Improving the acute and perioperative hemodynamic assessment

Kaufmann, Thomas

DOI:
10.33612/diss.128650137

Improving the acute and perioperative hemodynamic assessment

Thomas Kaufmann

Financial support by the Graduate School of Medical Sciences (GSMS)

Printed by	Ipskamp Printing Enschede, the Netherlands
Design & layout	Bianca Pijl, www.pijlldesign.nl Groningen, the Netherlands
ISBN	978-94-034-2698-3 *(print)* 978-94-034-2699-0 *(digital)*

rijksuniversiteit
groningen

Improving the acute and perioperative hemodynamic assessment

Proefschrift

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op

woensdag 8 juli 2020 om 16.15 uur

door

Thomas Kaufmann

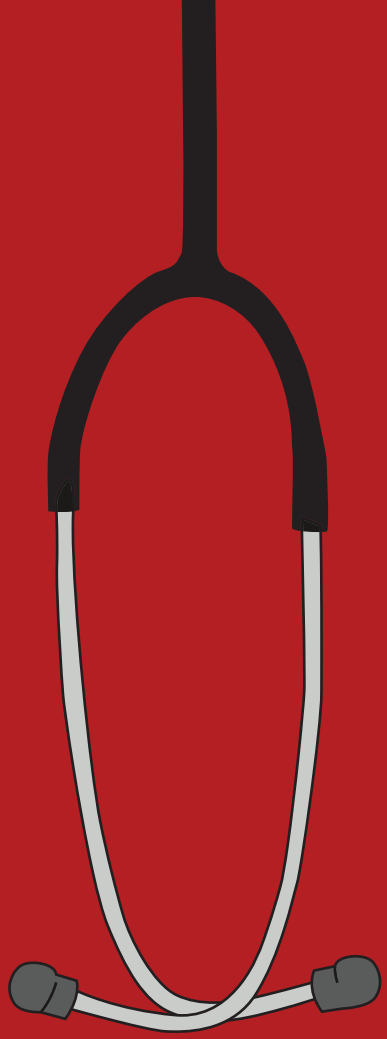
geboren op 11 augustus 1991
te Venlo

Promotores
Prof. dr. T.W.L. Scheeren
Prof. dr. J.C.C. van der Horst

Beoordelingscommissie
Prof. dr. J. Bakker
Prof. dr. W.F. Buhre
Prof. dr. P.H.J. van der Voort

Paranimfen
Pieter Steinkamp
Mark Kloosterhuis

Table of contents



1

General introduction and thesis outline

General introduction

Hemodynamic instability is often encountered in critically ill patients. Both patients undergoing high-risk surgery in the operating room (OR) and patients admitted to the Intensive Care Unit (ICU) may become hemodynamically unstable as a consequence of hypovolemia or vasodilation. While the pathophysiology and the underlying cause of hemodynamic instability may differ, both patients in the OR and the ICU have in common that hemodynamic instability leads to an increased risk of complications and mortality (1,2).

To improve the outcome of critically ill patients, caregivers aim to prevent hemodynamic instability. Hemodynamic instability results in insufficient perfusion of tissues and leads to organ dysfunction introducing a cascade that involves further loss of intravascular volume, i.e., hypovolemia, and vascular tone, i.e., vasodilation. The ultimate goal is to ensure sufficient oxygenation of the tissues through oxygen delivery and perfusion pressure. Oxygen delivery is the product of the volume of blood being pumped by the heart, i.e., cardiac output, and the amount of oxygen available in this blood, i.e., arterial oxygen content. Perfusion pressure is the difference between the inflow pressure, i.e., mean arterial pressure, and the outflow pressure of the tissues. In short, the delivery of oxygen to the tissues requires adequate flow and pressure in the blood vessels. An overview of the interaction of these hemodynamic variables, which can be used as a target to prevent hemodynamic instability or potential intervention target when instability occurs, is shown in Figure 1.

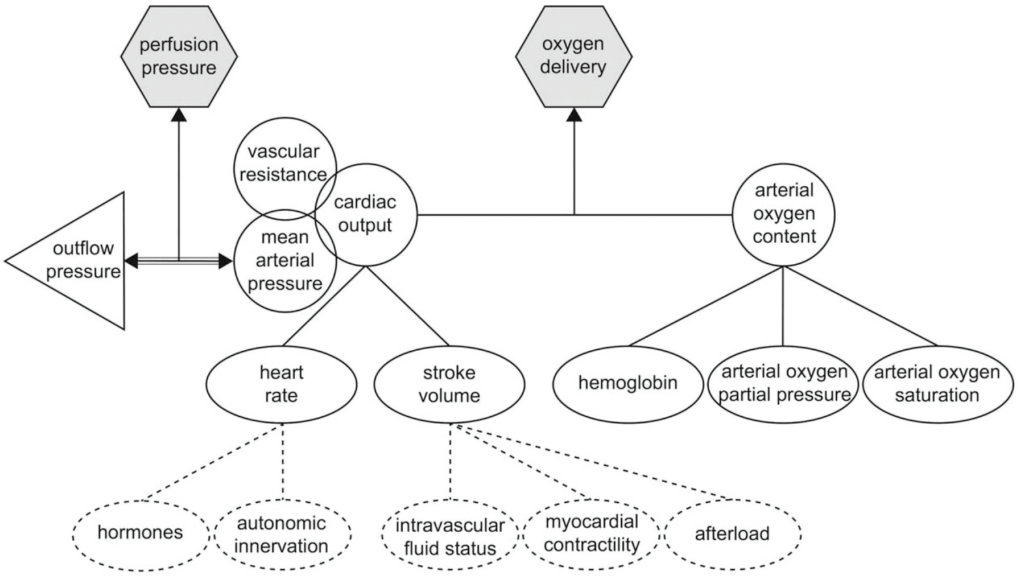


Figure 1. *Interaction of hemodynamic variables that can be targeted to prevent and treat hemodynamic instability. The aim is to optimize perfusion pressure and oxygen delivery to ensure adequate oxygenation of the organs and tissues (3)*

The real-time measurement of these hemodynamic variables is called hemodynamic monitoring, and a device that performs these measurements is called a hemodynamic monitor. Various types of hemodynamic monitors are commercially available, and many more are being developed. Monitoring devices range from fully invasive, e.g., a pulmonary artery catheter to measure cardiac output, to non-invasive, e.g., an oscillometric blood pressure cuff placed on the upper arm to intermittently measure blood pressure. Each of these monitors is indicated in distinctive clinical scenarios in different settings. The clinical heterogeneity of studies makes a comparison of the feasibility and applicability of monitors and judging the benefit of interventions using these monitors to obtain high-quality evidence difficult.

Thesis outline

The process of preventing hemodynamic instability by optimizing hemodynamic variables, and using the correct hemodynamic monitor to achieve this purpose remains a problematic aspect of perioperative and critical care medicine. Many different monitors are available that measure specific variables differently, and each monitor has to be validated correctly before it can be implemented in clinical practice. Failure to correctly validate monitors according to scientific standards or misinterpret conclusions of conducted studies can lead to erroneous decisions and, thus, potentially dangerous patient care.

This thesis has two aims. First, we aim to extend the available evidence on the applicability of hemodynamic monitoring to prevent and treat hemodynamic instability during the perioperative period and during admission to the ICU. Second, we aim to gain knowledge on how to improve the conduct of studies in perioperative and critical care medicine.

Throughout the thesis, the reader should be aware that a difference exists between the absence of evidence for monitoring versus evidence for the absence of monitoring in critically ill patients. Although evidence might lack for monitoring blood pressure and heart rate during surgery or admission to the ICU, no-one would perform surgery without monitoring these hemodynamic variables or treat a patient in the ICU without being informed on the circulation. Therefore, recommendations will be made even if high-quality evidence is absent. The goal is to increase knowledge and become aware of available evidence and areas where additional information is necessary. Proposing something as complex as hemodynamic monitoring to be either beneficial for every patient and setting or redundant overall will never be the conclusion of a thesis, as almost nothing is absolute.

Hemodynamic monitoring in the operating room and the intensive care

Hemodynamic monitoring during the perioperative period differs from basic to very advanced. Depending on the patient and the intervention, monitoring can vary from intermittent measurements of blood pressure alone to continuous measurements of cardiac output and surrogates of organ perfusion. More advanced hemodynamic monitors in perioperative medicine

are most often used when perioperative goal-directed therapy (PGDT) is applied. PGDT aims at optimizing global hemodynamics during the perioperative period by titrating interventions such as fluids, vasopressors, and inotropes to predefined hemodynamic targets (4). Assessment of the triggers and targets for PGDT and the response to interventions needed to reach these targets require hemodynamic monitoring beyond blood pressure and heart rate. Even though a vast amount of research has been conducted in this field, conclusions on the optimal monitor and the effect of applying to monitor combined with interventions on patient outcomes are currently inconsistent. Some authors of systematic reviews with meta-analysis state that using PGDT improves perioperative outcomes such as mortality or morbidity, other systematic reviews with meta-analysis conclude that there is no such benefit (5–7). Specific types of surgery or specific hemodynamic monitoring devices are often incorporated in these reviews (8,9). In **chapter 2**, we present a systematic overview of all studies that evaluated PGDT interventions in perioperative medicine.

Based on the available evidence for PGDT, several national and international anesthesiology societies have written practice guidelines that support the use of PGDT as part of perioperative care (10–12). Despite the availability of practice guidelines, the adoption and implementation of PGDT have been slow and incomplete (13). In **chapter 3**, we present a narrative review of the evidence for PGDT, which helps explain the considerations regarding the application of the PGDT intervention. We suggest the current best practice approach to using PGDT and discuss several future directions of PGDT, which may help to evolve this research field further.

Hemodynamic monitoring in the ICU is even more complicated than in the OR. First, patients with hemodynamic instability usually are admitted acutely. Second, hemodynamic instability evolves. It can be the presenting condition of a patient admitted to the ICU, but it can also develop early after admission or later after several days, depending on the underlying disease or complication. Third, several patient characteristics in patients with hemodynamic instability hamper hemodynamic monitoring with many monitors. For example, the presence of atrial fibrillation limits reliable assessments of the arterial pressure waveform which is done by several monitors.

Despite these difficulties, hemodynamic monitoring in the ICU is often used, especially in patients with persistent circulatory shock. Most often, to measure blood pressure, heart rate, and cardiac output, to diagnose the underlying cause or the volume state, and to optimize blood pressure and cardiac output using interventions. In the acute setting of circulatory shock, physicians mostly rely on clinical examination. Clinical examination of critically ill patients with circulatory shock is challenging. Patients may present with varying states of circulating blood volume, cardiac contractility, sympathetic nervous activity, vascular tone, and microcirculatory function depending on the type of shock. Assessment is even more difficult if comorbidities are present (14), which is increasingly the case in all patients. Clinical examination is practiced daily for clinical care, although it is not based on high-quality evidence (15). This lack of evidence leads to clinical examination in patients with circulatory shock being considered 'best practice' in circulatory shock guidelines (1).

To help improve clinical examination for hemodynamic estimates, we need to understand the current clinical practice. Our research group has recently conducted a large prospective cohort study, the Simple Intensive Care Studies-I (SICS-I), where researchers performed a standardized clinical examination, and clinical, laboratory and hemodynamic variables were collected from all acutely admitted critically ill patients during the first 24 hours of admission (16,17). As part of this protocol, researchers were asked to estimate cardiac function (i.e., cardiac index) using the variables obtained with clinical examination. In **chapter 4**, we present a Bayesian network that maps the decision-making process underlying researchers' estimates of cardiac function. This study was a predefined substudy of the SICS-I (16).

In patients with circulatory shock, it is recommended by guidelines to initially target mean arterial blood pressure of 65 mmHg (1). Invasive arterial blood pressure monitoring using an arterial catheter is considered the clinical reference method in critically ill patients. Non-invasive oscillometric blood pressure measurement using an upper-arm cuff is a widely used alternative for invasive monitoring (18). It is not yet conclusive whether non-invasive blood pressure measurements show agreement with the invasive arterial clinical reference technique. While some argue that non-invasive measurements may safely replace invasive measurements (19), others show an unacceptable disagreement in critically ill patients with circulatory shock (20,21). In **chapter 5**, we compared blood pressure measurements obtained using upper-arm cuff oscillometry with arterial catheter-derived blood pressure measurements in a large prospective cohort of critically ill patients with and without receiving norepinephrine. To investigate the clinical relevance of the differences between both methods, the recently developed error-grid method was used (22).

Non-invasive echocardiography is considered the standard clinical method to first measure cardiac output and to evaluate the type of shock in patients with circulatory shock (1). Application of the ultrasonography technique in the ICU is called critical care ultrasonography (CCUS) and is performed by critical care personnel, as opposed to a certified sonographer or cardiologist. CCUS is focused on the clinical question at hand, e.g., a quick, reliable measurement of cardiac output, and does not require obtaining all acoustic views from the standard windows (23). It has become more popular over the last years with the increasing availability of ultrasonography devices, and various professional bodies now mandate competency in CCUS (24). Data on the feasibility of obtaining images in ICU patients, the quality of the images obtained, and the quality of the measurements performed, is sparse but of significant importance for the adaptation of CCUS (23). In **chapter 6**, we present data on the feasibility of having CCUS performed by medical students, as an example of non-experts, in the ICU. This study was a predefined substudy of the SICS-I (16).

