARCH PART II

Continuous QTc measurement in critically ill patients

There is no consensus in previous studies about the level of QTc interval threshold that should be considered clinically relevant [51-54]. Numerous authors used the 500-millisecond threshold [54-56] and Pickham et al. [56] suggested 15 consecutive minutes. As research with continuous QTc measurements is scarce, there is no evidence up till now for the time that a

prolonged QTc has to be present to create a higher risk of TdP. Further research is required to establish whether a prolonged QTc of 15 consecutive minutes is clinically relevant or that a longer period is more associated with arrhythmias. Additionally, it is debatable whether current accuracy standards are appropriate to evaluate continuous QTc measurement, since they don't take variability and clinical consequences into account.
Severity of illness was significantly associated with QTc prolongation and may therefore be an independent risk factor for QTc prolongation. Our results suggest that continuous QTc measurements are a valuable addition to the standard ECG monitoring. Future studies should investigate more in detail whether it is advisable to monitor critically ill patients with continuous monitoring of the QTc interval, especially when the severity of illness is high.
Continuously measured QTc measures a structurally higher QTc duration than manually measurement on standard 12-lead ECG. This might result in a false positive QTc prolongation. In addition, the Bazett formula will lead to less prolonged QTc measurements compared with the Fridericia formula. For use in clinical practice, it is essential and reassuring to realize that the continuously measured QTc might be longer than the QTc as manually measured in an individual lead.
Limitations of QTc monitoring include high heart rate, since a heart rate more than 150 beats/min leads to P and T waves approaching each other too much. A wide QRS complex may also confound the QT measurements. Therefore, continuous QTc measurement seems not applicable yet for these patients.

In conclusion, critically ill patients are at high risk of QTc prolongation. Part II of this thesis shows that intermittent QTc measurement may underestimate the occurrence of QTc prolongation compared to continuous QTc measurement. Our study implies that QTc prolongation may be even more prevalent than previous research revealed. Continuous QTc measurement in critically ill patients with a QRS duration shorter than 120 msec shows an acceptable accuracy and can be used in routine ICU care next to manually measured QTc with a 12-lead ECG.

KEY FINDINGS OF THE THESIS PART III

Subcutaneous continuous glucose measurement in critically ill patients

The aim of the randomized controlled trial (RCT) in *chapter 8* was to assess the safety, efficacy, workload and costs of a subcutaneous continuous glucose measurement (CGM) system-guided blood glucose regulation in comparison with frequent point of care (POC) blood glucose-guided regulation in a mixed population of critically ill patients.
A total of 178 adult critically ill patients with an expected stay of more than 24 hours and in need of insulin therapy were randomized. In all study participants, blood glucose regulation was performed by a sliding scale algorithm with a blood glucose target of 5.0 to 9.0 mmol/L, which was integrated into the patient data management system (PDMS, Meta- Vision; iMDsoft, Tel Aviv, Israel). All participants received a subcutaneous CGM device. The CGM data in the intervention group were used to feed the glucose algorithm, whereas the data of the

intermittent POC glucose measurements in the control group were used in the same algorithm. The CGM data were blinded for the nurses and physicians in the control group. Safety was assessed with the incidence of severe hypoglycemia (<2.2 mmol/L) and efficacy with the percentage time in target range (5.0 to 9.0 mmol/L). In addition, we assessed nursing workload and costs. Due to one incorrect randomization and 21 CGM device failures, a total of 156 patients were included in the intention to treat analysis. The incidence of severe hypoglycemia (< 2.2 mmol/L) verified by arterial blood gas analysis (BGA) was zero in both the intervention and the control group. The number of severe hypoglycemia detected by the CGM-device was 4% in the intervention and 5% in the control group, p= 1.0. The percentage time in target range was 69 ± 26 % for the intervention group and 66 ± 26% for the control group (p= 0.47). A significant reduction in daily nursing workload for glucose control was found in the intervention group (17 versus 36 minutes; p= 0.001). Mean daily costs per patient were significantly reduced by €12 (95% CI -32 to -18, p= 0.02) in the intervention group. The accuracy of the CGM device was assessed by one third of the samples and presented with the median absolute relative difference (Median ARD). The median ARD was significantly lower than the POC device (13.7% versus 7.1%, p= 0.001).

To gain more insight into the efficacy and accuracy, we assessed the current evidence concerning the clinical benefits and accuracy of CGM devices in critically ill patients (*chapter 9*). A total of 37 studies were included in this scoping review, of which five RCTs assessed the efficacy of subcutaneous CGM devices. Generally, CGM devices do not appear to improve glycemic control in a clinically significant manner, although larger sample sizes may lead to better results. A study of 24 mainly neurosurgical patients showed improved glycemic control without hypoglycemic events, with a fully automated closed-loop system and an adapted glucose algorithm [57]. However, the investigators used different glucose target ranges in the intervention and control group (intervention 108 to 144 mg/dL versus control 126 to 180 mg/dL).

The accuracy was assessed in all studies of which 59% used a subcutaneous CGM device and 31% used an intravascular device. A total of 12 (38%) studies included a mixed ICU population, whereas 9 (28%) studies included only cardiac surgery patients. The majority of the studies (66%) used an arterial sample as a reference measurement, which is considered as the golden standard. The median number of paired samples was 672 with an interquartile range (IQR) of [346 - 1028] in 8 to 156 critically ill patients.

Accuracy is established on the comparison between CGM device glucoses and simultaneously obtained reference glucoses. Currently, there is no clear consensus for determining and reporting CGM accuracy. Consequently, studies used a wide variety of methods to assess the accuracy of CGM devices, making it difficult to compare the results. Recent expert consensus recommendations state that a mean of median ARD <14% is acceptable and that 98% of the readings >100 mg/dL should be within 12.5% and the remaining 2% within 20% of the reference glucose [58]. Furthermore, clinical accuracy assessed with Clark error grid analysis (CEG) requires that at least 95% of the paired values should be in zone A, a maximum of 5% can be in zone B, and no values can be in zones C, D and E [59]. None of the included studies

in *chapter 9* fulfill all these recommendations. Bland-Altman analyses showed an overall low mean bias, indicating a low systematic error, however wide limits of agreement (LOA), indicating a high random error. Generally, the intravascular CGM devices seem to have a better accuracy than the subcutaneous device. However, 80% of the studies with intravascular devices included only elective (cardiac surgery) patients which impede the external validity of the results.

Five studies sought to identify factors influencing the accuracy of CGM devices. The accuracy was worse in patients with diabetes, vasopressors treatment, higher SOFA scores, glucose levels in the hypoglycemic range and glycemic variability [60-62].

None of the studies with subcutaneous CGM devices reported severe adverse events. Adverse events consisted of local bleeding, redness or bruises, whereas the use of intravascular devices led to (device related) thrombosis.

We conducted a post-hoc analysis (*chapter 10*) of our complete dataset, since our scoping review has shown that previous studies are relatively small and the effect of severity of illness, diabetes, medication, hypoglycemia and rapid glucose changes on accuracy is still not clear. The primary aim of this retrospective analysis in *chapter 10* was to assess the accuracy and reliability of a subcutaneous CGM device in critically ill patients. The secondary aim was to establish patient related factors influencing the accuracy and reliability. We analyzed 2,840 paired measurements from 155 critically ill patients, with a median of 27.0 [20.0 - 32.0] paired values per patient. The time between the CGM (index test) and BGA values (reference test) had a median of 2.0 [1.0 - 4.0] minutes. A total of 2,788 (98.2%) values were assessed within 0 - 5 minutes from each other.

The overall median ARD was 13.3 [6.9 - 22.1]% and 1,166 (41.1%) single paired values had a median ARD ≥ 14%. The median ARD aggregated per patient was 13.4% and ≥14% in 71 (46%) patients. The median ARD was significantly higher in both the hypoglycemic (n= 25 samples) and the hyperglycemic range (n= 376 samples), 32.4 [12.1 - 53.4]% and 18.7 [10.7 - 28.3]% respectively, p= 0.001. The Bland-Altman analysis showed a mean bias of -0.82 mmol/L with a LOA of -3.88 mmol/L and an upper LOA of 2.24 mmol/L. A total of 1,626 (57.3%) values met the ISO-2013 standards and 1,334 (47%) CGM values were within 12.5% from the reference value. The CEG analysis showed a total of 71% of the values in zone A and 28% in zone B.

The reliability of the CGM device was assessed according to the recommendations of Finfer et al. The CGM device should continuously measure and display glucose in real time > 95% of the time. Skips in data acquisition due to device failure should not be more than 30 minutes at a time [58]. The original RCT included 178 patients of which 21 (12%) patients had a CGM device failure and therefore no measurements. Five device failures could be traced to user errors: one empty battery, one incorrect sensor insertion and three other protocol/device violations. Two device failures were caused by CGM sensor failures and the other 14 device failures were unclear This reliability analysis showed a median real-time data display of 100 [94.5 - 100.0]%) of the total CGM monitoring time. In 73% of analyzed patients the CGM device showed an overall real-time data display ≥95% of the total monitoring time. The CGM device measured without any interruptions in 81 (52%) patients.

Patient characteristics which could potentially influence CGM device accuracy were univariably tested between patients with a median ARD <14% versus median ARD ≥14% and CGM measured ≤95% versus >95%. Glucose variability measured with both CGM and BGA was significantly lower in the median ARD <14% group. There were no significant differences in patient characteristics of patients in which the CGM device measured above or below 95% of the time.

STRENGTHS AND LIMITATIONS PART III

Subcutaneous continuous glucose measurement in critically ill patients

Efficacy

A strength of our RCT in *chapter 8* is the relatively large sample size and the wide variety in case mix. However, some limitations of this study require further consideration. First, the study was conducted in a single Dutch ICU, which impedes the generalizability of the results. Second, the ICU nurses did not verify the severe hypoglycemia that was showed by CGM (intervention group) in two of the three patients despite specific instructions to do so. In addition, hypoglycemic events are barely present due to the successful strict glucose control at our ICU (for example, 18 events out of 56,324 glucose measurements in 2016). Thus, the available data are insufficient to define the accuracy of the CGM in the hypoglycemic range. Third, the computerized algorithm was designed for intermittent glucose measurements and not for continuous glucose measurements. As a consequence, the intervention group most likely did not fully benefit from the more frequent glucose measurements by CGM. An algorithm based on 10 - 15-minute glucose input might have resulted in an improved time in target range. Finally, the accuracy of the CGM device in *chapter 8* was assessed by only one third of the available data and was significantly lower than the POC device (median ARD 13.7% versus median ARD 7.1%, p= 0.001). Consequently, the accuracy of the device was not adequately assessed, and measurement errors could have occurred. However, adverse events did not take place.

Accuracy

Chapter 10 is the largest study on accuracy CGM in severe critically ill patients so far. Due to the considerable number of paired samples, we could assess the accuracy of the CGM device in glucose values in the hypoglycemic and hyperglycemic range. This is in contrast to several previous studies that had fewer values in the hypoglycemic range. Another strength is that we not only performed an analysis with the original target ranges of Clarke and coworkers, but also a modified analysis with our own ICU target ranges. Previous studies did not modify the CEG ranges to the glucose target ranges and treatment goals of their ICU, although this was recommended by Clarke et al. [59]. Our analyses showed that the CEG analysis with the original target range of 3.9 to 10.0 mmol/L led to a slight overestimation of the accuracy in the clinical acceptable zones A and B compared to our own ICU target ranges. Additionally, the original CEG analysis is not developed for use in critically ill patients and continuous glucose

measurement [63]. Continuous glucose-error grid analysis has important shortcomings since it requires very frequent reference measurements and is difficult to interpret [64].
A considerable limitation is that this study was not primarily designed for an analysis of the accuracy and reliability. This could have resulted in a reduced number of paired samples in individual patients. Another consequence of the retrospective design is the lack of additional information about the calibration times.

CLINICAL IMPLICATIONS AND FUTURE RESEARCH PART III

Subcutaneous continuous glucose measurement in critically ill patients

Despite the shortcomings in accuracy, CGM is a promising technique since 4 to 15% of hypoglycemic events are undetected [65] and hypoglycemic events occur more frequently when there is a longer time gap between glucose measurements [66]. Moreover, the management of strict glycemic control patients requires frequent glucose monitoring which comes with its associated financial expenses [67;68]. The results of the RCT in *chapter 8* suggest that CGM is as safe and effective as intermittent point-of-care measurements and reduces both nursing workload and daily costs. To make full use of the continuous measurements, a new computerized algorithm should be designed for (semi) continuous glucose measurements. It can be expected that an adapted algorithm will further improve the performance of CGM in the guidance of glycemic control. However, the current evidence on the use of CGM devices in critically ill patients is still scarce and more research is needed.
The accuracy of CGM devices needs to be improved and thoroughly assessed before CGM can be a fully alternative to the current POC devices. Furthermore, the consensus recommendations should be evaluated. The current recommendations do not differentiate between pooled median/mean ARD of all paired data points and individual patients' median/mean ARDs. By requiring a pooled median/mean ARD to be below a certain cut-off, patients with substantially higher values may go undetected. It is advisable to set a requirement for the dispersion around the average median/mean ARD as well. Additionally, it is debatable whether the ISO-2013 and Finfer accuracy standards are fully suitable for an intensive care setting. During ICU admission several reference glucose measurements with BGA are performed, as opposed to CGM in the outpatient setting (home) where no reference measurement is available. Consequently, potential measurement errors will be detected earlier and may not even lead to adverse events.
A relatively large number of devices (12%) experienced a failure which resulted in no measurements. Despite the training of the nurses and information provided in advance, user errors have occurred. The high failure rate limit the routine use in clinical practice.
The accuracy of the CGM device in our analysis was significantly worse in glucose (reference) values in the hypoglycemic and hyperglycemic range or in patients with higher glucose variability. Hyperglycemia and a relatively high glucose variability are more likely to occur at the onset of the intensive insulin treatment during the first hours of the ICU admission. When using a CGM device for glycemic control, additional reference measurements are required.

Future research should focus on the assessment of the CGM device accuracy in the hypoglycemic and hyperglycemic ranges.
When CGM accuracy is improved in all glycemic ranges and accuracy metrics are standardized, the ultimate goal would be the development of a closed-loop glucose control system in critically ill patients.

In conclusion, subcutaneous CGM to guide insulin treatment in critically ill patients is minimally invasive, in contrast to intravascular CGM devices. None of the studies with subcutaneous CGM devices reported severe adverse events. CGM has potential benefits, such as improvement of time in target, reduction of glycemic variability and less staff workload and costs. However, the current evidence on the use of CGM devices in critically ill patients is not sufficient and more research is needed, especially in a mixed ICU population. In addition, the accuracy of subcutaneous CGM devices seems adequate to guide alarms, but they require improvement when using for strict glucose control. Finally, the accuracy metrics must be standardized and adapted for the use of CGM in critically ill patients.

KEY FINDINGS OF THE THESIS PART IV

Evidence-based practice in intensive care medicine

The ICU is an environment where a safe and effective treatment is highly important since critically ill patients are at high risk for complications, harm and death. In order to deliver this complex medical treatment, critical care physicians and nurses must be well aware of current clinical practice guidelines. Additionally, they have to be able to critically assess scientific research and choose the best available treatment or diagnostic procedure/tool in their clinical practice for the individual patient [69]. Evidence-based practice (EBP) is an approach to optimize clinical decision-making by using the best available scientific evidence in combination with a clinician's expertise and patient preferences [70].The application of EBP requires, among organizational facilitators, a positive attitude towards EBP, skills and knowledge of the methods of clinical epidemiology [71]. Several questionnaires are available to measure these traits, of which the McColl questionnaire is the most widely used in previous research among physicians and nurses [72]. This questionnaire measures Attitudes Towards EBP; the ability to access scientific literature; self-rated knowledge of EBP journals, websites, and terms used in EBP. The McColl questionnaire has originally been developed for general practitioners [73] and none of the previous studies sought to assess the reliability and validity of the McColl questionnaire and the subscales among nurses [55;74;75]. Moreover, little is known about the validity and reliability of the Dutch version of the McColl questionnaire. Therefore, the aim of the cross-sectional study in *chapter 11* was to assess the construct validity and reliability of the Dutch version of the McColl questionnaire among nurses in a teaching hospital.
The original McColl questionnaire was translated into Dutch through the two-panel approach. This method potentially leads to conceptually equivalent versions practically performing in the same way as the original [76]. A total of 198 registered nurses (vocational or bachelor degree), student nurses, specialized nurses, advanced nurse practitioners, nurses with a

master's degree and nurse managers were invited to participate in the study of which 51 (27%) completed the questionnaire.

Internal consistency reliability (Cronbach's coefficient alpha) of the 'Attitudes Towards EBP', 'Understanding Journals/Websites' and 'Understanding EBP Terms' were 0.45 (insufficient), 0.69 (almost acceptable) and 0.89 (good), respectively. Combining the subscales 'Understanding Journals/Websites' and 'Understanding EBP Terms' resulted in a Cronbach's alpha of 0.89.

The construct validation of the presumed subscales was assessed by the known-groups validation method [77]. The hypothesis of this study was that higher educated nurses (nurses with a bachelor's degree, a post-Baccalaureate degree and nurses with a master's degree) would have higher item and subscale scores than vocational nurses. Furthermore, it was hypothesized that scientific database use (never versus ≥1 time) and any EBP training (yes/no) would result in higher item and subscale scores [78-81]. The subscales 'Understanding EBP Terms' and 'Understanding Journals/Websites' showed a satisfactory construct validity with significantly higher sum scores for higher educated nurses and nurses who received EBP training.

STRENGTHS AND LIMITATIONS PART IV

Evidence-based practice in intensive care medicine

A strength of the study in *chapter 11* is the translation through to the two panel approach that involved nursing professionals. This resulted in a Dutch version conveying the original meaning of the questions, which might be better suited for international comparisons. Another strength is the assessment of the internal consistency reliability and construct validity, since only a few studies reported both validity and reliability [72]. A limitation of this study was the limited sample size (n= 51) Consequently, we could not use factor analysis to explore the proposed dimensionality of the instrument. Inherent to surveys where respondents both self-select and self-report the items, social desirability bias may have occurred. The response rate of 27% may also have contributed to sampling bias, and we had no data of the non-respondents. Therefore, we could not check the representativeness of the sample. Nevertheless, response rates to online surveys are generally around 30% and the respondents represented a diverse sample of registered nurses in a teaching hospital [82].

CLINICAL IMPLICATIONS AND FUTURE RESEARCH PART IV

Evidence-based practice in intensive care medicine

This study shows that the attitude items of the McColl Questionnaire are neither valid nor reliable to assess attitudes toward EBP among nurses in a teaching hospital. These items should be thoroughly revised and validated. An alternative for the assessment of attitudes toward EBP is the Evidence-Based Practice Attitude Scale, which has promising psychometric properties [83]. Another finding from this study is that the combined knowledge subscale (Journals/Websites and EBP Terms) is both reliable and valid for the assessment of EBP knowledge among nurses. Logically, the journals and websites ought to be adapted according

to the country in which the survey is conducted [84]. In addition, future studies should establish the psychometric properties of the knowledge scale more extensively, since results can differ between hospitals, level of education and countries. Finally, valid and reliable tools, which evaluate EBP attitudes, skills and knowledge could be combined with valid and reliable tools that measure actual EBP behavior and barriers to the use of EBP [72]. Such an instrument is essential to enable the strategy and evaluation of a custom-made EBP implementation program in healthcare institutions [71].
Although previous research showed positive attitudes towards EBP among many healthcare professionals [71], EBP still suffers from substantial resistance from others. A commonly heard accusation is that EBP ignores clinical expertise, clinical judgment and clinician-patient interaction [85;86]. As clinicians apply EBP, as originally defined by D. Sacket , it is about "integrating individual clinical expertise with the best external evidence" [70]. Thus "individual expertise is as important as external evidence" [86]. This recent and substantial resistance towards EBP demonstrates the necessity to structurally assess and influence attitudes towards EBP.

In conclusion, the subscale Attitudes Towards EBP of the Dutch McColl questionnaire among nurses showed a poor reliability and construct validity. The adapted McColl sub scales 'Understanding Journals/Websites' and EBP terms can be merged into one scale, which showed good psychometric properties. This knowledge tool could be combined into a questionnaire, which assesses barriers and Attitudes Towards EBP and EBP utilization. These combined instruments should be validated in more detail in larger samples for further validity and reliability testing.