More advanced monitoring devices may become indicated if there is a more complex shock state (25). Besides measuring absolute values of cardiac output, it is also important to monitor changing trends in response to interventions such as a fluid challenge. One method of more advanced hemodynamic monitoring is the uncalibrated pulse wave analysis method, which estimates cardiac output from the arterial pressure waveform of an indwelling arterial line. In general, the

reliability of this method must be interpreted with caution in patients with extensive changes in vascular tone, which may occur in circulatory shock (25). Recently, in order to improve the measurement performance, a new algorithm to estimate cardiac output was developed for one of these devices. In **chapter 7**, we compare the agreement and trending ability of cardiac output measurements made with arterial pressure waveform analysis to cardiac output measurements made with echocardiography. This study was a predefined substudy of the Simple Intensive Care Studies-I (SICS-I) (16).

How to improve the methodology and conduct of studies?

The second aim of this thesis was to gain knowledge on how to improve the conduct of studies in perioperative and critical care medicine. In the available literature, as well as in our studies, we found limitations regarding methodology and study conduct. Therefore, the quality of evidence should be considered when appreciating published studies. Many reasons exist as to why the quality of evidence of studies may be decreased, such as limitations in the methods, the inconsistency of results, indirectness of evidence, imprecision, or publication bias (26). Several of these problems are correctable during the design and conduct of studies, and it is essential to do so as weaknesses in studies can produce misleading results, results that can be implemented in guidelines, results that even change daily practice and could lead to harm and waste valuable resources as well (27). Over the last decades, a number of initiatives have been proposed and implemented to improve study planning and conduct as well as reporting of studies, such as checklists for reporting studies (28,29), the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to appreciate quality of evidence (26), and guidance on the use of core outcome sets when developing trials (30). These measures have helped partially to improve the methodology and conduct of studies.

When studies are performed in multiple hospitals or when a systematic review pools the results of various studies, it is essential that the studied patient populations are comparable for each hospital. Illness severity models are used in the ICU to characterize disease severity and degree of organ dysfunction, and predict outcome in critically ill patients (31). Examples of these models, which also provide mortality prediction, include the Acute Physiology and Chronic Health Evaluation-IV (APACHE-IV) (32) and the Simplified Acute Physiology Score-II (SAPS-II) (33). For research purposes, these models are used to compare different study populations. Therefore, these models must be appropriately developed and validated for optimal use. Until now, no study has systematically assessed which models have been developed. In **chapter 8**, we present a scoping review of all available mortality prediction models for adult critically ill patients.

Even though several measures have been implemented to improve research, further improvements are still possible and needed. In **chapter 9,** we comment on a recently published randomized controlled trial that was stopped for futility and propose that full protocols of trials and other clinical studies are being published in peer-reviewed journals before initiation.

References

1 Cecconi M, De Backer D, Antonelli M, Beale R, Bakker J, Hofer C, et al. Consensus on circulatory shock and hemodynamic monitoring. Task force of the European Society of Intensive Care Medicine. Intensive Care Med. 2014 Dec 13;40(12):1795–815.

2 Vincent J-L, Pelosi P, Pearse R, Payen D, Perel A, Hoeft A, et al. Perioperative cardiovascular monitoring of high-risk patients: a consensus of 12. Crit Care. 2015 Dec 8;19(1):224.

3 Saugel B, Vincent JL, Wagner JY. Personalized hemodynamic management. Curr Opin Crit Care. 2017;23(4):334–41.

4 Saugel B, Reuter DA. Perioperative Goal-Directed Therapy Using Invasive Uncalibrated Pulse Contour Analysis. Front Med. 2018;5(January):1–5.

5 Giglio M, Dalfino L, Puntillo F, Rubino G, Marucci M, Brienza N. Haemodynamic goal-directed therapy in cardiac and vascular surgery. A systematic review and meta-analysis. Interact Cardiovasc Thorac Surg. 2012;15(5):878–87.

6 Cecconi M, Corredor C, Arulkumaran N, Abuella G, Ball J, Grounds RM, et al. Clinical review: Goal-directed therapy-what is the evidence in surgical patients? The effect on different risk groups. Crit Care. 2013 Mar 5;17(2):209.

7 Benes J, Giglio M, Brienza N, Michard F. The effects of goal-directed fluid therapy based on dynamic parameters on post-surgical outcome: A meta-analysis of randomized controlled trials. Crit Care. 2014;18(5):1–11.

8 Michard F, Giglio MT, Brienza N. Perioperative goal-directed therapy with uncalibrated pulse contour methods: Impact on fluid management and postoperative outcome. Br J Anaesth. 2017 Jun 11;119(1):22–30.

9 Sun Y, Chai F, Pan C, Romeiser JL, Gan TJ. Effect of perioperative goal-directed hemodynamic therapy on postoperative recovery following major abdominal surgery-a systematic review and meta-analysis of randomized controlled trials. Crit Care. 2017 Jun 12;21(1):141.

10 National Institute for Health and Clinical Excellence. Medical technologies guidance MTG3: CardioQ-ODM oesophageal doppler monitor. 2011.

11 Vallet B, Blanloeil Y, Cholley B, Orliaguet G, Pierre S, Tavernier B. Guidelines for perioperative haemodynamic optimization. Ann Fr Anesth Reanim. 2013 Oct;32(10):e151–8.

12 European Society of Anaesthesiology. Perioperative Goal-Directed Fluid Therapy Protocol Summary. Available from: http://html.esahq.org/patientsafetykit/resources/downloads/05_Checklists/Various_Checklists/Perioperative_Goal_Directed_Therapy_Protocols.pdf

13 Cannesson M, Pestel G, Ricks C, Hoeft A, Perel A. Hemodynamic monitoring and management in patients undergoing high risk surgery: A survey among North American and European anesthesiologists. Crit Care. 2011;15(4):R197.

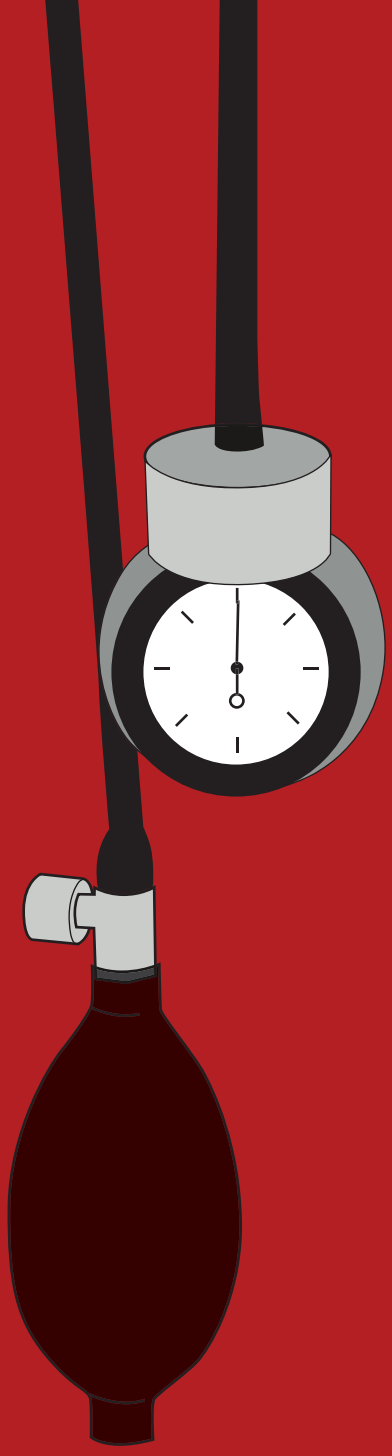
14 Saugel B, Malbrain MLNG, Perel A. Hemodynamic monitoring in the era of evidence-based medicine. Crit Care. 2016 Dec 20;20(1):401.

15 Hiemstra B, Eck RJ, Keus F, van der Horst ICC. Clinical examination for diagnosing circulatory shock. Curr Opin Crit Care. 2017 Aug;23(4):293–301.

16 Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, et al. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med. 2019 Feb;45(2):190–200.

17 Hiemstra B, Eck RJ, Wiersema R, Kaufmann T, Koster G, Scheeren TWL, et al. Clinical Examination for the Prediction of Mortality in the Critically Ill. Crit Care Med. 2019 Jul;47(10):1301–9.

18 Chatterjee A, Depriest K, Blair R, Bowton D, Chin R. Results of a survey of blood pressure monitoring by intensivists in critically ill patients: A preliminary study. Crit Care Med. 2010;38(12):2335–8.
19 Lakhal K, Ehrmann S, Boulain T. Noninvasive BP Monitoring in the Critically Ill: Time to Abandon the Arterial Catheter? Chest. 2018;153(4):1023–39.
20 Pytte M, Dybwik K, Sexton J, Straume B, Waage Nielsen E. Oscillometric brachial mean artery pressures are higher than intra-radial mean artery pressures in intensive care unit patients receiving norepinephrine. Acta Anaesthesiol Scand. 2006;50(6):718–21.
21 Riley LE, Chen GJ, Latham HE. Comparison of non-invasive blood pressure monitoring with invasive arterial pressure monitoring in medical ICU patients with septic shock. Blood Press Monit. 2017;22(4):202–7.
22 Saugel B, Grothe O, Nicklas JY. Error grid analysis for arterial pressure method comparison studies. Anesth Analg. 2018;126(4):1177–85.
23 Koster G, Van Der Horst ICC. Critical care ultrasonography in circulatory shock. Curr Opin Crit Care. 2017 Aug;23(4):326–33.
24 Vieillard-Baron A, Millington SJ, Sanfillipo F, Chew M, Diaz-Gomez J, McLean A, et al. A decade of progress in critical care echocardiography: a narrative review. Intensive Care Med. 2019;45(6):770–88.
25 Saugel B, Vincent J-L. Cardiac output monitoring: how to choose the optimal method for the individual patient. Curr Opin Crit Care. 2018;24(3):165–72.
26 Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, et al. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008 May 3;336(7651):995–8.
27 Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014 Jan;383(9912):166–75.
28 Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. BMJ. 2010 Mar 23;340(mar23 1):c332–c332.
29 Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009 Jul 21;339(5):b2535.
30 Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: Version 1.0. Trials. 2017;18(S3):1–50.
31 Vincent J-L, Moreno R. Clinical review: Scoring systems in the critically ill. Crit Care. 2010;14(2):207.
32 Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006 May;34(5):1297–310.
33 Gall JR, Lemeshow S, Saulnier F, Le Gall JR, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. JAMA J Am Med Assoc. 1993 Aug 14;270(24):2957–63.



2

Perioperative goal-directed therapy: A systematic review without meta-analysis

Thomas Kaufmann*
Ramon P. Clement*
Thomas W.L. Scheeren
Bernd Saugel
Frederik Keus
Iwan C.C. van der Horst

** Both authors contributed equally to this work*

Acta Anaesthesiologica Scandinavica

Abstract

Background

Perioperative goal-directed therapy aims to optimize haemodynamics by titrating fluids, vasopressors and/or inotropes to predefined hemodynamic targets. Perioperative goal-directed therapy is a complex intervention composed of several independent component interventions. Trials on perioperative goal-directed therapy show conflicting results. We aimed to conduct a systematic review and meta-analysis to investigate the benefits and harms of perioperative goal-directed therapy.

Methods

PubMED, EMBASE, Web of Science and Cochrane Library were searched. Trials were included if they had a perioperative goal-directed therapy protocol. The primary outcome was all-cause mortality. The first secondary outcome was serious adverse events excluding mortality. Risk of bias was assessed, and GRADE was used to evaluate quality of evidence.

Results

One hundred and twelve randomized trials were included of which one trial (1%) had low risk of bias. Included trials varied in patients: types of surgery which was expected due to inclusion criteria; in intervention and comparison: timing of intervention, monitoring devices, hemodynamic variables, target values, use of fluids, vasopressors and/or inotropes as well as combinations of these within protocols; and in outcome: mortality was reported in 87 trials (78%). Due to substantial clinical heterogeneity also within the various types of surgery a meta-analysis of data, including subgroup analyses, as defined in our protocol was considered inappropriate.

Conclusion

Clinical heterogeneity in patients, interventions and outcomes in peri-operative goal-directed therapy trials is too large to perform meta-analysis on all trials. Future trials and meta-analyses highly depend on universally agreed definitions on aspects beyond type of surgery of the complex intervention and its evaluation.

Introduction

Perioperative goa-directed therapy refers to the hemodynamic optimization during perioperative care by titrating fluids, vasopressors, and/or inotropes to predefined hemodynamic goals (1). The main purpose of perioperative goal-directed therapy is to maintain or restore sufficient oxygen delivery by providing adequate organ and tissue perfusion. However, both under-resuscitation with insufficient organ perfusion and over-resuscitation may lead to adverse outcomes (2).

In 1988, Shoemaker was the first to report lower mortality and morbidity rates associated with perioperative goa-directed therapy compared with standard care, followed by a plethora of other trials (3). Shoemaker used invasive pulmonary artery catheter (PAC) monitoring to guide his interventions. In the past, such catheters were the most widely used technique although a clear survival benefit was never proven (4). More recently, less-invasive and even non-invasive monitoring devices are used to reduce the risks associated with more invasive techniques. The British National Institute for Health and Care Excellence (NICE) and a report commissioned by the Centers for Medicare and Medicaid Services in the USA recommended the use of oesophageal Doppler monitoring (ODM) for optimizing hemodynamics in patients undergoing major surgery (5,6).