GENERAL CONCLUSION

This thesis attempted to assess the psychometric properties of a number of measurements devices and tools that are used in intensive care medicine and nursing. It is important to realize that measurement tools such as questionnaires need to be officially translated and validated into the language in which the tool will be used. In addition, the psychometric properties of a measurement tool or device must be assessed in the (patient) population in which it will be applied. This thesis is a contribution towards the process of assessing the psychometric properties of several measurement tools/devices. Therefore, additional research needs to be conducted to gain further insight into the validation of measurement scales: clinimetric research is an ongoing process.
Critical care physicians and nurses need to be aware of the basic principles of clinimetrics, in order to be able to deliver complex, effective and high quality medical and nursing treatment. An approach to apply clinimetrics in clinical practice is the use of EBP, which is an integration of the best current evidence, clinical expertise and patients' preferences. Despite EBP being embraced by many healthcare professionals, some reservations remain. In additions its implementation is still challenging and unsatisfactory since many barriers, such as a lack of time and skills, persist. It is important to stress that a positive attitude towards EBP is crucial to

ensure a successful implementation in daily clinical practice. Currently in 2017, EBP is suffering from substantial resistance from clinicians and policy makers. This demonstrates the necessity to structurally assess and influence attitudes towards EBP.

REFERENCES

1. Lee KM, Lee J, Chung CY, Ahn S, Sung KH, Kim TW, et al. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research Clin Orthop Surg 2012 Jun;4(2):149-55.
2. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-8.
3. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006 Oct;59(10):1033-9.
4. Gelinas C, Puntillo KA, Joffe AM, Barr J. A validated approach to evaluating psychometric properties of pain assessment tools for use in nonverbal critically ill adults. Seminars in Respiratory and Critical Care Medicine 2013 Apr;34(2):153-68.
5. Bouter LM, van Dongen MCJM, Zielhuis GA. In:Diagnostiek en prognostiek. Epidemiologisch onderzoek Opzet en interpretatie. 5de herziene druk ed. Houten: Bohn Stafleu van Loghum; 2005. p. 245-95.
6. Ahlers SJ, van der Veen AM, van Dijk M, Tibboel D, Knibbe CA. The use of the Behavioral Pain Scale to assess pain in conscious sedated patients. Anesth Analg 2010 Jan 1;110(1):127-33.
7. Aissaoui Y, Zeggwagh AA, Zekraoui A, Abidi K, Abouqal R. Validation of a behavioral pain scale in critically ill, sedated, and mechanically ventilated patients. Anesth Analg 2005 Nov;101(5):1470-6.
8. Buttes P, Keal G, Cronin SN, Stocks L, Stout C. Validation of the critical-care pain observation tool in adult critically ill patients. Dimens Crit Care Nurs 2014 Mar;33(2):78-81.
9. Chanques G, Pohlman A, Kress JP, Molinari N, de JA, Jaber S, et al. Psychometric comparison of three behavioural scales for the assessment of pain in critically ill patients unable to self-report. Crit Care 2014;18(5):R160.
10. Chen J, Lu Q, Wu XY, An YZ, Zhan YC, Zhang HY. Reliability and validity of the Chinese version of the behavioral pain scale in intubated and non-intubated critically ill patients: Two cross-sectional studies. Int J Nurs Stud 2016 May 27;61:63-71.
11. Echegaray-Benites C, Kapoustina O, Gelinas C. Validation of the use of the Critical-Care Pain Observation Tool (CPOT) with brain surgery patients in the neurosurgical intensive care unit. Intensive Crit Care Nurs 2014 Oct;30(5):257-65.
12. Gelinas C, Johnston C. Pain assessment in the critically ill ventilated adult: validation of the Critical-Care Pain Observation Tool and physiologic indicators. Clin J Pain 2007 Jul;23(6):497-505.
13. Li Q, Wan X, Gu C, Yu Y, Huang W, Li S, et al. Pain assessment using the critical-care pain observation tool in Chinese critically ill ventilated adults. J Pain Symptom Manage 2014 Nov;48(5):975-82.
14. Nurnberg DD, Saboonchi F, Sackey PV, Bjorling G. A preliminary validation of the Swedish version of the Critical-Care Pain Observation Tool in adults. Acta Anaesthesiol Scand 2011 Apr;55(4):379-86.
15. Payen JF, Bru O, Bosson JL, Lagrasta A, Novel E, Deschaux I, et al. Assessing pain in critically ill sedated patients by using a behavioral pain scale. Crit Care Med 2001 Dec;29(12):2258-63.
16. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and psychological measurement; 1973 Oct 1.
17. Altman DG. Some common problems in medical research. In: Practical statistics for medical research. New York: Chapman & Hall/CRC; 1991. p. 396-435.
18. Juarez P, Bach A, Baker M, Duey D, Durkin S, Gulczynski B, et al. Comparison of two pain scales for the assessment of pain in the ventilated adult patient. Dimens Crit Care Nurs 2010 Nov;29(6):307-15.
19. Kanji S, MacPhee H, Singh A, Johanson C, Fairbairn J, Lloyd T, et al. Validation of the Critical Care Pain Observation Tool in Critically Ill Patients With Delirium: A Prospective Cohort Study. Crit Care Med 2016 Jan 16.
20. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability. Measurement in Medicine.Cambridge: Cambridge university press; 2011. p. 96-145.

21. Morete MC, Mofatto SC, Pereira CA, Silva AP, Odierna MT. Translation and cultural adaptation of the Brazilian Portuguese version of the Behavioral Pain Scale. Rev Bras Ter Intensiva 2014 Oct;26(4):373-8.
22. Ahlers SJ, van Gulik L, van der Veen AM, van Dongen HP, Bruins P, Belitser SV, et al. Comparison of different pain scoring systems in critically ill patients in a general ICU. Crit Care 2008;12(1):R15.
23. Vazquez M, Pardavila MI, Lucia M, Aguado Y, Margall MA, Asiain MC. Pain assessment in turning procedures for patients with invasive mechanical ventilation. Nurs Crit Care 2011 Jul;16(4):178-85.
24. de Vet HC. Terwee CB, Mokking LB, Knoll DL. Field-testing: item reduction and data structure. In: Measurement in Medicine.Cambridge: Cambridge university press; 2011. p. 65-84.
25. Streiner D.L, Norman G.R. Health measurement scales a practical guide to their development and use. third edition ed. Hamilton: Oxford university press; 2003.
26. Paulson-Conger M, Leske J, Maidl C, Hanson A, Dziadulewicz L. Comparison of two pain assessment tools in nonverbal critical care patients. Pain Manag Nurs 2011 Dec;12(4):218-24.
27. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Springer Netherlands; 2017 Jun 7.
28. Gelinas C, Harel F, Fillion L, Puntillo KA, Johnston CC. Sensitivity and specificity of the critical-care pain observation tool for the detection of pain in intubated adults after cardiac surgery. J Pain Symptom Manage 2009 Jan;37(1):58-67.
29. Joffe AM, McNulty B, Boitor M, Marsh R, Gelinas C. Validation of the Critical-Care Pain Observation Tool in brain-injured critically ill adults. J Crit Care 2016 Dec;36:76-80.
30. Severgnini P, Pelosi P, Contino E, Serafinelli E, Novario R, Chiaranda M. Accuracy of Critical Care Pain Observation Tool and Behavioral Pain Scale to assess pain in critically ill conscious and unconscious patients: prospective, observational study. J Intensive Care 2016;4:68.
31. Gelinas C. Pain assessment in the critically ill adult: Recent evidence and new trends. Intensive Crit Care Nurs 2016 Jun;34:1-11.
32. Baron R, Binder A, Biniek R, Braune S, Buerkle H, Dall P, et al. Evidence and consensus based guideline for the management of delirium, analgesia, and sedation in intensive care medicine. Revision 2015 (DAS-Guideline 2015) - short version. Ger Med Sci 2015;13:Doc19.
33. de Vet HC. Terwee CB, Mokking LB, Knoll DL. Validity. In: Measurement in Medicine.Cambridge: Cambridge university press; 2011. p. 150-99.
34. Streiner D L, Norman G R. Health measurement scales a practical guide to their development and use. third ed. New York: Oxford University Press; 2003. p. 61-79.
35. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Field-testing: item reduction and data structure.In: Measurement in Medicine. first ed. Cambridge: Cambridge university press; 2011. p. 65-95.
36. MacCallum RC, Widaman KF, Preacher KJ, Hong S. Sample Size in Factor Analysis: The Role of Model Error. Multivariate Behav Res 2001 Oct 1;36(4):611-37.
37. Streiner D L, Norman G R. Reliability. In: Health measurement scales a practical guide to their development and use. 3 ed. New York: Oxfor university press; 2003. p. 126-52.
38. Maaskant J, Raymakers-Janssen P, Veldhoen E, Ista E, Lucas C, Vermeulen H. The clinimetric properties of the COMFORT scale: A systematic review. Eur J Pain 2016 Nov;20(10):1587-611.
39. Varndell W, Fry M, Elliott D. A systematic review of observational pain assessment instruments for use with nonverbal intubated critically ill adult patients in the emergency department: an assessment of their suitability and psychometric properties. J Clin Nurs 2017 Jan;26(1-2):7-32.
40. Ouimet S, Kavanagh BP, Gottfried SB, Skrobik Y. Incidence, risk factors and consequences of ICU delirium. Intensive Care Med 2007 Jan;33(1):66-73.
41. van den Boogaard M, Schoonhoven L, van der Hoeven JG, van AT, Pickkers P. Incidence and short-term consequences of delirium in critically ill patients: A prospective observational cohort study. Int J Nurs Stud 2012 Jul;49(7):775-83.
42. Gelinas C. Pain assessment in the critically ill adult: Recent evidence and new trends. Intensive Crit Care Nurs 2016 Jun;34:1-11.
43. Vink P, Lucas C, Maaskant JM, van Erp WS, Lindeboom R, Vermeulen H. Clinimetric properties of the Nociception Coma Scale (-Revised): A systematic review. Eur J Pain 2017 Jun 2.

44. Chilkoti G, Wadhwa R, Saxena AK. Technological advances in perioperative monitoring: Current concepts and clinical perspectives. J Anaesthesiol Clin Pharmacol 2015 Jan;31(1):14-24.
45. Berger RD, Kasper EK, Baughman KL, Marban E, Calkins H, Tomaselli GF. Beat-to-beat QT interval variability: novel evidence for repolarization lability in ischemic and nonischemic dilated cardiomyopathy. Circulation 1997 Sep 2;96(5):1557-65.
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-74.
47. Rautaharju PM, Surawicz B, Gettes LS, Bailey JJ, Childers R, Deal BJ, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. Circulation 2009 Mar 17;119(10):e241-e250.
48. Statters DJ, Malik M, Ward DE, Camm AJ. QT dispersion: problems of methodology and clinical significance. J Cardiovasc Electrophysiol 1994 Aug;5(8):672-85.
49. Surawicz B. Will QT dispersion play a role in clinical decision-making?. J Cardiovasc Electrophysiol 1996 Aug;7(8):777-84.
50. Pickham D, Helfenbein E, Shinn JA, Chan G, Funk M, Drew BJ. How many patients need QT interval monitoring in critical care units? Preliminary report of the QT in Practice study. J Electrocardiol 2010 Nov;43(6):572-6.
51. Al-Khatib SM, LaPointe NM, Kramer JM, Califf RM. What clinicians should know about the QT interval. JAMA 2003 Apr 23;289(16):2120-7.
52. Drew BJ, Ackerman MJ, Funk M, Gibler WB, Kligfield P, Menon V, et al. Prevention of torsade de pointes in hospital settings: a scientific statement from the American Heart Association and the American College of Cardiology Foundation. Circulation 2010 Mar 2;121(8):1047-60.
53. Molnar J, Zhang F, Weiss J, Ehlert FA, Rosenthal JE. Diurnal pattern of QTc interval: how long is prolonged? Possible relation to circadian triggers of cardiovascular events. J Am Coll Cardiol 1996 Jan;27(1):76-83.
54. Roden DM. Drug-induced prolongation of the QT interval. N Engl J Med 2004 Mar 4;350(10):1013-22.
55. Kozik TM, Wung SF. Acquired long QT syndrome: frequency, onset, and risk factors in intensive care patients. Crit Care Nurse 2012 Oct;32(5):32-41.
56. Pickham D, Helfenbein E, Shinn JA, Chan G, Funk M, Weinacker A, et al. High prevalence of corrected QT interval prolongation in acutely ill patients is associated with mortality: results of the QT in Practice (QTIP) Study. Crit Care Med 2012 Feb;40(2):394-9.
57. Leelarathna L, English SW, Thabit H, Caldwell K, Allen JM, Kumareswaran K, et al. Feasibility of fully automated closed-loop glucose control using continuous subcutaneous glucose measurements in critical illness: a randomized controlled trial. Crit Care 2013 Jul 24;17(4):R159.
58. Finfer S, Wernerman J, Preiser JC, Cass T, Desaive T, Hovorka R, et al. Clinical review: Consensus recommendations on measurement of blood glucose and reporting glycemic control in critically ill adults. Crit Care 2013 Jun 14;17(3):229.
59. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987 Sep;10(5):622-8.
60. Siegelaar SE, Barwari T, Hermanides J, van der Voort PH, Hoekstra JB, DeVries JH. Microcirculation and its relation to continuous subcutaneous glucose sensor accuracy in cardiac surgery patients in the intensive care unit. J Thorac Cardiovasc Surg 2013 Nov;146(5):1283-9.
61. van Hooijdonk RT, Leopold JH, Winters T, Binnekade JM, Juffermans NP, Horn J, et al. Point accuracy and reliability of an interstitial continuous glucose-monitoring device in critically ill patients: a prospective study. Crit Care 2015 Feb 5;19:34.
62. Wollersheim T, Engelhardt LJ, Pachulla J, Moergeli R, Koch S, Spies C, et al. Accuracy, reliability, feasibility and nurse acceptance of a subcutaneous continuous glucose management system in critically ill patients: a prospective clinical trial. Ann Intensive Care 2016 Dec;6(1):70.

63. Clarke WL, Anderson S, Farhy L, Breton M, Gonder-Frederick L, Cox D, et al. Evaluating the clinical accuracy of two continuous glucose sensors using continuous glucose-error grid analysis. Diabetes Care 2005 Oct;28(10):2412-7.
64. Wentholt IM, Hoekstra JB, DeVries JH. A critical appraisal of the continuous glucose-error grid analysis. Diabetes Care 2006 Aug;29(8):1805-11.
65. Boyd JC, Bruns DE. Effects of measurement frequency on analytical quality required for glucose measurements in intensive care units: assessments by simulation models. Clin Chem 2014 Apr;60(4):644-50.
66. Juneja R, Roudebush CP, Nasraway SA, Golas AA, Jacobi J, Carroll J, et al. Computerized intensive insulin dosing can mitigate hypoglycemia and achieve tight glycemic control when glucose measurement is performed frequently and on time. Crit Care 2009;13(5):R163.
67. Aragon D. Evaluation of nursing work effort and perceptions about blood glucose testing in tight glycemic control. Am J Crit Care 2006 Jul;15(4):370-7.
68. Gartemann J, Caffrey E, Hadker N, Crean S, Creed GM, Rausch C. Nurse workload in implementing a tight glycaemic control protocol in a UK hospital: a pilot time-in-motion study. Nurs Crit Care 2012 Nov;17(6):279-84.
69. Phillips C. Relationships between duration of practice, educational level, and perception of barriers to implement evidence-based practice among critical care nurses. Int J Evid Based Healthc 2015 Dec;13(4):224-32.
70. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-2.
71. Ubbink DT, Guyatt GH, Vermeulen H. Framework of policy recommendations for implementation of evidence-based practice: a systematic scoping review. BMJ Open 2013 Jan 24;3(1).
72. Oude RK, Zwolsman SE, Ubbink DT, Mol BW, van DN, Vermeulen H. Tools to assess evidence-based practice behaviour among healthcare professionals. Evid Based Med 2013 Aug;18(4):129-38.
73. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. BMJ 1998 Jan 31;316(7128):361-5.
74. Maaskant JM, Knops AM, Ubbink DT, Vermeulen H. Evidence-based practice: a survey among pediatric nurses and pediatricians. J Pediatr Nurs 2013 Apr;28(2):150-7.
75. Ubbink DT, Vermeulen H, Knops AM, Legemate DA, Oude RK, Heineman MJ, et al. Implementation of evidence-based practice: outside the box, throughout the hospital. Neth J Med 2011 Feb;69(2):87-94.
76. Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. Value Health 2004 Sep;7 Suppl 1:S27-S30.
77. Hattie J, Cooksey W. Procedures for assessing the validities of tests using the "known-groups" method. Applied psychological measurement 1984;8(3):295-305.
78. Heydari A, Mazlom SR, Ranjbar H, Scurlock-Evans L. A study of Iranian nurses' and midwives' knowledge, attitudes, and implementation of evidence-based practice: the time for change has arrived. Worldviews Evid Based Nurs 2014 Oct;11(5):325-31.
79. Rose BL, Aquila AM, Bartos S, McCurry T, Cunningham CE, Lane T, et al. A Cross-sectional Study on Evidence-Based Nursing Practice in the Contemporary Hospital Setting: Implications for Nurses in Professional Development. J Nurses Prof Dev 2017 Mar;33(2):64-9.
80. Saunders H, Vehvilainen-Julkunen K. Nurses' Evidence-Based Practice Beliefs and the Role of Evidence-Based Practice Mentors at University Hospitals in Finland. Worldviews Evid Based Nurs 2017 Feb;14(1):35-45.
81. Warren JI, McLaughlin M, Bardsley J, Eich J, Esche CA, Kropkowski L, et al. The Strengths and Challenges of Implementing EBP in Healthcare Systems. Worldviews Evid Based Nurs 2016 Feb;13(1):15-24.
82. Sheehan K B. E-mail survey response rates: a review. Journal of computer-mediated communication 2001;6(2).
83. van Sonsbeek MA, Hutschemaekers GJ, Veerman JW, Kleinjan M, Aarons GA, Tiemens BG. Psychometric properties of the Dutch version of the Evidence-Based Practice Attitude Scale (EBPAS). Health Res Policy Syst 2015 Nov 16;13:69.

84. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-74.
85. Zonder context geen bewijs. Over de illusie van evidence-based practice in de zorg. Den Haag: Raad voor Volksgezondheid en Samenleving; 2017.
86. Ioannidis JPA. Hijacked evidence-based medicine: stay the course and throw the pirates overboard. J Clin Epidemiol 2017 Apr;84:11-3.

APPENDIX

Table 1. Pain scores for both nurses general ICU patients[1]

	BPS n= 68		**CPOT n= 68**
Nurse 1		**Nurse 1**	
RestI	3.0 [3.0 – 3.0]	**RestI**	0.0 [0.0 – 1.0]
Oral care	4.0 [3.0 – 40]	**Oral care**	0.0 [0.0 – 0.0]
p-Value[2]	*0.000*	*p-Value[2]*	*0.007*
RestII	3.0 [3.0 – 3.0]	**RestII**	0.0 [0.0 – 0.0]
Turning	5.0 [4.0 – 6.8]	**Turning**	2.0 [0.0 – 3.0]
p-Value[3]	*0.000*	*p-Value[3]*	*0.000*
Nurse 2		**Nurse 2**	
RestI	3.0 [3.0 – 4.0]	**RestI**	0.0 [0.0 – 0.0]
Oral care	4.0 [3.0 – 4.0]	**Oral care**	0.0 [0.0 – 1.0]
p-Value[2]	*0.001*	*p-Value[2]*	*0.000*
RestII	3.0 [3.0 – 3.0]	**RestII**	0.0 [0.0 – 0.0]
Turning	5.0 [3.3 – 6.0]	**Turning**	2.0 [0.0 – 3.0]
p-Value[3]	*0.000*	*p-Value[3]*	*0.000*

[1] *Chapter 2*

[2] RestI vs. Oral care (Wilcoxon Signed Ranks Test)

[3] RestII vs. Turning (Wilcoxon Signed Ranks Test)