While most literature on perioperative goal-directed therapy evaluated patients having major gastrointestinal surgery, the data expand to orthopedic, cardiothoracic and vascular surgery (7-9). Several meta-analyses have been published on the use of perioperative goal-directed therapy in various types of patients, but conclusions are inconsistent (10-13). Furthermore, perioperative goal-directed therapy is not widely implemented in clinical practice across Europe (14).

Recently, reviews evaluated specific types of surgery or hemodynamic monitoring devices (13,15). To provide a more extensive overview of perioperative goal-directed therapy, the aim of this systematic review was to investigate the benefits and harms of perioperative goal-directed therapy in patients having all types of surgery regardless of the protocol used.

Methods

This systematic review was conducted following our protocol registered on the PROSPERO database (CRD42016035548) following the recommendations of The Cochrane Handbook for Systematic Reviews of Interventions and Grading of Recommendations Assessment, Development and Evaluation (GRADE) (16,17). We reported this review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Table S1) (18).

Eligibility criteria

All trials published in English, irrespective of blinding, publication status or sample size were considered for assessment of benefits and harms. Quasi-randomized trials where the method

of allocating participants to a treatment was not strictly random and observational trials were excluded. All trials evaluating any type of perioperative goal-directed therapy were considered for inclusion, irrespective the type of surgery, the goal-directed therapy algorithm, the types of vasopressors and inotropes used, the hemodynamic variable and its value targeted. All trials were included irrespective of the control intervention.

Perioperative goal-directed therapy was defined by any hemodynamic monitoring along with interventions aimed at optimizing haemodynamics during the perioperative period to achieve a specified predetermined hemodynamic target value. Trials had to describe the interventions, including the hemodynamic monitoring device, the hemodynamic variable and target value and the types and amounts of fluids and/or inotropes used. Such a clear description was required for the intervention group, but not for the control group.

The primary outcome was all-cause mortality (at longest follow-up). The first secondary outcome was serious adverse events (SAE) excluding mortality (to avoid double counts). SAE is a composite outcome summarizing all serious events necessitating an intervention and/or operation and/or prolonged hospital stay excluding mortality according to ICH-GCP definitions.19 Other secondary outcomes were hospital and ICU length of stay. Finally, we also considered the surrogate outcome of the total amounts of fluids administered.

Search strategy

We searched The Cochrane Central Register of Controlled Trials (CENTRAL) of The Cochrane Library, PubMed/MEDLINE, Web of Science and EMBASE. Furthermore, references of identified trials and (systematic) reviews were cross-searched. In addition, Google Scholar (Google Inc.) was used for 'cited reference search' by backwards snowballing (Supplement S1). We used no time restrictions. The final search was performed on 2 May 2018.

Study selection and data extraction

Two authors (TK, RPC) independently selected trials for inclusion. Excluded trials based on full text are listed with reasons for exclusion (Figure 1). Two authors independently performed data extraction, including trial characteristics (lead author, publication year, numbers of patients enrolled), participant characteristics (baseline characteristics, inclusion and exclusion criteria and types of surgery), intervention characteristics (hemodynamic variable and its value targeted, monitoring devices, interventions used) and all outcomes.

Bias risk assessment

The risks of bias were assessed by two authors (TK, RPC), independently, without masking of trial names following The Cochrane Handbook for Systematic Reviews of Interventions (16). Any differences in opinion were resolved through discussion. The following risk of bias domains was assessed: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data, selective outcome reporting and other biases such as vested interests. Trials classified as low risk of bias in all domains were considered

as trials with low overall risk of bias. Trials with one or more of these bias risk domains scored as unclear or high risk of bias were considered trials with high overall risk of bias (16).

Statistical analysis

We followed the instructions in The Cochrane Handbook for Systematic Reviews of Interventions (16) for the intended meta-analyses using Review Manager 5.3.5 (20). For the intended trial sequential analysis (TSA), we used TSA software (version 0.9.5.10 beta) (21).

For dichotomous variables, the risk ratio (RR) with the TSA-adjusted confidence interval (CI) was calculated if there were two or more trials for an outcome. For rare events (<5% in the control group), we calculated odds ratios (OR) or Peto's OR in case of very rare events (<2% in the control group) with TSA-adjusted CI. The outcomes were reported as proportions for each group. For continuous outcomes, we reported mean differences (MD) or weighted mean differences (WMD) with TSA-adjusted CI.

Both a fixed-effect and a random-effects model were used for meta-analysis. In case of discrepancy between the two models, both results were reported. Considering the anticipated abundant clinical heterogeneity (in populations, alternative and control interventions, and settings) the random-effects model was emphasized except if one or two trials dominated the available evidence.

Statistical heterogeneity was explored by the chi-squared test with significance set at P-value of 0.10, and the quantity of heterogeneity was measured by I-squared (22).

We planned on performing the following subgroup analyses: (a) trials with overall low risk of bias compared to trials with overall high risk of bias; (b) the intervention effect in the trials depending on the type of surgery. Only subgroup analyses showing statistical significant test of interactions ($P < 0.05$) were considered to provide evidence of an intervention effect pending the subgroup.

A funnel plot was used to explore small trial bias and to use asymmetry in funnel plot of trial size against treatment effect to assess this bias if data on more than 10 trials were available (16).

GRADE

We used the GRADE system to assess the quality of the body of evidence associated with each of the major outcomes in our review using GRADE software (17). The quality measure of a body of evidence considers within-trial risk of bias, indirectness, heterogeneity, imprecision and risk of publication bias.

Results

Our search strategy identified 2852 unique citations. After removal of duplicates, 1836 remaining hits were screened based on title and abstract. In all, 258 full-text articles were assessed for eligibility. After exclusion of 146 hits, 112 trials were selected for inclusion in this systematic review, of which 15 trials were identified through cross-reference searching (snowballing; Figure 1).

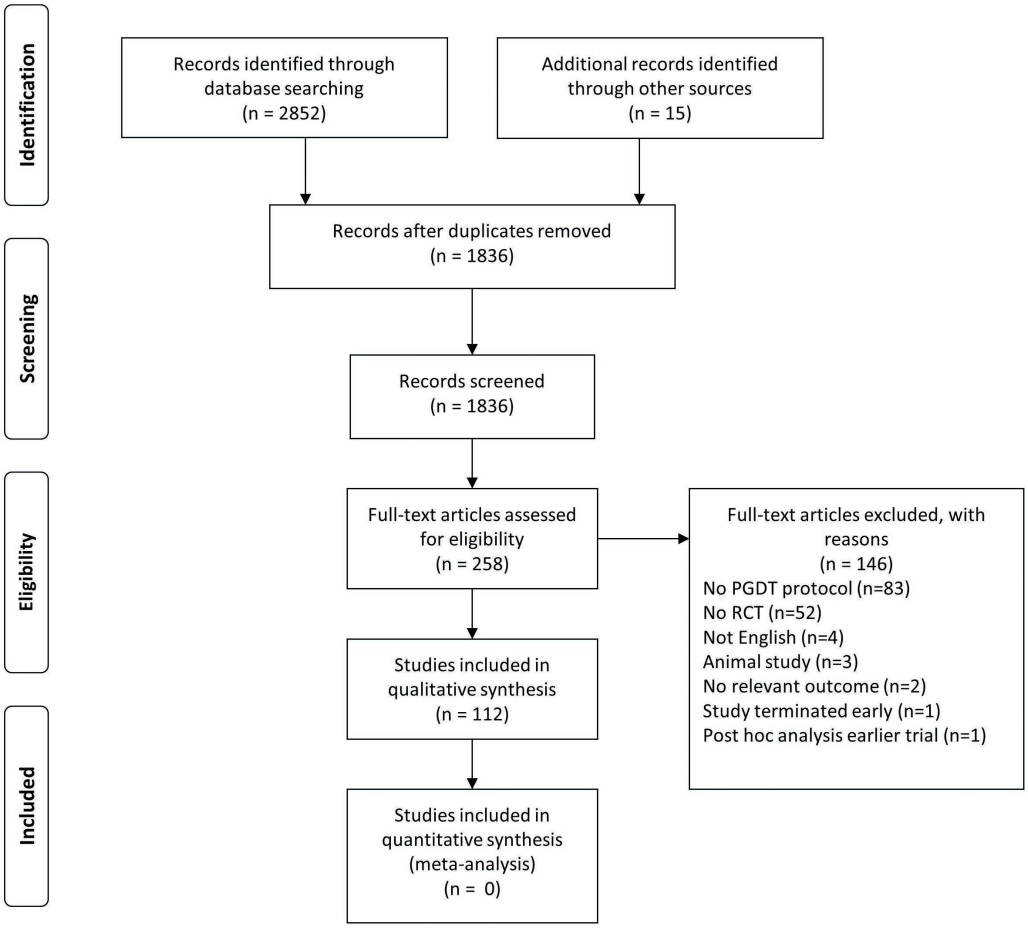


Figure 1. Flowchart of study selection

Characteristics of the included trials

In total, 112 trials with 13 562 patients were included in this systematic review (3,4,7-9,23-129). The characteristics of the 112 included trials are listed in the supplements (Table S2A-E). Nine trials used a three-arm parallel group design; all other trials used a two-arm parallel group design.

Risk of bias

Adequate sequence generation was used in 82 trials (73%), allocation concealment was used in 58 trials (52%), blinding of participants and personnel was used in 23 trials (21%) and blinding of outcome assessors was used in 61 trials (64%). Complete outcome data were reported in 88 trials (79%). There was a low risk of bias regarding selective outcome reporting in 103 trials (92%), and 72 trials (64%) had no other risks of bias. One trial (1%) had a low overall risk of bias and 111 trials (99%) had high overall risk of bias (Figure 2 and Figure S1).

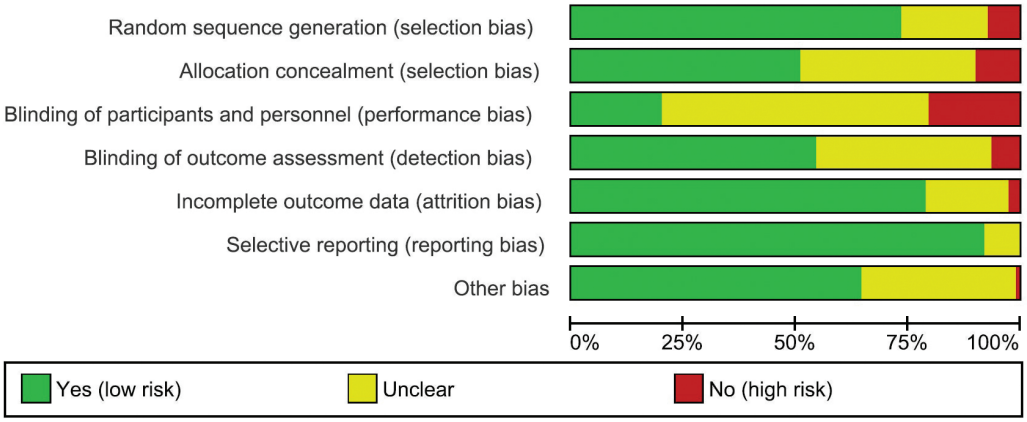


Figure 2. *Risk of bias graph. Review authors' judgements about each risk of bias item presented as percentages across all included trials*

Characteristics of the patients included in the trials

The types of surgery that the patients in the trials underwent included abdominal surgery (43 trials; 38%), cardiothoracic surgery (16 trials; 14%), high-risk surgery (13 trials; 12%), orthopedic surgery (eight trials; 7%), vascular surgery (nine trials; 8%), liver surgery (six trials; 5%), plastic surgery (five trials; 4%), neurosurgery (three trials; 3%), trauma surgery (three trials; 3%), thoracic surgery (two trials; 2%) and other surgery or left unspecified in five trials (4%).

Timing of the intervention

The timing of the conduct of the perioperative goal-directed therapy intervention varied between trials. The majority of 70 trials (63%) only intervened during surgery, one trial (1%) only intervened before surgery, 11 trials (10%) performed the intervention exclusively post-surgery and 30 trials (27%) used extended durations.

Hemodynamic target variable

We identified a total of 30 different variables used as targets in the trials to guide the interventions. These consisted of static variables such as heart rate (HR), mean arterial pressure (MAP), cardiac output (CO), systolic blood pressure (SBP), central venous pressure (CVP), pulmonary artery occlusion pressure (PAOP), extravascular lung water (EVLW), global end diastolic volume (GEDV), intrathoracic blood volume (ITBV), left ventricular stroke work (LVSW), right ventricular

stroke work (RVSW) and stroke volume (SV). Calculated variables based on above-mentioned variables such as oxygen delivery (DO_2), oxygen extraction ratio (O_2ER), pulmonary vascular resistance (PVR), stroke volume index (SVI), systemic vascular resistance index (SVRI) and oxygen consumption (VO_2) were also used. The dynamic variables used included stroke volume variation (SVV), pulse pressure variation (PPV), systolic pressure variation (SPV), corrected flow time (FTc), plethysmography variability index (PVI). Other variables used included base deficit (BD), central venous oxygen saturation ($S_{cv}O_2$), hematocrit (Ht), hemoglobin (Hb), lactate, PaO_2/FiO_2-ratio (PF-ratio) and urine production. Even when identical variables were used, we found different cut-off target values in trials.