Table 2 Inter-rater reliability chapters 2, 3 and 4

ASSESSMENT	CHAPTER 2 GENERAL ICU ICC [95% CI][1]	CHAPTER 3 CTC[2] ICC [95% CI][1]	CHAPTER 4 MIXED ICU[3] ICC [95% CI][1]
BPS	**n= 68 patients**	**n= 72 patients**	**n= 108 patients**
Overall	**0.74 95% CI [0.68 - 0.79]**	**0.744 [0.687–0.791]**	**-**
Rest I	0.70 95% CI [0.56 - 0.80]	0.239 [0.010 – 0.444]	-
Oral care	0.71 95% CI [0.57 - 0.81]	0.712 [0.577 – 0.810]	-
Rest II	0.80 95% CI [0.70 - 0.88]	0.398 [0.186 – 0.575]	-
Turning	0.60 95% CI [0.42 - 0.73]	0.750 [0.629 – 0.836]	-
CPOT			
Overall	**0.75 95% CI [0.69 - 0.79]**	**0.74 [0.69 – 0.79]**	**0.57 [0.47 – 0.65]**
Rest I	0.72 95% CI [0.58 - 0.81]	0.34 [0.12 – 0.53]	-
Oral care	0.72 95% CI [0.58 - 0.82]	0.83 [0.74 – 0.89]	-
Rest II	0.80 95% CI [0.70 - 0.88]	0.47 [0.26 – 0.63]	0.38 [0.20 – 0.53]
Turning	0.62 95% CI [0.45 - 0.75]	0.62 [0.46 – 0.75]	0.56 [0.42 – 0.68]

[1] Intraclass correlation coefficient One-way random with 95% confidence interval [2]
[2] Post cardiac surgery
[3] Mixed ICU patients with 75% surgery patients and 25% general (medical) ICU patients

Table 3. Bland Altman analyses

	Mean bias	sd[1]	Upper LOA[2]	Lower LOA
General ICU BPS[3]	0.055	1.097	2.205	-2.095
General ICU CPOT	0.048	1.139	2.280	-2.185
CTC BPS[4]	0.038	0.900	1.802	-1.726
CTC CPOT	-0.031	0.923	1.777	-1.840
Mixed ICU CPOT[5]	-0.060	1,085	2.066	-2.187

[1] standard deviation of de difference between the raters
[2] Limit of agreement
[3] General ICU- patients (*chapter 2*)
[4] Patients post-cardiac surgery (*chapter 3*)
[5] General en post-cardiac surgery (*chapter 4*)

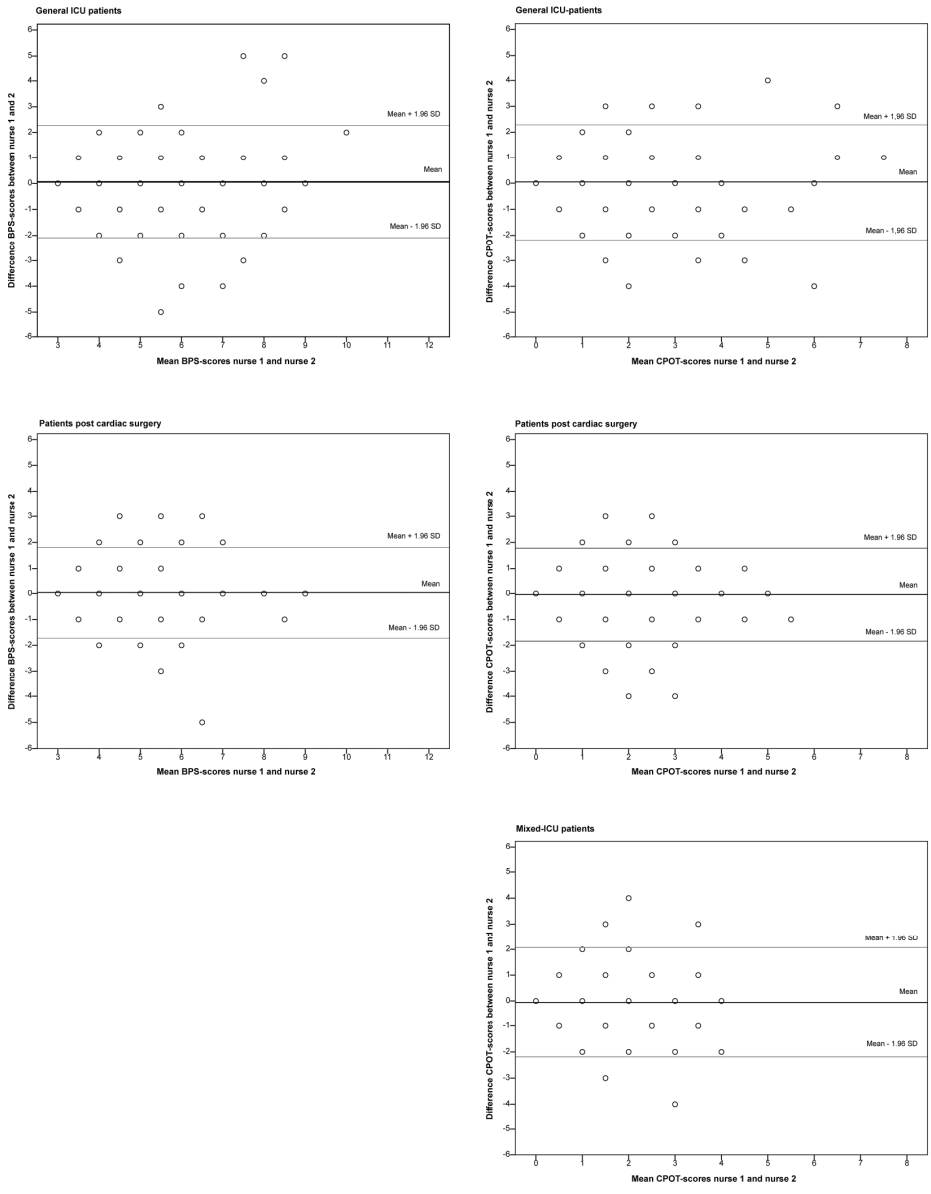


Figure 1. Bland-Altman plots

Bland Altman analyses for CPOT- en BPS-scores between nurse 1 and nurse 2. Each circle represents a pain score assessed by both nurses at the same time; many circles overlap. The limits of agreement showed that 95% of the BPS-scores of the nurses differed up to two points without a systematic difference. All Difference scores (y-axis) are normally distributed and there is no significant correlation between the Mean scores (x-axis) and Difference scores (Y-axis).

Table 4. Standard error of measurement BPS and CPOT

	ICC[1] BPS	BPS Mean ± sd[2]	BPS SEM	ICC1 CPOT	CPOT Mean ± sd	CPOT SEM
General ICU[3]	**0.74^**			**0.75^**		
Nurse 1 (N1)		4.04 ±1.57	**0.80***		1.01 ± 1.63	**0.82***
Nurse 2 (N2)		4.00 ±1,46	**0.74***		0.96 ± 1,56	**0.78***
Difference N1&N2[4]		0.58 ± 0.93	**0.66#**		0.61 ± 0.96	**0.68#**
All scores[5]		4.02 ± 1.51	**0.77***		0.98 ± 1.59	**0.80***
RestI	0.70	3.53±1.10	**0.60***	0.72	0.49 ± 1.15	**0.61***
Oral care	0.71	3.97 ± 1.37	**0.74***	0.72	0.96 ± 1.60	**0.85***
RestII	0.80	3.39 ± 0.94	**0.42***	0.80	0.33 ± 1.01	**0.45***
Turning	0.60	5.20 ± 1.81	**1.14***	0.62	2.13 ± 1.81	**1.12***
CTC[6]	**0.74^**			**0.74^**		
Nurse 1		3.89 ± 1.23	**0.63***		0.77 ± 1.24	**0.63***
Nurse 2		3.83 ± 1.28	**0.65***		0.80 ± 1.33	**0.68***
Difference N1&N2		0.46 ± 0.80	**0.57#**		0.49 ± 0.76	**0.54#**
All scores[4]		3.87 ± 1.25	**0.64***		0.78 ± 1.29	**0.66***
RestI	0.24	3.20 ± 0.50	**0.44***	0.34	0.11 ± 0.50	**0.41***
Oral care	0.71	3.91 ± 1.17	**0.63***	0.83	0.85 ± 1.33	**0.55***
RestII	0.40	3.47 ± 0.89	**0.69***	0.47	0.37 ± 0.86	**0.63***
Turning	0.75	4.91 ± 1.50	**0.75***	0.62	1.81 ± 1.48	**0.91***
Mixed ICU[7]	-			**0.57^**		
Nurse 1		-	**-**		0.79 ± 1.13	**0.82***
Nurse 2		-	**-**		0.85 ± 1.20	**0.87***
Difference N1&N2		-	**-**		0.60 ± 0.90	**0.64#**
All scores		-	**-**		0.82 ± 1.16	**0.84***
RestI	-	-	**-**	-		
Oral care	-	-	**-**	-		
RestII	-	-	**-**	0.38	0.40 ± 0.86	**0.68***
Turning	-	-	**-**	0.56	1.24 ± 1.27	**0.93***

* SEM = SD$\sqrt{}$ (1- ICC) # $SEM_{consistency} = SD_{difference}/\sqrt{2}$ [20]
^ Intraclass correlation coefficient of all assessments (overall)

[1] Intraclass correlation coefficient One-way random [2]

[2] Standard deviation
[3] General (medical) ICU patients (*chapter 2*)

[4] Absolute difference between both nurses (all assessments)
[5] scores all assessment and both nurses

[6] Post-cardiac surgery (*chapter 3*)
[7] Mixed ICU with 75% post surgery patients and 25% general (medical) ICU patients (*chapter 4*)

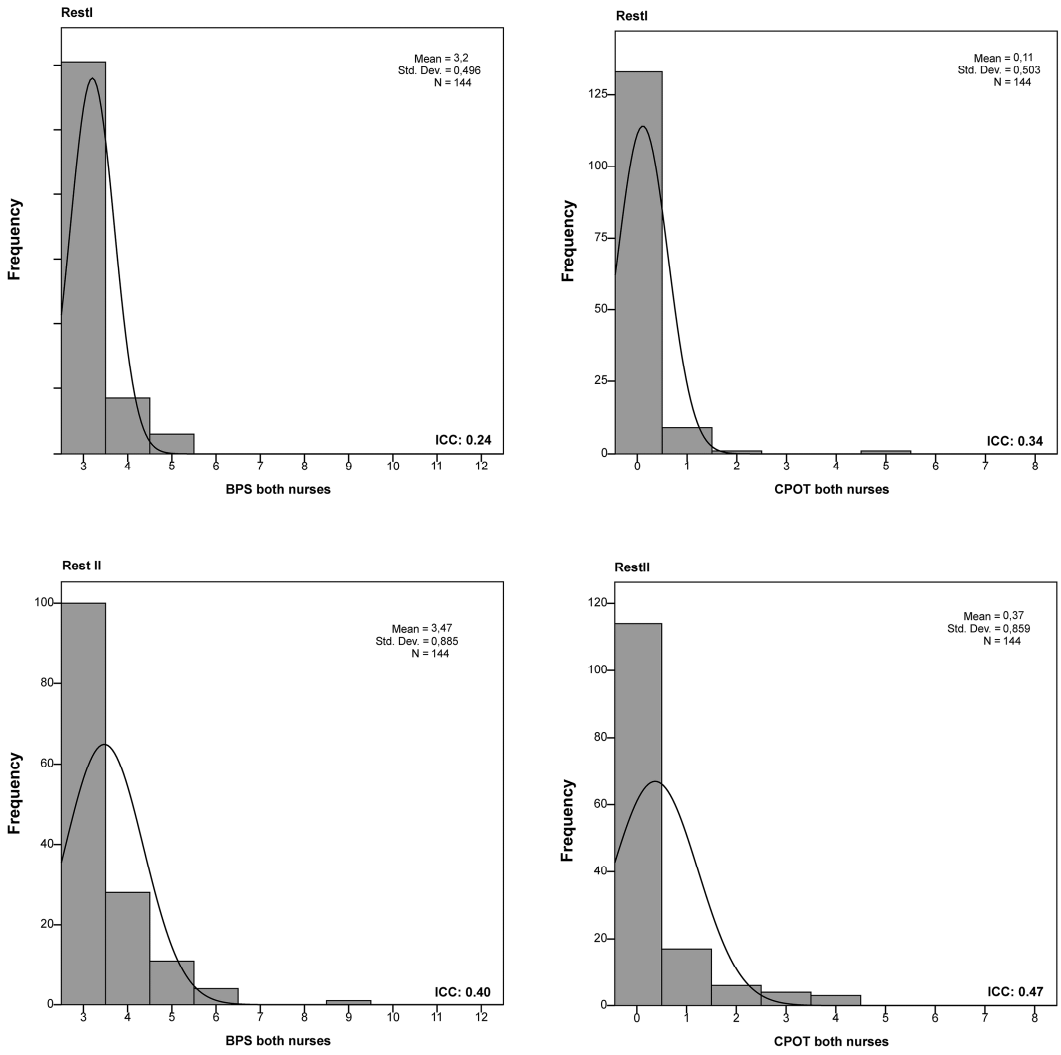


Figure 2. Histograms BPS and CPOT (patients post-cardiac surgery)

Table 5a. Internal consistency BPS and CPOT general ICU patients (*chapter 2*)

Corrected Item-total correlation BPS*			
General ICU BPS domains	**Scale mean if item deleted**	**Corrected item-total correlation**	**Cronbach's alpha if item deleted**
Face	3.15	.498	.664
Upper limps	3.46	.570	.531
Compliance ventilator	3.90	.520	.623

* During turning nurse1

Cronbach's alpha during turning (nurse1): 0.697 Cronbach's alpha during restl (nurse1): 0.810

Corrected Item-total correlation CPOT*			
General ICU CPOT domains	**Scale mean if item deleted**	**Corrected item-total correlation**	**Cronbach's alpha if item deleted**
Face	1.45	.515	.729
Movement	1.60	.536	.712
Muscle tension	1.60	.689	.631
Compliance ventilator	1.81	.500	.731

* During turning nurse1

Cronbach's alpha during turning (nurse1): 0.758 Cronbach's alpha during restl (nurse1): 0.810

Table 5b. Internal consistency BPS and CPOT general ICU patients (*chapter 2*)

Inter-item correlation*			
General ICU BPS domains	**Face**	**Upper limps**	**Compliance ventilator**
Face	1.000	.463	.396
Upper limps	.463	1.000	.508
Compliance ventilator	.396	.508	1.000

* During turning nurse1

Inter-item correlation*				
General ICU CPOT domains	**Face**	**Movement**	**Muscle tension**	**Compliance ventilator**
Face	1.000	.333	.578	.337
Movement	.333	1.000	.534	.432
Muscle tension	.578	.534	1.000	.451
Compliance ventilator	.337	.432	.451	1.000

* During turning nurse1

NEDERLANDSE SAMENVATTING

S. Rijkenberg

NEDERLANDSE SAMENVATTING

Metingen zijn onlosmakelijk verbonden met de behandeling op de intensive care (IC).
Intensivisten en verpleegkundigen gebruiken metingen om een diagnose te stellen, om het effect van de behandeling te evalueren en tenslotte, om een prognose te stellen. Van opname tot ontslag of overlijden; elke fase in de medische behandeling van de kritiek zieke patiënt vereist metingen. Metingen zijn bijvoorbeeld klinisch chemische en hematologisch bepalingen, het meten van glucosewaarden in het bloed met een glucosemeter of het beoordelen van een echo. Maar ook het beoordelen van het hartritme via een elektrocardiogram (ECG of hartfilmpje) of monitor zijn metingen. Een meetinstrument kan ook een score zijn zoals een delierscore of een pijnscore instrument.

Vanwege het belang voor de patiënt is het essentieel dat meetinstrumenten valide en betrouwbaar zijn. Een meetinstrument is valide wanneer het daadwerkelijk meet wat het beoogt te meten (validiteit). Een pijnscore moet bijvoorbeeld pijn meten en geen onrust of een delier. Betrouwbaarheid is de mate waarin patiënten van elkaar kunnen worden onderscheiden, ondanks de meetfout [1]. Meetfouten zijn systematische of toevallige (*random*) fouten van het meetinstrument [1]. Een continue glucosemeter kan bijvoorbeeld systematisch één punt lager meten dan de glucosewaarde in het bloed gemeten met de gouden standaard methode. Systematische meetfouten leiden tot vertekening (*bias*) van de werkelijkheid en tasten de validiteit van het meetinstrument aan. Toevallige meetfouten zijn onvermijdelijk en ontstaan bijvoorbeeld doordat de verpleegkundigen verschillen in expertise met het meetinstrument. Toevallige fouten hebben een negatieve invloed op de reproduceerbaarheid (precisie) van het meetinstrument: de mate waarin het instrument steeds dezelfde resultaten bij dezelfde patiënt onder dezelfde condities geeft [2].

Er zijn veel meetinstrumenten voor IC-patiënten beschikbaar, maar lang niet alle meetinstrumenten hebben goede meeteigenschappen en zijn goed onderzocht. Dus hoe beoordeelt een intensivist of IC-verpleegkundige dan of een meetinstrument valide en betrouwbaar is?

De klinische epidemiologie heeft een antwoord op deze vraag. Klinische epidemiologie is een wetenschap die ziekte en genezing beschrijft en verklaart middels epidemiologische technieken [3]. Een methode om dit in de klinische praktijk toe te passen is *Evidence-Based Practice* (EBP). EBP helpt om behandelingsbeslissingen te rationaliseren door consequent gebruik te maken van de meest recente inzichten uit goed uitgevoerd wetenschappelijk onderzoek. Deze inzichten worden gecombineerd met de klinische vaardigheden van de professional en uiteraard de voorkeur of toestand van de patiënt. EBP heeft zijn wortels in de geneeskunde (*Evidence-Based Medicine*), maar speelt inmiddels ook een belangrijke rol binnen het verpleegkundige discipline. EBP bestaat doorgaans uit vijf stappen, waaronder het kritisch beoordelen van de wetenschappelijke kwaliteit van de gevonden literatuur. Dit vereist kennis van de begrippen en methoden uit de klinische epidemiologie en oefening met het lezen en beoordelen van wetenschappelijke artikelen. Klinimetrie is een onderdeel van EBP en houdt zich bezig met de kwaliteit van metingen en meetinstrumenten in de klinische praktijk [1]. Het

kritisch beoordelen van een klinimetrische studie is niet eenvoudig: training en ervaring zijn hiervoor noodzakelijk.

Het doel van dit proefschrift is om de klinimetrische eigenschappen van een aantal meetinstrumenten op de IC te beoordelen. In **deel één** van dit proefschrift (hoofdstuk 2 en 3) heb ik de validiteit en betrouwbaarheid van twee pijnobservatieschalen voor mechanisch beademde IC-patiënten onderzocht. De helft van de patiënten op de Intensive Care (IC) ervaart matige tot ernstige pijn tijdens rust en routinehandelingen. Onderzoek toont aan dat het draaien van patiënten één van de meest pijnlijke handelingen op de IC is. Onbehandelde, acute pijn kan nadelige gevolgen hebben, waaronder het doormaken van een myocardinfarct na chirurgie, een gebrek aan slaap en het ontwikkelen van een posttraumatische stressstoornis. Richtlijnen adviseren om structureel pijn te meten en te behandelen. Indien een patiënt niet zelf kan aangeven hoeveel pijn hij/zij ervaart, kan gebruik gemaakt worden van een pijnobservatiescore. Recente richtlijnen adviseren om bij deze patiënten de *Behavioral Pain Scale* (BPS) of de *Critical-Care Pain Observation Tool* (CPOT) te gebruiken. Dit zijn gevalideerde pijnobservatiescores voor beademde IC-patiënten die niet kunnen communiceren. Het doel van deze studies was te onderzoeken welk meetinstrument het beste onderscheid kan maken tussen een pijnlijke handeling (draaien) en niet-pijnlijke handeling (mondzorg) bij algemene IC-patiënten en bij IC-patiënten na cardiochirurgie.
De betrouwbaarheid van alle pijnmetingen was voldoende, maar tijdens rust was de betrouwbaarheid van de BPS en CPOT onvoldoende in patiënten na cardiochirurgie. Beide scores bleken tijdens rust niet in staat om te kunnen discrimineren tussen pijn en geen pijn. In beide populaties leken de BPS en CPOT valide, maar de BPS liet ook een stijging van de score zien bij een niet-pijnlijke handeling. Daarom vinden wij de CPOT beter geschikt voor het dagelijkse gebruik op de IC. Beide meetinstrumenten bleken echter niet valide in geagiteerde patiënten.
Er bestond nog geen officiële gevalideerde Nederlandse vertaling van de CPOT. Daarom hebben wij de Engelse CPOT door middel van een ' vertaling en terugvertaling' (*forward-backward translation*) – vertaald in het Nederlands. Van deze Nederlandse versie van de CPOT (CPOT-NL) hebben we de validiteit, de betrouwbaarheid en het afkappunt voor pijn onderzocht in een gemengde IC-populatie (algemene IC-patiënten en patiënten na cardiochirurgie). Driekwart van deze patiënten was opgenomen vanwege cardiochirurgie. Helaas bleek in de betrouwbaarheid van de CPOT-NL tijdens rust en tijdens de pijnlijke handeling laag. Wanneer er weinig variatie in de pijnscores is, is de meetfout te groot en het discriminerende vermogen van de score matig. De score bleek onvoldoende in staat om onderscheid te maken tussen pijn en geen pijn. Het afkappunt van de CPOT-NL voor pijn was 2 punten met een sensitiviteit van 39%. Dus de kans dat de CPOT-NL pijn detecteert bij patiënten met pijn is slechts 39% en hierdoor is de kans dat de verpleegkundige de pijn onderschat te groot. De specificiteit was 85%: de kans dat de CPOT-NL ten onrechte aangeeft dat een patiënt pijn heeft is klein. De kans op het toedienen van te veel pijnmedicatie is daardoor ook klein.