Monitoring devices

In total, 18 different hemodynamic monitoring devices were used in the trials (Table S2B). The devices varied from the pulmonary artery catheter to several less-invasive methods such as a central venous catheter already in place or newly introduced for fibreoptic oxymetry (CeVOX), calibrated transpulmonary thermodilution (PiCCO, VolumeView/EV1000), lithium dilution (LiDCO), pulse contour analysis (FloTrac, ProAQT, LiDCOrapid, Datex Ohmeda monitor, IBPplus monitor), endotracheal bioimpedance (ECOM), ultrasound derived techniques (ODM) and non-invasive methods such as thoracic bioimpedance (IQ), bioreactance (NICOM/Cheetah), and non-invasive pulse contour analysis (CNAP, Nexfin, Pulse CO-Oximetry).

Fluids

All trials used some combination of fluids and vasoactive medications to achieve the targeted hemodynamic value. Regarding fluid interventions, 61 trials (54%) used only colloids, 10 trials (9%) used only crystalloids and 32 trials (29%) used combinations of both. A total of nine trials (8%) did not specify the types of fluids used.

Vasopressors and/or inotropes

Different vasopressors and/or inotropes were used in the trials: dobutamine (46 trials; 41%), norepinephrine (34 trials; 30%), ephedrine (27 trials; 24%), dopamine (19 trials; 17%), phenylephrine (16 trials; 14%), epinephrine (15 trials; 13%), dopexamine (five trials; 4%), metaraminol (two trials; 2%), milrinone (two trials; 2%), cafedrine (one trial; 1%), theodrenaline (one trial; 1%) and vasopressin (one trial; 1%). In 28 trials (25%), the types of vasopressors and/or inotropes used were either unspecified or were left at the discretion of the anesthesiologist.

Outcomes

We observed vast clinical heterogeneity included in the trials for type of surgery but also in all individual components of the complex intervention of perioperative goal-directed therapy, including the timing of the intervention, the type of monitoring device, the hemodynamic variables assessed, the hemodynamic value targeted, the types and amounts of fluids given, the types of vasopressors and/or inotropes used (Table 1). Due to such observed clinical heterogeneity, it was deemed inappropriate to pool any of the data into a pooled intervention effect estimate, and therefore, no meta-analysis was conducted following recommendations of The Cochrane Handbook for Systematic Reviews of Interventions (16).

Primary outcome

Eighty-seven trials (78%) reported mortality, the timing of which varied from perioperative mortality up to 1-year mortality or was left unspecified. Furthermore, causation of mortality in this outcome varied from all-cause to specific causes such as cardiac death. Mortality varied from 0% to 34% in the intervention groups and from 0% to 44% in the control groups (Figure 3). Statistical heterogeneity measured by I^2 was 12%. A funnel plot suggested no arguments for bias (Figure 4). We also evaluated subgroups classified according to type of surgery (Figure 5). We did not perform meta-analysis in these subgroups as there still was significant clinical heterogeneity based on all other variables.

Secondary outcomes

The SAEs that were reported varied from one single-specific complication, such as postoperative ileus or kidney failure, to multiple predefined adverse outcomes in several organ systems (Table S2C). None of the trials reported SAE according to the ICH-GCP definitions.

Hospital stay was reported in 78 trials (70%) and ICU stay was reported in 40 trials (36%). Hospital stay varied from 2 to 31 days and length of ICU stay varied from 0 to 15 days (Table S2E).

The types of fluids given in the intervention and control groups varied substantially (Table S2B). Details on total amounts of fluids given were reported in 104 trials (93%). The amounts of fluids given were either reported as total amounts given during the trial or as total amounts of each type of fluid or as mL/kg/h. Total amounts of fluids given ranged from no additional fluid to nearly 22 L of crystalloids in severe trauma patients (Table S2E).

GRADE

The quality of the evidence was assessed as very low for all outcomes based on risk of bias limitations, indirectness, inconsistency, imprecision and other considerations (Table 2).

Table 1. Key characteristics of the included trials

Author	Year	Type of surgery	Device	Timing	Prime goal	Types of fluids	Vasoactive medication
Schultz	1985	Orthopedic	PAC	pre/intra/post	CI 3.0-3.5	N/S	N/S
Shoemaker	1988	High risk	PAC	pre/intra/post	CI >4.5	Crystalloids/Colloids	Inotropes/Vasopressors
Berlauk	1991	Vascular	PAC	pre/intra	CI >2.8	Colloids	Inotropes
Fleming	1992	Trauma	PAC	pre/intra/post	CI >4.52	Crystalloids/Colloids	Inotropes
Boyd	1993	High risk	PAC	pre/intra/post	DO_2 >600	Colloids	Inotropes/Vasopressors
Bishop	1995	Trauma	PAC	pre/intra/post	CI >4.5	Crystalloids/Colloids	Inotropes
Mythen	1995	Cardiothoracic	ODM	intra	Maximum SV	Crystalloids	Vasopressors
Bender	1997	Vascular	PAC	pre/intra/post	CI >2.8	Crystalloids	Inotropes
Sinclair	1997	Orthopedic	ODM	intra	CFT <0.35	Crystalloids/Colloids	N/S
Ziegler	1997	Vascular	PAC	pre	PAOP >12	Crystalloids	Inotropes
Ueno	1998	Liver	PAC	post	CI >4.5	Colloids	Inotropes/Vasopressors
Valentine	1998	Vascular	PAC	pre/intra/post	CI >2.8	Crystalloids	Inotropes
Wilson	1999	High risk	PAC	pre/intra/post	PAOP =12	Crystalloids/Colloids	Inotropes/Vasopressors
Lobo	2000	High risk	PAC	intra/post	DO_2 >600	Crystalloids/Colloids	Inotropes/Vasopressors
Pölönön	2000	Cardiothoracic	PAC	post	SvO_2 >70	N/S	Inotropes/Vasopressors
Velmahos	2000	Trauma	Bioimpedance	pre/intra/post	SBP >100	Crystalloids/Colloids	Inotropes/Vasopressors
Bonazzi	2002	Vascular	PAC	pre/intra/post	CI >3.01	Crystalloids	Inotropes
Conway	2002	Abdominal	ODM	intra	CFT <0.35	Colloids	N/S
Gan	2002	High risk	ODM	intra	CFT <0.35	Crystalloids/Colloids	N/S
Venn	2002	Orthopedic	CVL/ODM	intra	CFT <0.35; CVP >14	Crystalloids/Colloids	N/S
Sandham	2003	High risk	PAC	intra	DO_2 550-600	N/S	N/S
McKendry	2004	Cardiothoracic	ODM	post	SVI >35	Colloids	Inotropes/Vasopressors
Pearse	2005	High risk	LiDCO	post	SVV <10	Colloids	Inotropes/Vasopressors
Szakmany	2005	Abdominal	PiCCO	intra	ITBVI 850-950	Colloids	N/S
Wakeling	2005	Abdominal	ODM	intra	SVV <10	Colloids	N/S
Lobo	2006	High risk	PAC	intra/post	DO_2 >600	Crystalloids/Colloids	Inotropes
Noblett	2006	Abdominal	ODM	intra	CFT <0.35	Colloids	Vasopressors
Donati	2007	Abdominal	CVL	intra/post	O_2ER <27	Colloids	Inotropes
Goepfert	2007	Cardiothoracic	PiCCO	intra/post	GEDVI >640	Colloids	Vasopressors
Lopes	2007	High risk	IBPPlus	intra	PPV <10	Colloids	Vasopressors
Buettner	2008	High risk	PiCCO	intra	SPV <10	Crystalloids/Colloids	Vasopressors
Harten	2008	Abdominal	LiDCO	intra	PPV <10	Colloids	Inotropes
Kapoor	2008	Cardiothoracic	FloTrac	intra	SVV <10	Colloids	Inotropes
Senagore	2009	Abdominal	ODM	intra	SVV >10	Colloids	N/S
Smetkin	2009	Cardiothoracic	PiCCO	intra	ITBVI 850-1000	Colloids	Inotropes/Vasopressors
Benes	2010	Abdominal	FloTrac	intra	SVV <10	Colloids	Inotropes/Vasopressors
Forget	2010	Abdominal	Masimo/PVI	intra	PVI <13	Crystalloids/Colloids	Vasopressors
Jammer	2010	Abdominal	CVL	intra	$ScvO_2$ >75	Colloids	N/S
Jhanji	2010	Abdominal	LiDCO	post	SVV <10	Colloids	Inotropes
Mayer	2010	Abdominal	FloTrac	intra	CI >2.5	Crystalloids/Colloids	Inotropes/Vasopressors
Van der Linden	2010	Vascular	FloTrac	intra	CI >2.5	Colloids	Inotropes
Wenkui	2010	Abdominal	Serum lactate	intra	Lactate <1.6	Colloids	Vasopressors
Cecconi	2011	Orthopedic	FloTrac	intra	SVV <10	Crystalloids/Colloids	Inotropes
Pillai	2011	Abdominal	ODM	intra	SVV <10	Colloids	N/S

Brandstrup	2012	Abdominal	ODM	intra	SVV <10	Colloids	Vasopressors
Challand	2012	Abdominal	ODM	intra	SVV <10	Colloids	N/S
Jain	2012	Plastic	LiDCO	intra	SVV <10	N/S	N/S
Lenkin	2012	Cardiothoracic	PAC/PiCCO	intra	PAOP 12-18; GEDVI 680-850	Crystalloids/Colloids	Inotropes/Vasopressors
Yassen	2012	Liver	PAC	post	SVI >55; RVEDVI >110	Crystalloids	N/S
Zhang Jun	2012	Abdominal	Datex	intra	PPV <11	Crystalloids/Colloids	Vasopressors
Bartha	2013	Orthopedic	LiDCO	pre/intra	DO_2 >600	Colloids	Inotropes/Vasopressors
Bisgaard AAA	2013	Vascular	LiDCO	intra/post	Sustained rise of SVI >10	Colloids	Inotropes/Vasopressors
Bisgaard LLA	2013	Vascular	LiDCO	intra/post	Sustained rise of SVI >10	Colloids	Inotropes/Vasopressors
Bundgaard	2013	Abdominal	ODM	intra	SV rise <10	Colloids	Vasopressors
El Sharkawy	2013	Liver	ODM	intra/post	CFT <0.35	Colloids	N/S
Figus	2013	Plastic	ODM	intra	SV rise <10	Colloids	N/S
Goepfert	2013	Cardiothoracic	PiCCO	intra/post	SVV <10	Crystalloids/Colloids	Vasopressors
Jones	2013	Liver	LiDCO	post	Maximum SV	Colloids	Inotropes
McKenny	2013	Abdominal	ODM	intra	SV rise <10	Colloids	N/S
Ramsingh	2013	Abdominal	FloTrac	intra	SVV <12	Crystalloids/Colloids	N/S
Salzwedel	2013	Abdominal	ProAQT	intra	PPV <10	N/S	Inotropes/Vasopressors
Scheeren	2013	High risk	FloTrac	intra	SVV <10	Colloids	N/S
Zakhaleva	2013	Abdominal	ODM	intra	CFT <0.35	Colloids	N/S
Zhang	2013	Cardiothoracic	FloTrac	intra	SVV <10	Colloids	Inotropes/Vasopressors
Zheng	2013	Abdominal	FloTrac	intra	CI >2.5	Colloids	Vasopressors
Fayed	2014	Liver	ODM	intra	CFT <0.35	Colloids	Vasopressors
Pearse	2014	Abdominal	LiDCO	intra/post	Maximum SV	Colloids	Inotropes
Peng	2014	Orthopedic	FloTrac	intra	SVV <10 or <14	Colloids	Vasopressors
Pestana	2014	Abdominal	NICOM	intra/post	CI >2.5; MAP >65	Crystalloids/Colloids	Inotropes/Vasopressors
Phan	2014	Abdominal	ODM	intra	CFT <0.35	Colloids	N/S
Pösö	2014	Abdominal	TTE/FloTrac	pre/intra	SVV <12	Colloids	Inotropes/Vasopressors
Thomson	2014	Cardiothoracic	LiDCO	post	SV rise <10	Crystalloids/Colloids	N/S
Zeng	2014	Abdominal	FloTrac	intra	SVV <13	Colloids	Vasopressors
Ackland	2015	High risk	LiDCO	post	SV rise <10	Colloids	Inotropes/Vasopressors
Benes	2015	Orthopedic	CNAP	intra	PPV <13	Crystalloids/Colloids	Vasopressors
Colantonio	2015	Abdominal	FloTrac	intra	CI >2.5	Colloids	Inotropes
Correa-Gallego	2015	Liver	FloTrac	intra	SVV <2SD baseline	Colloids	N/S
Fellahi	2015	Cardiothoracic	ECOM	intra	SVV <11	Colloids	Inotropes/Vasopressors
Funk AAA	2015	Vascular	FloTrac	intra	SVV <13	Colloids	Vasopressors
Funk FFR	2015	Plastic	FloTrac	intra/post	SVV <13	Colloids	Vasopressors
Kumar	2015	High risk	FloTrac	intra	CI >2.5	Crystalloids/Colloids	Inotropes/Vasopressors
Lai	2015	Abdominal	LiDCO	intra	SVV <10	Colloids	N/S
Lee	2015	Cardiothoracic	NICOM	intra	ΔSVI <10	Colloids	Inotropes/Vasopressors