Ondanks de teleurstellende klinimetrische eigenschappen van de CPOT-NL, blijkt uit de wetenschappelijke literatuur dat de CPOT het meest geschikte instrument is om pijn te meten bij beademde IC- patiënten die zelf geen pijn kunnen aangeven. Het afkappunt voor pijn varieert in de verschillende onderzoeken en de CPOT kan geen onderscheid maken tussen weinig, matige- of ernstige pijn. Verpleegkundigen moeten naast het frequent gebruik van de CPOT, ook letten op andere signalen die op pijn kunnen wijzen. Het is belangrijk om de klinimetrische eigenschappen van een instrument te meten in de populatie waarin hij gebruikt wordt. Het is de vraag of de gemengde IC-populatie geschikt was om de CPOT-NL te onderzoeken, want uit dit proefschrift en literatuuronderzoek blijkt dat algemene IC-patiënten en patiënten na cardiochirurgie misschien te veel verschillen.

In **deel twee** van dit proefschrift is de frequentie van QT-tijd verlenging en het continu meten van de QT-tijd met behulp van de monitor onderzocht. De QT-tijd kan door middel van een ECG of een monitor worden gemeten. Een verlenging van de QT-tijd kan ernstige hartritmestoornissen veroorzaken, die zelfs kunnen leiden tot het overlijden van een patiënt. Een verlengde QT-tijd kan aangeboren zijn, maar kan ook worden veroorzaakt door bijvoorbeeld ziektes van het hart, een verstoring van de elektrolyten balans (zoals zout en kalium) of bepaalde medicijnen. Kritiek zieke patiënten krijgen vaak dergelijke medicijnen en hebben een verhoogde kans op een verstoorde elektrolyten balans. Om patiënten goed te bewaken liggen alle IC-patiënten aan een monitor die continu de bloeddruk en het hartritme controleert. De QT-tijd wordt meestal één keer per dag gemeten met een ECG. Sinds 2008 is er nieuwe software die naast het hartritme ook continu de QT-tijd kan meten. IC-patiënten hebben een verhoogde kans op het krijgen van een verlengde QT-tijd en uit het onderzoek in hoofdstuk 6 bleek dit dan ook bij meer dan de helft van de patiënten voor te komen. Het continu meten van de QT-tijd, in combinatie met een dagelijkse meting van de QT-tijd met een ECG, lijkt een belangrijk onderdeel van het bewaken van de IC-patiënt omdat eerder onderzoek de incidentie van QT-verlenging mogelijk onderschat. De continue QT-meting liet systematisch een iets langere QT-tijd zien in vergelijking met de gouden standaard meting (de handmatige beoordeling van het ECG), waardoor de kans op het missen van een verlengde QT-tijd klein lijkt. Daarom kon geconcludeerd worden dat de accuratesse van de continue QT meting voldoende is bij patiënten met een normale QRS-breedte.

In **deel drie** van de proefschrift is de accuratesse, effectiviteit en veiligheid van een continue subcutane glucosemeter bij IC-patiënten onderzocht. Stress hyperglycemie komt, ongeacht de aanwezigheid van diabetes, bij meer dan 90% van de IC- patiënten voor. Om de glucosewaarden te reguleren bepalen verpleegkundigen frequent de glucosewaarden in het bloed en behandelen de patiënt, indien nodig, met insuline. Op onze IC maken we voor de glucoseregulatie gebruik van een computer gestuurd algoritme waarbij de glucosewaarden intermitterend gemeten worden via arteriële bloedgassen en *point of care* metingen met de *AccuChek* glucosemeter. Deze glucosewaarden komen automatisch in ons patiënt data

management systeem (PDMS). Het PDMS geeft vervolgens de verpleegkundige een opdracht om de pompstand te wijzigen. Dit leidde in 2016 tot ruim 56 duizend glucose bepalingen. Subcutane continue glucose meters (CGM) meten via een dunne naald in de huid en zijn ontwikkeld voor thuisgebruik. Ze worden inmiddels ruim 10 jaar gebruikt in studieverband op de IC, maar zijn nog steeds geen standaard care. Het voordeel van een CGM-meter is dat te lage- of te hoge glucosewaarden mogelijk eerder gedetecteerd worden. Daarnaast zou het potentieel minder tijd en geld kunnen kosten. Uit onze onderzoeken bleek dat subcutane continue glucosemeting minimaal invasief is in tegenstelling tot intravasculaire continue glucosemeters. Uit het literatuuronderzoek bleek dat er geen ernstige bijwerkingen van de CGM-meters gerapporteerd zijn. De CGM-meter heeft potentiële voordelen, zoals de verbetering van de glucoseregulatie, het verminderen van de variabiliteit van de glucose waarden en een reductie van kosten, maar dit is nog niet aangetoond in wetenschappelijk onderzoek. Daarnaast was de accuratesse van het onderzochte CGM-apparaat nog onvoldoende voor dagelijks gebruik op de IC.

Deel vier van dit proefschrift beschrijft de klinimetrische eigenschappen van een meetinstrument om EBP-attitudes en EBP-kennis en vaardigheden onder verpleegkundigen te meten. Hoewel uit eerdere onderzoeken blijkt dat de McColl vragenlijst de meest gebruikte EBP- vragenlijst onder verpleegkundigen en artsen is, waren de klinimetrische eigenschappen onvoldoende beschreven. In hoofdstuk 11 zijn de validiteit en betrouwbaarheid van de Nederlandse McColl (McColl-NL) onderzocht in verpleegkundigen in een perifeer opleidingsziekenhuis. Hiervoor is eerst de Engelstalige McColl in het Nederlands vertaald door middel van de *two-panel* methode. Vervolgens is de McColl-NL uitgezet onder 198 verpleegkundigen, waarvan 27% de vragenlijst heeft ingevuld. De betrouwbaarheid en validiteit van de subschaal 'EBP-attidude' waren onvoldoende. De betrouwbaarheid van de subschalen 'Kennis EBP-tijdschriften/websites' en 'begrip EBP-termen' was respectievelijk 'bijna acceptabel' en 'goed'. Een combinatie van deze kennis subschalen resulteerde in een goede betrouwbaarheid. De validiteit van de subschalen 'Kennis EBP-tijdschriften/websites' en 'begrip EBP-termen' was voldoende. De gecombineerde kennis schaal (tijdschriften/websites en termen) kan worden gebruikt in combinatie met een EBP-attitude vragenlijst met goede klinimetrische eigenschappen. De klinimetrische eigenschappen van een dergelijke gecombineerde schaal zou onderzocht moeten worden in grotere populaties.

Het is belangrijk om te beseffen dat meetinstrumenten, zoals vragenlijsten, officieel moeten worden vertaald en gevalideerd in de taal waarin het instrument zal worden gebruikt. Bovendien moeten de klinimetrische eigenschappen van een meetinstrument of apparaat worden beoordeeld in de (patiënten) populatie waarin het instrument zal worden toegepast. Dit proefschrift levert een bijdrage aan het proces van het beoordelen van de klinimetrische eigenschappen van een aantal meetinstrumenten / apparaten op de IC. Aanvullend onderzoek

om meer inzicht te krijgen in de validatie van meetschalen is noodzakelijk: klinimetrisch onderzoek is een continu proces.

REFERENTIES

1. Offringa M, Assendelft WJJ, Scholten RJPM. Inleiding in evidence-based medicine. Derde herziene druk. Bohn Stafleu van Loghum; 2008. Hfst. 4 p. 134-149
2. Bouter LM, van Dongen MCJM, Zielhuis GA. In: Diagnostiek en prognostiek. Epidemiologisch onderzoek Opzet en interpretatie. 5de herziene druk ed. Houten: Bohn Stafleu van Loghum; 2005. p. 245-95.
3. Vandenbroucke JP, Hofman A. Grondslagen der epidemiologie. Zesde druk. Elsevier.1999. p 1-10

ACKNOWLEDGEMENTS

This thesis would not have been possible without the time, support and dedication of many people involved throughout this research project including those highlighted in the list of contributing authors.

Firstly, I would like to express my sincere gratitude to Professor Peter van der Voort for allowing me to conduct this research under his guidance. I am especially grateful for the confidence and the freedom he gave me to accomplish this work. I would also like to thank Professor Nardo van der Meer who provided me with valuable suggestions for this thesis and the opportunity to perform this PhD at TIAS School for Business and Society.

I extend my sincere thanks to all nurses, physicians and participating patients of the Department of Intensive Care OLVG (East), as well as all those who contributed directly or indirectly to this dissertation. In particular, I wish to thank Rob Bosman, Emmy Rood and Addy Bianchi for the numerous data extractions from the patient clinical information system as well as for the custom made digital study forms and study orders.

My colleagues and roommates Sigrid van Steen, Sophie Buitinck and Marissa van Ingen, many thanks for your support, advice and "*gezelligheid*".

My colleagues Joep Maeijer and Manja Herrebrugh from the audiovisual department of the OLVG, thank you for formatting the figures throughout this thesis. Above all, many thanks for the support , coffee breaks, walks in the park and great companionship over the years.

Finally, my warmest gratitude goes to Som Toohey for his endless English revisions, including the acknowledgements, and support.