Mikor	2015	Abdominal	CeVOX	intra	$ScvO_2$ >75	Colloids	Vasopressors
Moppett	2015	Orthopedic	LiDCO	intra	SV rise <10	Colloids	N/S
Parke	2015	Cardiothoracic	FloTrac	post	SVV <13	N/S	N/S
Xiao	2015	Other	LiDCO	intra	SV rise <10	Crystalloids	Vasopressors
Yu	2015	Abdominal	Masimo/PVI	intra	PVI <13	Crystalloids/Colloids	Vasopressors
Broch	2016	Abdominal	Nexfin	intra/post	PPV <10	Crystalloids/Colloids	Inotropes/Vasopressors
Hand	2016	Plastic	FloTrac	intra	MAP >75	N/S	Inotropes/Vasopressors
Kapoor	2016	Cardiothoracic	FloTrac	post	CI >2.5; CVP >6; SVV <10	N/S	N/S
Kumar	2016	Abdominal	FloTrac	intra	SVV <10%	Crystalloids/Colloids	Inotropes/Vasopressors
Osawa	2016	Cardiothoracic	LiDCO	intra	CI >3.0	Crystalloids	Inotropes
Pavlovic	2016	Other	PiCCO	intra	PPV or SVV <12	Crystalloids/Colloids	Inotropes/Vasopressors
Picard	2016	Neuro	ODM	intra	CFT <0.33	Colloids	Inotropes/Vasopressors
Schmid	2016	Abdominal	PiCCO	intra/post	GEDVI >640	Crystalloids/Colloids	Inotropes/Vasopressors
Elgendy	2017	Abdominal	FloTrac	intra/post	SVV >12%	Colloids	Inotropes/Vasopressors
Gomez-Izquierdo	2017	Abdominal	ODM	intra	SV rise <10%	Colloids	Vasopressors
Kapoor	2017	Cardiothoracic	FloTrac	intra/post	$ScvO_2$ >70%	Crystalloids/Colloids	Inotropes
Kaufmann	2017	Thoracic	ODM	intra	SVV <10%	Crystalloids	Vasopressors
Li	2017	Not specified	FloTrac	intra	VCCI <40%	N/S	N/S
Liang	2017	Other	FloTrac	intra	SVV 8-13%	Colloids	Inotropes
Luo	2017	Neuro	FloTrac	Intra	SVV<15%	Colloids	Inotropes/Vasopressors
Reisinger	2017	Abdominal	ODM	Intra/post	SV rise <10%	Colloids	Vasopressors
Sethi	2017	Abdominal	CVL	intra	CVP 8-12	Crystalloids	Vasopressors
Stens	2017	Abdominal	Nexfin	intra	PPV <12%	Crystalloids/Colloids	Inotropes/Vasopressors
Wu	2017	Neuro	FloTrac	intra	SVV <12%	Colloids	Inotropes/Vasopressors
Xu	2017	Thoracic	FloTrac	Intra	SVV <13%	Colloids	Inotropes/Vasopressors
Calvo-Vecino	2018	Abdominal	ODM	intra	SV rise <10%	Crystalloids/Colloids	N/S
Demirel	2018	Abdominal	Masimo/PVI	intra	PVI <14%	Crystalloids/Colloids	Vasopressors
Kim	2018	Plastic	FloTrac	intra	SVV <12%	Colloids	Inotropes/Vasopressors
Liu	2018	Other	FloTrac	Intra/post	CI >2.5	Crystalloids/Colloids	Inotropes

Abbreviations: CFT, corrected velocity time; CI, cardiac index; CVL, central venous line; CVP, central venous pressure; DO_2, delivery of oxygen; GEDVI, global end diastolic volume index; intra, intraoperative; ITBVI, intra thoracic blood volume index; MAP, mean arterial pressure; N/S, not specified; O_2ER, oxygen extraction ratio; ODM, oesophageal Doppler monitor; PAC, pulmonary artery catheter; PAOP, pulmonary artery occlusion pressure; post, postoperative; PPV, pulse pressure variation; pre, preoperative; PVI, pleth variability index; RVEDVI, right ventricular end diastolic volume index; $S_{cv}O_2$, central venous oxygen saturation; SV, stroke volume; SVI, stroke volume index; SVV, stroke volume variation; VCCI, vena cava collapsibility index

Table 2. *GradePRO summary of findings table of the outcomes of interest*

Certainty assessment [a]							Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		
Mortality at maximum follow up								
87	randomized trials	serious [b]	not serious [c]	very serious [d]	very serious [e]	none [f]	⊕◯◯◯ VERY LOW	CRITICAL
Serious adverse effects - not reported [g]								
-	-	-	-	-	-	-	-	CRITICAL
Hospital length of stay								
78	randomized trials	serious [b]	not serious	very serious [d]	not serious	none	⊕◯◯◯ VERY LOW	IMPORTANT
ICU length of stay								
40	randomized trials	serious [b]	not serious	very serious [d]	not serious	none	⊕◯◯◯ VERY LOW	IMPORTANT
Total fluids administered								
93	randomized trials	serious [b]	not serious	very serious [d,h]	not serious	none	⊕◯◯◯ VERY LOW	NOT IMPORTANT

[a] *Since we did not pool the results in this review, we could not provide a summary of findings consisting of event rates and effect. We merely graded the evidence provided by the individual trials as a whole to illustrate the level of evidence in PGDT studies.*

[b] *Most trials had unclear or high risk of bias in one or more domains rendering it necessary to downgrade the level of evidence.*

[c] *Although there was a great variation in direction of effect, we did not downgrade the level of evidence since most confidence intervals overlap as shown in figure 3 and 5. We computed a P-value and an I^2-value which showed no signs of statistical heterogeneity.*

[d] *While mortality is not a surrogate outcome, we deemed it necessary to classify as very serious due to the significant differences in PGDT algorithms and patient categories.*

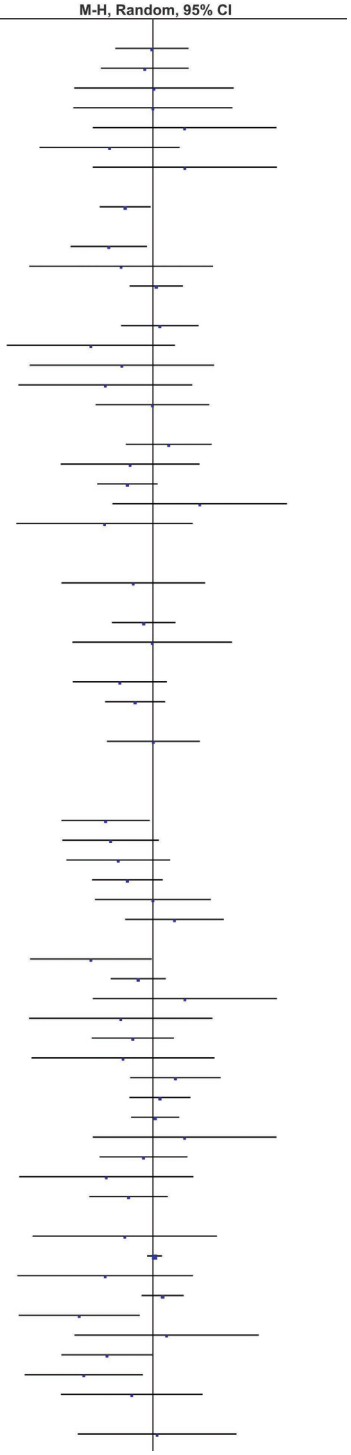
[e] *Most trials were inadequately powered to detect a difference in the outcome mortality considering the very low event rate.*

[f] *Funnel plots showed no clear asymmetry.*

[g] *No trials reported the serious adverse events according to the ICH-GCP definitions and therefore this was not evaluated.*

[h] *Surrogate outcome.*

Study or Subgroup	Intervention Events	Intervention Total	Control Events	Control Total	Weight	Odds Ratio M-H, Random, 95% CI
High risk of bias						
Ackland 2015	5	95	5	92	2.2%	0.97 [0.27, 3.46]
Bartha 2013	3	74	4	75	1.6%	0.75 [0.16, 3.47]
Bender 1997	1	51	1	53	0.5%	1.04 [0.06, 17.08]
Benes 2010	1	60	1	60	0.5%	1.00 [0.06, 16.37]
Benes 2015	1	40	0	40	0.4%	3.08 [0.12, 77.80]
Berlauk 1991	1	45	2	21	0.6%	0.22 [0.02, 2.53]
Bisgaard AAA 2013	1	32	0	32	0.4%	3.10 [0.12, 78.87]
Bisgaard LLA 2013	0	20	0	20		Not estimable
Bishop 1995	9	50	24	65	3.9%	0.38 [0.16, 0.90]
Bonazzi 2002	0	50	0	50		Not estimable
Boyd 1993	3	53	12	54	2.0%	0.21 [0.06, 0.79]
Buettner 2008	0	40	1	40	0.4%	0.33 [0.01, 8.22]
Calvo-Vecino 2018	10	209	9	211	3.7%	1.13 [0.45, 2.83]
Cecconi 2011	0	20	0	20		Not estimable
Challand 2012	5	89	4	90	2.0%	1.28 [0.33, 4.93]
Colantonio 2015	0	38	4	42	0.5%	0.11 [0.01, 2.14]
Conway 2002	0	28	1	29	0.4%	0.33 [0.01, 8.53]
Correa-Gallego 2015	0	69	2	66	0.4%	0.19 [0.01, 3.94]
Donati 2007	2	68	2	67	1.0%	0.98 [0.13, 7.20]
El Sharkawy 2013	0	29	0	30		Not estimable
Elgendy 2017	5	43	3	43	1.6%	1.75 [0.39, 7.85]
Fellahi 2015	1	48	2	44	0.7%	0.45 [0.04, 5.11]
Fleming 1992	8	33	15	34	3.0%	0.41 [0.14, 1.15]
Forget 2010	2	41	0	41	0.4%	5.25 [0.24, 112.88]
Funk AAA 2015	0	20	2	20	0.4%	0.18 [0.01, 4.01]
Goepfert 2007	0	39	0	40		Not estimable
Gómez-Izquierdo 2017	0	64	0	64		Not estimable
Harten 2008	1	14	2	15	0.6%	0.50 [0.04, 6.22]
Jammer 2010	0	121	0	120		Not estimable
Jhanji 2010	9	90	6	45	2.8%	0.72 [0.24, 2.17]
Jones 2013	1	46	1	45	0.5%	0.98 [0.06, 16.12]
Kapoor 2008	0	13	0	14		Not estimable
Kapoor 2016	2	60	6	60	1.4%	0.31 [0.06, 1.60]
Kapoor 2017	6	66	12	76	3.0%	0.53 [0.19, 1.51]
Kumar 2015	0	20	0	20		Not estimable
Lai 2015	3	109	3	111	1.4%	1.02 [0.20, 5.16]
Lee 2015	0	29	0	29		Not estimable
Lenkin 2012	0	20	0	20		Not estimable
Liang 2017	0	30	0	30		Not estimable
Lobo 2000	3	19	9	18	1.5%	0.19 [0.04, 0.88]
Lobo 2006	2	25	7	25	1.3%	0.22 [0.04, 1.21]
Lopes 2007	2	17	5	16	1.1%	0.29 [0.05, 1.80]
Luo 2017	4	73	9	72	2.3%	0.41 [0.12, 1.38]
Mayer 2010	2	30	2	30	0.9%	1.00 [0.13, 7.60]
McKendry 2004	4	85	2	89	1.3%	2.15 [0.38, 12.05]
McKenny 2013	0	51	0	50		Not estimable
Mikor 2015	1	38	8	41	0.8%	0.11 [0.01, 0.94]
Moppett 2015	8	51	15	63	3.5%	0.60 [0.23, 1.54]
Mythen 1995	1	30	0	30	0.4%	3.10 [0.12, 79.23]
Noblett 2006	0	52	1	51	0.4%	0.32 [0.01, 8.06]
Osawa 2016	3	62	6	64	1.8%	0.49 [0.12, 2.06]
Parke 2015	0	70	1	74	0.4%	0.35 [0.01, 8.67]
Pavlovic 2016	5	20	3	23	1.5%	2.22 [0.46, 10.79]
Pearse 2005	9	62	7	60	2.9%	1.29 [0.45, 3.71]
Pearse 2014	12	368	11	365	4.3%	1.08 [0.47, 2.49]
Peng 2014	1	40	0	40	0.4%	3.08 [0.12, 77.80]
Pestana 2014	3	72	4	70	1.6%	0.72 [0.15, 3.33]
Phan 2014	0	50	2	50	0.4%	0.19 [0.01, 4.10]
Polonon 2000	3	196	7	197	1.9%	0.42 [0.11, 1.66]
Poso 2014	0	26	0	20		Not estimable
Reisinger 2017	0	27	1	31	0.4%	0.37 [0.01, 9.46]
Sandham 2003	163	997	155	997	12.8%	1.06 [0.84, 1.35]
Scheeren 2013	0	26	2	26	0.4%	0.18 [0.01, 4.05]
Schmid 2016	22	92	16	88	5.2%	1.41 [0.69, 2.91]
Schultz 1985	1	35	10	35	0.9%	0.07 [0.01, 0.61]
Senagore 2009	1	42	0	22	0.4%	1.63 [0.06, 41.59]
Sethi 2018	2	46	9	48	1.5%	0.20 [0.04, 0.97]
Shoemaker 1988	1	28	18	60	0.9%	0.09 [0.01, 0.69]
Sinclair 1997	1	20	2	20	0.6%	0.47 [0.04, 5.69]
Smetkin 2009	0	20	0	20		Not estimable
Stens 2017	1	81	1	94	0.5%	1.16 [0.07, 18.89]



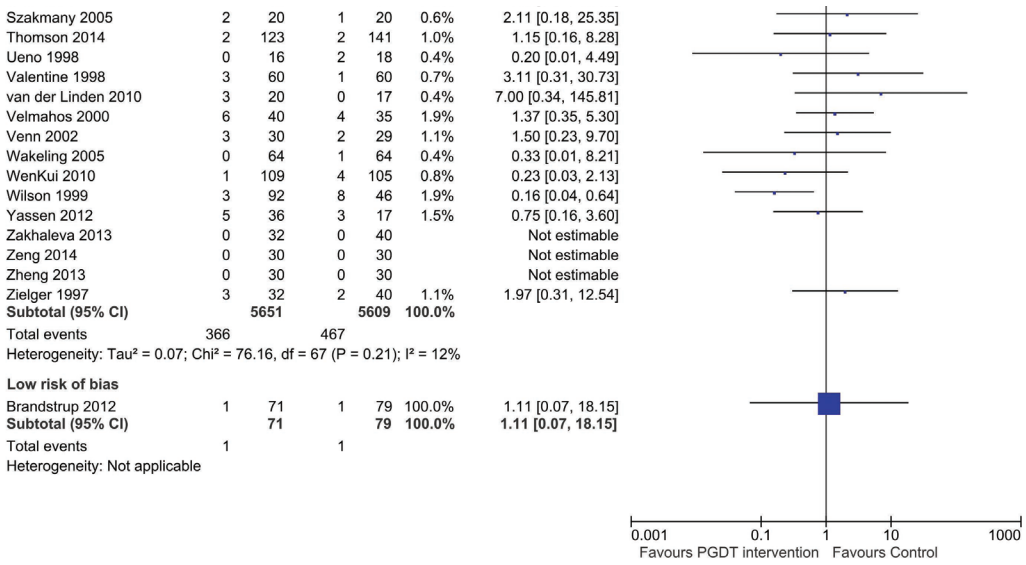


Figure 3. Forest plot of all-cause mortality at maximum follow-up. Subgroups were constructed and classified according to high or low risk of bias. Due to clinical heterogeneity, meta-analysis was deemed inappropriate

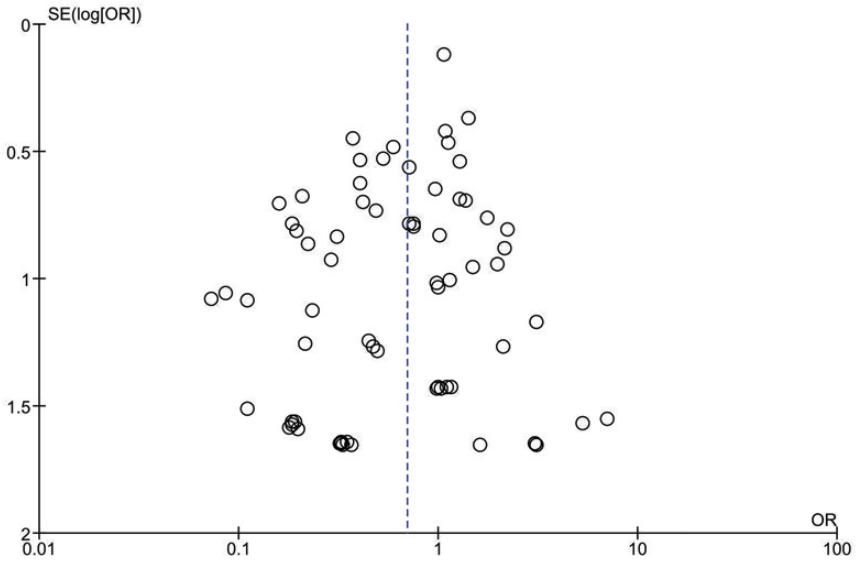
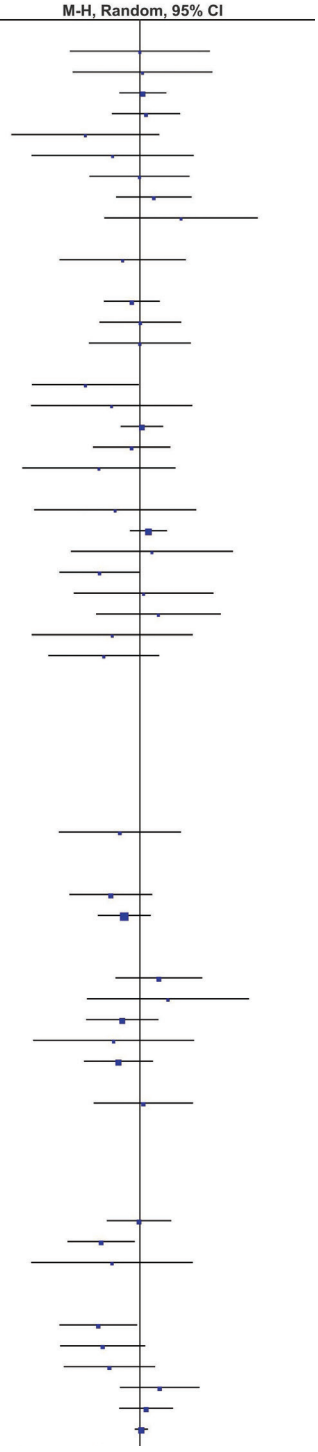


Figure 4. Funnel plot of all-cause mortality at maximum follow-up

Discussion

We conducted a systematic review of perioperative goal-directed therapy and found 112 randomized trials. Only one trial (1%) had a low overall risk of bias. Perioperative goal-directed therapy was tested in patient groups of at least 10 different types of surgery. There was very large clinical heterogeneity, not only in the types of patients but even more in the complex intervention of perioperative goal-directed therapy including its component interventions and also in the outcomes evaluated. This includes variations in the five main components that compose the complex intervention of perioperative goal-directed therapy: the type of monitoring devices (n = 18), the hemodynamic variables (n = 30), the target value (n = 112), the types of fluids (n = 5), and the inotropes and vasopressors used (combined n = 13). Theoretically, various perioperative goal-directed therapy interventions can be composed when accounting for all possibilities of each component intervention (even when ignoring the variations in hemodynamic values targeted). Even though low statistical heterogeneity (I^2 = 12%) was observed, pooling of data was clearly inappropriate considering such clinical heterogeneity, and thus, we refrained from any meta-analysis, including in subgroups, following recommendations of The Cochrane Handbook for Systematic Reviews of Interventions. Even when disregarding the type of surgery, the perioperative goal-directed therapy interventions show substantial heterogeneity in other domains. Final conclusions of any beneficial or harmful effect of perioperative goal-directed therapy in general therefore remain to be decided and homogeneity on either the type of surgery or the five main components of perioperative goal-directed therapy is needed to extrapolate the observations to clinical practice.

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Random, 95% CI
Abdominal surgery						
Benes 2010	1	60	1	60	1.3%	1.00 [0.06, 16.37]
Brandstrup 2012	1	71	1	79	1.3%	1.11 [0.07, 18.15]
Calvo-Vecino 2018	10	209	9	211	11.8%	1.13 [0.45, 2.83]
Challand 2012	5	89	4	90	5.5%	1.28 [0.33, 4.93]
Colantonio 2015	0	38	4	42	1.1%	0.11 [0.01, 2.14]
Conway 2002	0	28	1	29	1.0%	0.33 [0.01, 8.53]
Donati 2007	2	68	2	67	2.5%	0.98 [0.13, 7.20]
Elgendy 2017	5	43	3	43	4.5%	1.75 [0.39, 7.85]
Forget 2010	2	41	0	41	1.1%	5.25 [0.24, 112.88]
Gómez-Izquierdo 2017	0	64	0	64		Not estimable
Harten 2008	1	14	2	15	1.6%	0.50 [0.04, 6.22]
Jammer 2010	0	121	0	120		Not estimable
Jhanji 2010	9	90	6	45	8.3%	0.72 [0.24, 2.17]
Lai 2015	3	109	3	111	3.8%	1.02 [0.20, 5.16]
Mayer 2010	2	30	2	30	2.4%	1.00 [0.13, 7.60]
McKenny 2013	0	51	0	50		Not estimable
Mikor 2015	1	38	8	41	2.2%	0.11 [0.01, 0.94]
Noblett 2006	0	52	1	51	1.0%	0.32 [0.01, 8.06]
Pearse 2014	12	368	11	365	14.5%	1.08 [0.47, 2.49]
Pestana 2014	3	72	4	70	4.3%	0.72 [0.15, 3.33]
Phan 2014	0	50	2	50	1.1%	0.19 [0.01, 4.10]
Poso 2014	0	26	0	20		Not estimable
Reisinger 2017	0	27	1	31	1.0%	0.37 [0.01, 9.46]
Schmid 2016	22	92	16	88	19.1%	1.41 [0.69, 2.91]
Senagore 2009	1	42	0	22	1.0%	1.63 [0.06, 41.59]
Sethi 2018	2	46	9	48	4.0%	0.20 [0.04, 0.97]
Stens 2017	1	81	1	94	1.3%	1.16 [0.07, 18.89]
Szakmany 2005	2	20	1	20	1.6%	2.11 [0.18, 25.35]
Wakeling 2005	0	64	1	64	1.0%	0.33 [0.01, 8.21]
WenKui 2010	1	109	4	105	2.1%	0.23 [0.03, 2.13]
Zakhaleva 2013	0	32	0	40		Not estimable
Zeng 2014	0	30	0	30		Not estimable
Zheng 2013	0	30	0	30		Not estimable
Subtotal (95% CI)		**2305**		**2266**	**100.0%**	
Total events	86		97			
Heterogeneity: Tau² = 0.00; Chi² = 18.52, df = 25 (P = 0.82); I² = 0%						
Cardiothoracic surgery						
Fellahi 2015	1	48	2	44	5.2%	0.45 [0.04, 5.11]
Goepfert 2007	0	39	0	40		Not estimable
Kapoor 2008	0	13	0	14		Not estimable
Kapoor 2016	2	60	6	60	11.3%	0.31 [0.06, 1.60]
Kapoor 2017	6	66	12	76	28.2%	0.53 [0.19, 1.51]
Lee 2015	0	29	0	29		Not estimable
Lenkin 2012	0	20	0	20		Not estimable
McKendry 2004	4	85	2	89	10.3%	2.15 [0.38, 12.05]
Mythen 1995	1	30	0	30	2.9%	3.10 [0.12, 79.23]
Osawa 2016	3	62	6	64	14.9%	0.49 [0.12, 2.06]
Parke 2015	0	70	1	74	3.0%	0.35 [0.01, 8.67]
Polonon 2000	3	196	7	197	16.4%	0.42 [0.11, 1.66]
Smetkin 2009	0	20	0	20		Not estimable
Thomson 2014	2	123	2	141	7.8%	1.15 [0.16, 8.28]
Subtotal (95% CI)		**861**		**898**	**100.0%**	
Total events	22		38			
Heterogeneity: Tau² = 0.00; Chi² = 4.67, df = 8 (P = 0.79); I² = 0%						
High-risk surgery						
Ackland 2015	5	95	5	92	9.9%	0.97 [0.27, 3.46]
Boyd 1993	3	53	12	54	9.6%	0.21 [0.06, 0.79]
Buettner 2008	0	40	1	40	3.0%	0.33 [0.01, 8.22]
Kumar 2015	0	20	0	20		Not estimable
Liang 2017	0	30	0	30		Not estimable
Lobo 2000	3	19	9	18	8.3%	0.19 [0.04, 0.88]
Lobo 2006	2	25	7	25	7.5%	0.22 [0.04, 1.21]
Lopes 2007	2	17	5	16	6.9%	0.29 [0.05, 1.80]
Pavlovic 2016	5	20	3	23	8.1%	2.22 [0.46, 10.79]
Pearse 2005	9	62	7	60	11.4%	1.29 [0.45, 3.71]
Sandham 2003	163	997	155	997	16.8%	1.06 [0.84, 1.35]