LIST OF CONTRIBUTING AUTHORS

Boom D.T.
Department of Intensive Care, OLVG, Amsterdam, The Netherlands

Bosman R.J.
Department of Intensive Care Medicine, OLVG, Amsterdam, the Netherlands

Eskes A.M.
Department of Surgery, Academic Medical Center, Amsterdam, the Netherlands

Feijen H.M.
Department of Intensive Care Medicine, OLVG, Amsterdam, the Netherlands

Endeman H.
Department of Intensive Care Medicine, OLVG, Amsterdam, the Netherlands

Hermanides J.
Department of Anesthesiology, Academic Medical Center Amsterdam, The Netherlands

Hoogstraaten E.
Department of Intensive Care Medicine, OLVG

Janssen G.H.J.
Department of Intensive Care Medicine, OLVG

Kreder S.
Department of Intensive Care Medicine, OLVG

Limpens J.
Medical Library, Academic Medical Center, Amsterdam, The Netherlands

Lindeboom R.
Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Medical Faculty, Academic Medical Center and University of Amsterdam, the Netherlands

Maaskant J.M.
Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre and University of Amsterdam, the Netherlands

van der Meer N.J.
TIAS School for Business and Society, Tilburg University, Tilburg, The Netherlands

van Oostveen C.
Spaarne Gasthuis Academy, Spaarne Gasthuis Hospital, Haarlem, the Netherlands

Oudemans-van Straaten H.M.O.
Department of Intensive Care Medicine, Free University Medical Centre, Amsterdam, the Netherlands

van der Voort P.H.J.
Department of Intensive Care Medicine, OLVG, Amsterdam, the Netherlands
TIAS School for Business and Society, Tilburg University, Tilburg, The Netherlands

Sechterberger M.K.
Department of Endocrinology, Academic Medical Center, Amsterdam, Netherlands

Van Stijn I.
Department of Intensive Care Medicine, OLVG

Stilma W.
Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Sturm J.
Rotterdam University of Applied Sciences, Rotterdam, the Netherlands

Vermeulen H.
IQ Healthcare Radboud institute of Health Sciences, Scientific Center for Quality of Healthcare, Nijmegen, the Netherlands

DeVries J.H.
Department of Endocrinology, Academic Medical Center, Amsterdam, Netherlands

Wester J.P.J.
Department of Intensive Care Medicine, OLVG

CURRICULUM VITAE

Werkervaring

2009 tot heden	**Wetenschapscoördinator Intensive Care Unit OLVG locatie Oost** *OLVG, Amsterdam*
2010 - 2015	**Wetenschapscoördinator verpleegkundig onderzoek Teaching Hospital OLVG** *Teaching Hospital, het Onze Lieve Vrouwe Gasthuis, Amsterdam*
2009	**Wetenschapscoördinator Teaching Hospital OLVG a.i.** *Het Onze Lieve Vrouwe Gasthuis, Amsterdam*
2006	**Specialistisch verpleegkundige Cardiac care** *Cardiologiecentrum Amsterdam-Zuid, Heart for Health, Amsterdam*
2005 - 2009	**Specialistisch verpleegkundige Cardiac Care Unit** *Het Onze Lieve Vrouwe Gasthuis, Amsterdam*
2004 - 2005	**Vervolg opleiding tot Cardiac care verpleegkundige** *Het Onze Lieve Vrouwe Gasthuis, Amsterdam Amstel Academie, Amstelveen*
2003 - 2004	**Verpleegkundige afdeling cardiologie** *Het Onze Lieve Vrouwe Gasthuis, Amsterdam*
2000 - 2003	**Verpleegkundige in opleiding** *Het Onze Lieve Vrouwe Gasthuis, Amsterdam*

Opleidingen en kwalificaties

2006 - 2010	**Masterstudie Evidence-Based Practice (EBP)** *Academisch Medisch Centrum (AMC-UvA), Amsterdam*
2004 – 2005	**Vervolgopleiding tot Cardiac care verpleegkundige** *Amstel Academie, Amsterdam*
2000 – 2003	**Bachelor opleiding Verpleegkunde** *Hogeschool INHOLLAND, Diemen*
1997 – 1999	**Nederlandse taal- en letterkunde** Propedeutisch examen Nederlandse taal- en letterkunde (1998) *Vrije Universiteit, Amsterdam*
1996 – 1997	**Bachelor opleiding HBO-Verpleegkunde** Propedeutisch examen opleiding tot Verpleegkundige (1997) *Hogeschool van Amsterdam, Amsterdam*
1990 – 1996	**Voorbereidend wetenschappelijk onderwijs** *Het Pascal College, Zaandam*

Cursussen

2017	**Multilevel analyse** EpidM *Afdeling epidemiologie & biostatistiek VU medisch centrum, Amsterdam*
2016	**WMO/GCP training (online)** *GCP Central B.V.*
2014	**Klinische predictiemodellen** EpidM *Afdeling epidemiologie & biostatistiek VU medisch centrum, Amsterdam*
2010	**Alumni onderwijs: sample size calculation and power analysis for intervention and diagnostic research** *Master EBP, Academisch Medisch Centrum (AMC-UvA), Amsterdam*
2010	**Alumni onderwijs: building and validating prediction models using logistic regression** *Master EBP, Academisch Medisch Centrum (AMC-UvA), Amsterdam*

LIST OF PUBLICATIONS

Rijkenberg S, van Steen S.C.J, DeVries J.H, van der Voort P.H.J.
Accuracy and reliability of a continuous glucose measurement device in critically ill patients
Accepted for publication in Journal of Clinical Monitoring and Computing, November 2017

van Steen S.C, **Rijkenberg S,** Sechterberger M.K, DeVries J.H, van der Voort P.H.J.
Glycemic effects of a low carbohydrate enteral formula compared with an enteral formula of standard composition in critically ill patients: an open-label randomized controlled clinical trial
Accepted for publication in Journal of Parenteral and Enteral Nutrition , November 2017

Rijkenberg S, Stilma W, Bosman R.J, van der Meer N.J, van der Voort P.H.J.
Pain Measurement in Mechanically Ventilated Patients After Cardiac Surgery: Comparison of the Behavioral Pain Scale (BPS) and the Critical-Care Pain Observation Tool (CPOT). J Cardiothorac Vasc Anesth. 2017 Aug;31(4):1227-1234.
Epub 2017 Mar 15. PubMed PMID: 28800982.

van der Voort P.H.J, Bosman R.J, Franssen E.J.F, **Rijkenberg S**.
Opening the Microcirculation with Ketanserin I.V. Improves Peripheral Temperature: An Observational Cohort Study. Journal of Anesthesiology and Critical Care Medicine. 2017 Feb;4(1).Enliven Archive, www.enlivenarchive.org

van Steen S.C, **Rijkenberg S**, Limpens J, van der Voort P.H.J, Hermanides J, DeVries J.H. The Clinical Benefits and Accuracy of Continuous Glucose Monitoring Systems in Critically Ill Patients-A Systematic Scoping Review. Sensors (Basel). 2017 Jan 14;17(1). Review. PubMed PMID: 28098809;
PubMed Central PMCID: PMC5298719.

Woldhek A.L, **Rijkenberg S**, Bosman R.J, van der Voort P.H.J.
Readmission of ICU patients: A quality indicator? J Crit Care. 2017 Apr;38:328-334.
Epub 2016 Dec 6. PubMed PMID: 27939901.

Rijkenberg S, van der Voort P.H.J
Can the critical-care pain observation tool (CPOT) be used to assess pain in delirious ICU patients? J. Thorac Dis. 2016
May;8(5):E285-7. PubMed PMID: 27162683; PubMed Central PMCID: PMC4842811.

Houwink A.P, **Rijkenberg S**, Bosman R.J, van der Voort P.H.J.The association between lactate, mean arterial pressure, central venous oxygen saturation and peripheral temperature and mortality in severe sepsis: a retrospective cohort analysis.
Crit Care. 2016 Mar 12;20:56. PubMed PMID: 26968689; PubMed Central PMCID: PMC4788911.

Stilma W, **Rijkenberg S**, Feijen H.M, Maaskant J.M, Endeman H.
Validation of the Dutch version of the critical-care pain observation tool.
Nurs Crit Care. 2015 Dec 22. [Epub ahead of print] PubMed PMID: 26689613.

Janssen G.H, **Rijkenberg S**, van der Voort P.H.J.
Validation of continuous QTc measurement in critically ill patients.
J Electrocardiol. 2016 Jan-Feb;49(1):81-6. doi: 10.1016/j.jelectrocard.2015.10.001. Epub 2015 Oct 22. PubMed PMID: 26520168.

Steenbergen S, **Rijkenberg S**, Adonis T, Kroeze G, van Stijn I, Endeman H.
Long-term treated intensive care patients outcomes: the one-year mortality rate, quality of life, health care use and long-term complications as reported by general practitioners.
BMC Anesthesiol. 2015 Oct 12;15:142. PubMed PMID: 26459381; PubMed Central PMCID:PMC4604105.

van der Voort P.H, de Metz J, Wester J.P, van Stijn I, Feijen H.M, Balzereit A, **Rijkenberg S**, Obster R, Bosman R.J. Telemedicine in a Dutch intensive care unit: A descriptive study of the first results. J Telemed Telecare. 2016 Apr;22(3):141-7. Epub 2015 Jul 2. PubMed PMID: 26141722.

Rijkenberg S, Stilma W, Endeman H, Bosman R.J, Oudemans-van Straaten H.M. Pain measurement in mechanically ventilated critically ill patients: Behavioral Pain Scale versus Critical-Care Pain Observation Tool. J Crit Care. 2015 Feb;30(1):167-72. Epub 2014 Sep 22. PubMed PMID: 25446372.

Boom D.T, Sechterberger M.K, **Rijkenberg S**, Kreder S, Bosman R.J, Wester J.P, van Stijn I, DeVries J.H, van der Voort P.H.J.
Insulin treatment guided by subcutaneous continuous glucose monitoring compared to frequent point-of-care measurement in critically ill patients: a randomized controlled trial. Crit Care. 2014 Aug 20;18(4):453. PubMed PMID: 25139609; PubMed Central PMCID: PMC4161875.

Hoogstraaten E, **Rijkenberg S**, van der Voort P.H.J.
Corrected QT-interval prolongation and variability in intensive care patients.
J Crit Care. 2014 Oct;29(5):835-9. Epub 2014 May 27. PubMed PMID: 24986247.

van der Voort P.H.J, Westra B, Wester J.P, Bosman R.J, van Stijn I, Haagen I.A, Loupatty F.J, **Rijkenberg S**.
Can serum L-lactate, D-lactate, creatine kinase and I-FABP be used as diagnostic markers in critically ill patients suspected for bowel ischemia.
BMC Anesthesiol. 2014 Dec 2;14:111. eCollection 2014. PubMed PMID: 25844063; PubMed Central PMCID: PMC4384375.