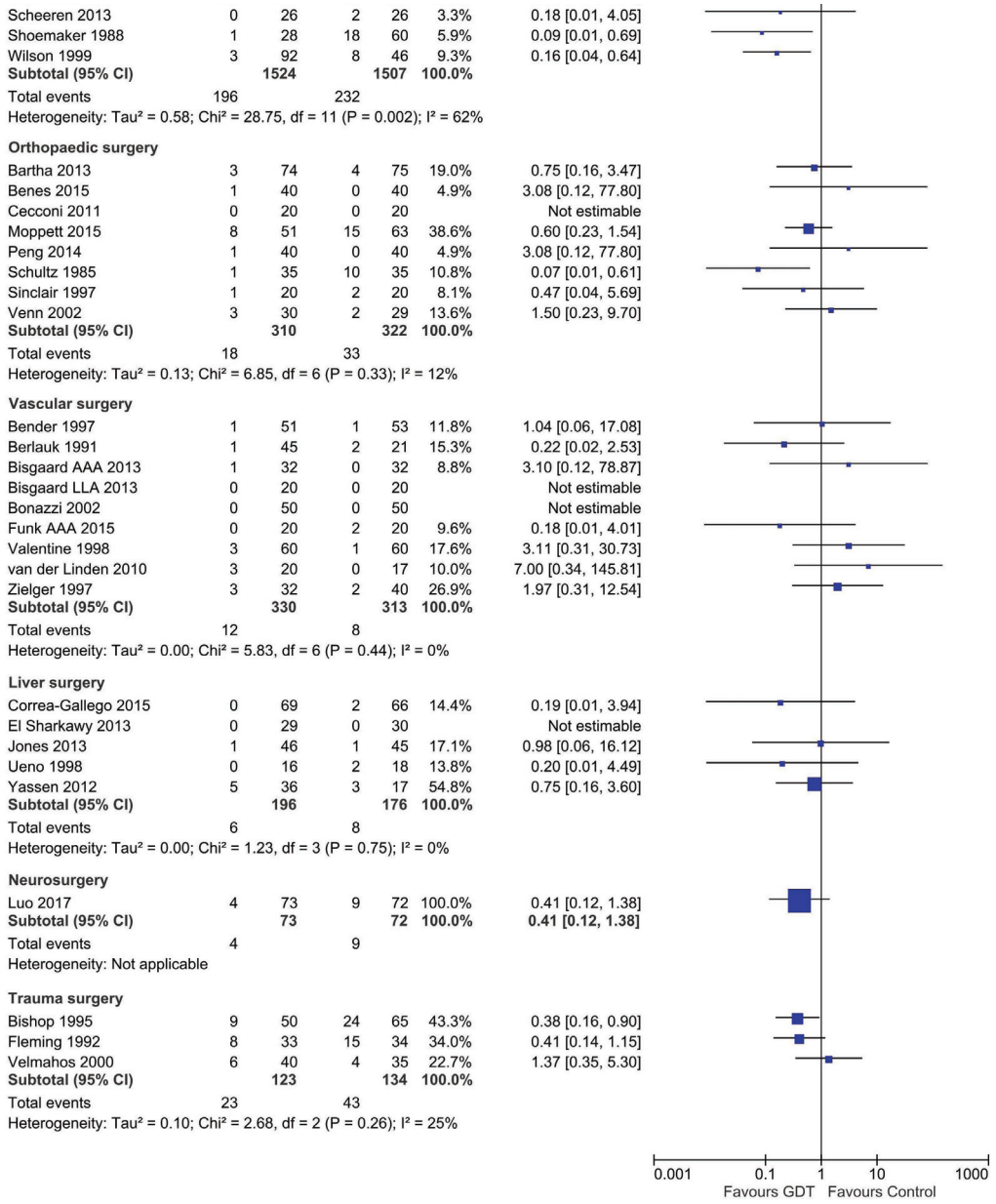


Figure 5. Forest plot of all-cause mortality at maximum follow-up with subgroups classified according to type of surgery. Due to clinical heterogeneity, meta-analysis was deemed inappropriate

Many meta-analyses have been conducted for evaluation of perioperative goal-directed therapy for specific patient categories (10-13). Others narrowed their research question to specific perioperative goal-directed therapy interventions, limiting clinical heterogeneity in the interventions used. For example, one systematic review only evaluated LidCO-based fluid management during hip surgery (83). Such narrowing has advantages in terms of population and device definitions at the costs of reduced power and reduced generalizability of the perioperative goal-directed therapy intervention. However, even within one trial or within a focused systematic review, patients may differ in several aspects. We chose to include trials irrespective of type of surgery because national guidelines recommend the use of perioperative goal-directed therapy in major surgery (5,6).

Experts recognize that evaluation of the perioperative goal-directed therapy intervention is impeded by large clinical heterogeneity, especially considering the changes over time (130,131). Earlier perioperative goal-directed therapy trials targeted supranormal hemodynamic values compared to a more restrictive fluid regimen applied in more recent trials. Researchers have adopted lessons learnt from earlier trials into new trials so that the perioperative goal-directed therapy intervention might have evolved (132). Furthermore, general advancements in medicine might have reduced mortality both in the intervention group and in the control group over time. Therefore, merging data from earlier trials with data from more recent trials into one pooled intervention effect estimate may be inappropriate.

During this systematic review, we made several deviations from our published protocol. First, there were no data on the composite outcome SAEs, and we were therefore unable to report this. We recognize the difficulty associated with the definition and registration of SAEs and did not take into account that most of the included trials were published before SAE definitions by ICH-GCP were developed. Third, we also included trials if most of the perioperative goal-directed therapy protocol was described. Last, since the total amount of administered fluids and length of hospital stay were reported inconsistently, we could not perform statistical analyses on both outcomes.

We aimed to present the extracted data for this systematic review as accessible as possible. However, we realize that, without meta-analyses, we ask the reader to analyze the data on their own using the supplementary table in order to fully appreciate this work (Table S2).

In general, randomized trials with low overall risk of bias are needed before the conclusion can be drawn that any intervention is beneficial (133). The beneficial effects of perioperative goal-directed therapy need confirmation by trials with low overall risks of bias (134). This systematic review, as in general, is obviously limited by the quality of the included trials (135): only one trial had low overall risks of bias (33).

Future research on perioperative goal-directed therapy will most likely benefit from a consensus on definitions of the standard combination of components that form the complex composite intervention of perioperative goal-directed therapy as well as on its individual components. Furthermore, initiatives such as Core Outcome Measures in Effectiveness Trials (COMET) have agreed on standard sets of outcomes for a specific area of research (136) and researchers of perioperative goal-directed therapy may likewise benefit from such a standard set of outcomes (133).

Conclusion

Perioperative goal-directed therapy has been tested in many randomized trials. Meta-analysis of data was considered inappropriate due to the vast amount of clinical heterogeneity. Therefore, a uniform conclusion on the effect of perioperative goal-directed therapy, in general, on outcome remains to be elucidated.

Future research on perioperative goal-directed therapy probably will benefit from consensus definitions and reports should at least include all perioperative goal-directed therapy components. Any reported pooled intervention effect estimates on perioperative goal-directed therapy should be interpreted with great caution due to the current heterogeneity. Various multicenter randomized trials on perioperative goal-directed therapy are ongoing (www.clinicaltrials.gov), including the large Fluid Optimisation in Emergency LAparotomy (FLO-ELA) trial targeting 7646 patients (www.floela.org) and the OPtimisation of Peri-operaTive Cardiovascular Management to Improve Surgical outcomE II (OPTIMISE II) trial targeting 2502 patients (www.optimiseii.org).

Supplementary material

Supplements are available online:
https://doi.org/10.1111/aas.13212

References

1 Cove ME, Pinsky MR. Perioperative hemodynamic monitoring. Best Pract Res Clin Anaesthesiol. 2012;26(4):453–62.
2 Shin CH, Long DR, McLean D, Grabitz SD, Ladha K, Timm FP, et al. Effects of Intraoperative Fluid Management on Postoperative Outcomes: A Hospital Registry Study. Ann Surg. 2018 Jun 10;267(6):1084–92.
3 Shoemaker WC, Appel PL, Kram HB, Waxman K, Lee TS. Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients. Chest. 1988 Dec;94(6):1176–86.
4 Sandham JD, Hull RD, Brant RF, Knox L, Pineo GF, Doig CJ, et al. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. N Engl J Med. 2003 Jan 2;348(1):5–14.
5 National Institute for Health and Clinical Excellence. Medical technologies guidance MTG3: CardioQ-ODM oesophageal Doppler monitor. March 2011. https://www.nice.org.uk/mtg3.
6 Agency for Healthcare Research and Quality. Esophageal Doppler ultrasound-based cardiac output monitoring for real-time therapeutic management of hospitalized patients: a review. January 2007. http://www.cms.gov/determinationprocess/downloads/id45TA.pdf.
7 Cecconi M, Fasano N, Langiano N, Divella M, Costa MG, Rhodes A, et al. Goal-directed haemodynamic therapy during elective total hip arthroplasty under regional anaesthesia. Crit Care. 2011;15(3):R132.
8 Kapoor PM, Kakani M, Chowdhury U, Choudhury M, Kiran U. Early goal-directed therapy in moderate to high-risk cardiac surgery patients. Ann Card Anaesth. 2008;11(1):27–34.
9 Bisgaard J, Gilsaa T, Rønholm E, Toft P. Haemodynamic optimisation in lower limb arterial surgery: room for improvement? Acta Anaesthesiol Scand. 2013 Feb;57(2):189–98.
10 Giglio M, Dalfino L, Puntillo F, Rubino G, Marucci M, Brienza N. Haemodynamic goal-directed therapy in cardiac and vascular surgery. A systematic review and meta-analysis. Interact Cardiovasc Thorac Surg. 2012;15(5):878–87.
11 Cecconi M, Corredor C, Arulkumaran N, Abuella G, Ball J, Grounds RM, et al. Clinical review: Goal-directed therapy-what is the evidence in surgical patients? The effect on different risk groups. Crit Care. 2012 Mar 5;17(2):209.
12 Benes J, Giglio M, Brienza N, Michard F. The effects of goal-directed fluid therapy based on dynamic parameters on post-surgical outcome: A meta-analysis of randomized controlled trials. Crit Care. 2014;18(5):1–11.
13 Michard F, Giglio MT, Brienza N. Perioperative goal-directed therapy with uncalibrated pulse contour methods: Impact on fluid management and postoperative outcome. Br J Anaesth. 2017 Jun 11;119(1):22–30.
14 Ahmad T, Beilstein CM, Aldecoa C, Moreno RP, Molnár Z, Novak-Jankovic V, et al. Variation in haemodynamic monitoring for major surgery in European nations: secondary analysis of the EuSOS dataset. Perioper Med. 2015;4(1):8.
15 Sun Y, Chai F, Pan C, Romeiser JL, Gan TJ. Effect of perioperative goal-directed hemodynamic therapy on postoperative recovery following major abdominal surgery-a systematic review and meta-analysis of randomized controlled trials. Crit Care. 2017 Jun 12;21(1):141.
16 Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011.
17 Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, et al. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008 May 3;336(7651):995–8.
18 Idu MM, Buth J, Cuypers P, Hop WCJ, Van de Pavoordt EDWM, Tordoir JMH. Economising vein-graft surveillance programs. Eur J Vasc Endovasc Surg. 1998 Jul 21;15(5):432–8.
19 ICH Harmonised Tripartite Guideline (1995) Clinical safety data management: definitions and standards for expedited reporting. Fed Regist 60:11284-11287.

20 Review Manager (RevMan). Version 5.3 for Windows. (updated to 5.3.5 on 13 June 2014) The Nordic Cochrane Centre, The Cochrane Collaboration: Copenhagen, 2008.

21 Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive - Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. Int J Epidemiol. 2009 Feb;38(1):287–98.

22 Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002 Jun 15;21(11):1539–58.

23 Berlauk JF, Abrams JH, Gilmour IJ, O'Connor SR, Knighton DR, Cerra FB. Preoperative optimization of cardiovascular hemodynamics improves outcome in peripheral vascular surgery. A prospective, randomized clinical trial. Ann Surg. 1991 Sep;214(3):289–97; discussion 298-9.

24 Wilson J, Woods I, Fawcett J, Whall R, Dibb W, Morris C, et al. Reducing the risk of major elective surgery: randomised controlled trial of preoperative optimisation of oxygen delivery. BMJ. 1999 Apr 24;318(7191):1099–103.

25 Wu J, Ma Y, Wang T, Xu G, Fan L, Zhang Y. Goal-directed fluid management based on the auto-calibrated arterial pressure-derived stroke volume variation in patients undergoing supratentorial neoplasms surgery. Int J Clin Exp Med. 2017;10(2):3106–14.

26 Xiao W, Duan Q, Zhao L, Chi X, Wang F, Ma D, et al. Goal-directed fluid therapy may improve hemodynamic stability in parturient women under combined spinal epidural anesthesia for cesarean section and newborn well-being. J Obstet Gynaecol Res. 2015 Oct;41(10):1547–55.

27 Xu H, Shu SH, Wang D, Chai XQ, Xie YH, Zhou W De. Goal-directed fluid restriction using stroke volume variation and cardiac index during one-lung ventilation: A randomized controlled trial. J Thorac Dis. 2017;9(9):2992–3004.

28 Yassen AM. Pressure versus volume indices to guide fluid infusion early after living donor liver transplantation: A prospective randomized controlled trial. Egypt J Anaesth. 2012;28(3):223–30.

29 Yu Y, Dong J, Xu Z, Shen H, Zheng J. Pleth variability index-directed fluid management in abdominal surgery under combined general and epidural anesthesia. J Clin Monit Comput. 2015 Feb;29(1):47–52.

30 Zakhaleva J, Tam J, Denoya PI, Bishawi M, Bergamaschi R. The impact of intravenous fluid administration on complication rates in bowel surgery within an enhanced recovery protocol: A randomized controlled trial. Color Dis. 2013 Jul;15(7):892–9.

31 Zeng K, Li Y, Liang M, Gao Y, Cai H, Lin C. The influence of goal-directed fluid therapy on the prognosis of elderly patients with hypertension and gastric cancer surgery. Drug Des Devel Ther. 2014;8:2113–9.

32 Zhang J, Chen C, Lei X, Feng Z, Zhu S. Goal-directed fluid optimization based on stroke volume variation and cardiac index during one-lung ventilation in patients undergoing thoracoscopy lobectomy operations: a pilot study. Clinics. 2013 Jul;68(7):1065–70.

33 Zhang J, Qiao H, He Z, Wang Y, Che X, Liang W. Intraoperative fluid management in open gastrointestinal surgery: goal-directed versus restrictive. Clinics. 2012 Oct;67(10):1149–55.

34 Bisgaard J, Gilsaa T, Rønholm E, Toft P. Optimising stroke volume and oxygen delivery in abdominal aortic surgery: a randomised controlled trial. Acta Anaesthesiol Scand. 2013 Feb;57(2):178–88.

35 Zheng H, Guo H, Ye J, Chen L, Ma H. Goal-Directed Fluid Therapy in Gastrointestinal Surgery in Older Coronary Heart Disease Patients: Randomized Trial. World J Surg. 2013 Dec;37(12):2820–9.

36 Ziegler DW, Wright JG, Choban PS, Flancbaum L. A prospective randomized trial of preoperative "optimization" of cardiac function in patients undergoing elective peripheral vascular surgery. Surgery. 1997 Sep;122(3):584–92.

37 Bishop MH, Shoemaker WC, Appel PL, Meade P, Ordog GJ, Wasserberger J, et al. Prospective, Randomized Trial of Survivor Values of Cardiac Index, Oxygen Delivery, and Oxygen Consumption as Resuscitation Endpoints in Severe Trauma. J Trauma Inj Infect Crit Care. 1995 May;38(5):780–7.

38 Bonazzi M, Gentile F, Biasi GM, Migliavacca S, Esposti D, Cipolla M, et al. Impact of perioperative haemodynamic monitoring on cardiac morbidity after major vascular surgery in low risk patients. A randomised pilot trial. Eur J Vasc Endovasc Surg. 2002 May;23(5):445–51.

39 Boyd O, Grounds RM, Bennett ED. A randomized clinical trial of the effect of deliberate perioperative increase of oxygen delivery on mortality in high-risk surgical patients. JAMA. 1993 Dec 8;270(22):2699–707.

40 Brandstrup B, Svendsen PE, Rasmussen M, Belhage B, Rodt SÅ, Hansen B, et al. Which goal for fluid therapy during colorectal surgery is followed by the best outcome: Near-maximal stroke volume or zero fluid balance? Br J Anaesth. 2012 Aug;109(2):191–9.

41 Broch O, Carstens A, Gruenewald M, Nischelsky E, Vellmer L, Bein B, et al. Non-invasive hemodynamic optimization in major abdominal surgery: a feasibility study. Minerva Anestesiol. 2016 Nov;82(11):1158–69.

42 Buettner M, Schummer W, Huettemann E, Schenke S, van Hout N, Sakka SG. Influence of systolic-pressure-variation-guided intraoperative fluid management on organ function and oxygen transport. Br J Anaesth. 2008 Aug;101(2):194–9.

43 Bundgaard-Nielsen M, Jans Ø, Müller RG, Korshin A, Ruhnau B, Bie P, et al. Does goal-directed fluid therapy affect postoperative orthostatic intolerance?: A randomized trial. Anesthesiology. 2013 Oct;119(4):813–23.

44 Calvo-Vecino JM, Ripollés-Melchor J, Mythen MG, Casans-Francés R, Balik A, Artacho JP, et al. Effect of goal-directed haemodynamic therapy on postoperative complications in low–moderate risk surgical patients: a multicentre randomised controlled trial (FEDORA trial). Br J Anaesth. 2018 Apr;120(4):734–44.

45 Challand C, Struthers R, Sneyd JR, Erasmus PD, Mellor N, Hosie KB, et al. Randomized controlled trial of intraoperative goal-directed fluid therapy in aerobically fit and unfit patients having major colorectal surgery. Br J Anaesth. 2012 Jan;108(1):53–62.

46 Colantonio L, Claroni C, Fabrizi L, Marcelli ME, Sofra M, Giannarelli D, et al. A randomized trial of goal directed vs. standard fluid therapy in cytoreductive surgery with hyperthermic intraperitoneal chemotherapy. J Gastrointest Surg. 2015 Apr;19(4):722–9.

47 Conway DH, Mayall R, Abdul-Latif MS, Gilligan S, Tackaberry C. Randomised controlled trial investigating the influence of intravenous fluid titration using oesophageal Doppler monitoring during bowel surgery. Anaesthesia. 2002 Sep;57(9):845–9.

48 Correa-Gallego C, Tan KS, Arslan-Carlon V, Gonen M, Denis SC, Langdon-Embry L, et al. Goal-Directed Fluid Therapy Using Stroke Volume Variation for Resuscitation after Low Central Venous Pressure-Assisted Liver Resection: A Randomized Clinical Trial. J Am Coll Surg. 2015 Aug;221(2):591–601.

49 Demirel İ, Bolat E, Altun AY, Özdemir M, Beştaş A. Efficacy of Goal-Directed Fluid Therapy via Pleth Variability Index During Laparoscopic Roux-en-Y Gastric Bypass Surgery in Morbidly Obese Patients. Obes Surg. 2018 Feb;28(2):358–63.

50 Donati A, Loggi S, Preiser JC, Orsetti G, Münch C, Gabbanelli V, et al. Goal-directed intraoperative therapy reduces morbidity and length of hospital stay in high-risk surgical patients. Chest. 2007 Dec;132(6):1817–24.

51 Yassen K, El Sharkawy O, Fayed N, Abd Elhafez H, Refaat E, Ibraheem AE, et al. Transoesophageal Doppler compared to central venous pressure for perioperative hemodynamic monitoring and fluid guidance in liver resection. Saudi J Anaesth. 2013 Oct;7(4):378.

52 Elgendy MA, Esmat IM, Kassim DY. Outcome of intraoperative goal-directed therapy using Vigileo/FloTrac in high-risk patients scheduled for major abdominal surgeries: A prospective randomized trial. Egypt J Anaesth. 2017;33(3):263–9.

53 Fayed NA, Murad WS. Goal directed preemptive ephedrine attenuates the reperfusion syndrome during adult living donor liver transplantation. Egypt J Anaesth. 2014 Apr;30(2):187–95.

54 Fellahi J-L, Brossier D, Dechanet F, Fischer M-O, Saplacan V, Gérard J-L, et al. Early goal-directed therapy based on endotracheal bioimpedance cardiography: a prospective, randomized controlled study in coronary surgery. J Clin Monit Comput. 2015 Jun;29(3):351–8.

55 Figus A, Wade RG, Oakey S, Ramakrishnan V V. Intraoperative esophageal Doppler hemodynamic monitoring in free perforator flap surgery. Ann Plast Surg. 2013 Mar;70(3):301–7.

56 Fleming A, Bishop M, Shoemaker W, Appel P, Sufficool W, Kuvhenguwha A, et al. Prospective trial of supranormal values as goals of resuscitation in severe trauma. Arch Surg. 1992 Oct;127(10):1175–9; discussion 1179-81.

57 Forget P, Lois F, de Kock M. Goal-directed fluid management based on the pulse oximeter-derived pleth variability index reduces lactate levels and improves fluid management. Anesth Analg. 2010 Oct;111(4):910–4.

58 Funk DJ, HayGlass KT, Koulack J, Harding G, Boyd A, Brinkman R. A randomized controlled trial on the effects of goal-directed therapy on the inflammatory response open abdominal aortic aneurysm repair. Crit Care. 2015 Jun 10;19(1):247.

59 Funk D, Bohn J, Mutch W, Hayakawa T, Buchel EW. Goal-directed fluid therapy for microvascular free flap reconstruction following mastectomy: A pilot study. Plast Surg (Oakville, Ont). 2015;23(4):231–4.

60 Gan TJ, Soppitt A, Maroof M, El-Moalem H, Robertson KM, Moretti E, et al. Goal-directed intraoperative fluid administration reduces length of hospital stay after major surgery. Anesthesiology. 2002 Oct;97(4):820–6.

61 Goepfert MSG, Reuter DA, Akyol D, Lamm P, Kilger E, Goetz AE. Goal-directed fluid management reduces vasopressor and catecholamine use in cardiac surgery patients. Intensive Care Med. 2007 Jan;33(1):96–103.

62 Goepfert MS, Richter HP, Zu Eulenburg C, Gruetzmacher J, Rafflenbeul E, Roeher K, et al. Individually optimized hemodynamic therapy reduces complications and length of stay in the intensive care unit: A prospective, randomized controlled trial. Anesthesiology. 2013 Oct;119(4):824–36.

63 Gómez-Izquierdo JC, Trainito A, Mirzakandov D, Stein BL, Liberman S, Charlebois P, et al. Goal-directed Fluid Therapy Does Not Reduce Primary Postoperative Ileus after Elective Laparoscopic Colorectal Surgery: A Randomized Controlled Trial. Anesthesiology. 2017;127(1):36–49.

64 Hand WR, Stoll WD, McEvoy MD, McSwain JR, Sealy CD, Skoner JM, et al. Intraoperative goal-directed hemodynamic management in free tissue transfer for head and neck cancer. Head Neck. 2016 Apr;38(6):E1974–80.

65 Harten J, Crozier JEM, McCreath B, Hay A, McMillan DC, McArdle CS, et al. Effect of intraoperative fluid optimisation on renal function in patients undergoing emergency abdominal surgery: A randomised controlled pilot study (ISRCTN 11799696). Int J Surg. 2008 Jun;6(3):197–204.

66 Jain AK, Khan AM. Stroke volume variation as a guide for fluid resuscitation in patients undergoing large-volume liposuction. Plast Reconstr Surg. 2012 Sep;130(3):462e-9e.

67 Jammer I, Ulvik A, Erichsen C, Lødemel O, Østgaard G. Does central venous oxygen saturation-directed fluid therapy affect postoperative morbidity after colorectal surgery?: A randomized assessor-blinded controlled trial. Anesthesiology. 2010 Nov;113(5):1072–80.

68 Jhanji S, Vivian-Smith A, Lucena-Amaro S, Watson D, Hinds CJ, Pearse RM. Haemodynamic optimisation improves tissue microvascular flow and oxygenation after major surgery: A randomised controlled trial. Crit Care. 2010;14(4):R151.

69 Jones C, Kelliher L, Dickinson M, Riga A, Worthington T, Scott MJ, et al. Randomized clinical trial on enhanced recovery versus standard care following open liver resection. Br J Surg. 2013 Jul;100(8):1015–24.

70 Kapoor P, Magoon R, Rawat R, Mehta Y. Perioperative utility of goal-directed therapy in high-risk cardiac patients undergoing coronary artery bypass grafting: "A clinical outcome and biomarker-based study." Ann Card Anaesth. 2016;19(4):638.

71 Kapoor P, Magoon R, Rawat R, Mehta Y, Taneja S, Ravi R, et al. Goal-directed therapy improves the outcome of high-risk cardiac patients undergoing off-pump coronary artery bypass. Ann Card Anaesth. 2017;20(1):83–9.

72 Kaufmann KB, Stein L, Bogatyreva L, Ulbrich F, Kaifi JT, Hauschke D, et al. Oesophageal Doppler guided goal-directed haemodynamic therapy in thoracic surgery-a single centre randomized parallel-arm trial. Br J Anaesth. 2017;118(6):852–61.

73 Kim HJ, Kim EJ, Lee HJ, Min JY, Kim TW, Choi EC, et al. Effect of goal-directed haemodynamic therapy in free flap reconstruction for head and neck cancer. Acta Anaesthesiol Scand. 2018 Mar 25;1–12.

74 Kumar L, Kanneganti YS, Rajan S. Outcomes of implementation of enhanced goal directed therapy in high-risk patients undergoing abdominal surgery. Indian J Anaesth. 2015 Apr;59(4):228–33.

75 Ackland GL, Iqbal S, Paredes LG, Toner A, Lyness C, Jenkins N, et al. Individualised oxygen delivery targeted haemodynamic therapy in high-risk surgical patients: A multicentre, randomised, double-blind, controlled, mechanistic trial. Lancet Respir Med. 2015 Jan;3(1):33–41.

76 Kumar L, Rajan S, Baalachandran R. Outcomes associated with stroke volume variation versus central venous pressure guided fluid replacements during major abdominal surgery. J Anaesthesiol Clin Pharmacol. 2016;32(2):182.

77 Lai CW, Starkie T, Creanor S, Struthers RA, Portch D, Erasmus PD, et al. Randomized controlled trial of stroke volume optimization during elective major abdominal surgery in patients stratified by aerobic fitness. Br J Anaesth. 2015 Oct;115(4):578–89.

78 Lee S, Lee SH, Chang BC, Shim JK. Efficacy of goal-directed therapy using bioreactance cardiac output monitoring after valvular heart surgery. Yonsei Med J. 2015 Jul;56(4):913–20.

79 Lenkin AI, Kirov MY, Kuzkov V V., Paromov K V., Smetkin AA, Lie M, et al. Comparison of goal-directed hemodynamic optimization using pulmonary artery catheter and transpulmonary thermodilution in combined valve repair: A randomized clinical trial. Crit Care Res Pract. 2012;2012:821218.

80 Li B, Zhang L, Zhang S, Wang B. Application of ultrasound in goal-directed fluid management of anesthesia in elderly patients. Biomed Res. 2017;2017(Special Issue HealthScienceandBioConvergence TechnologyEdition-II):S482–6.

81 Liang M, Li Y, Lin L, Lin X, Wu X, Gao Y, et al. Effect of goal-directed fluid therapy on the prognosis of elderly patients with hypertension receiving plasmakinetic energy transurethral resection of prostate. Int J Clin Exp Med. 2017;10(1):1290–6.

82 Liu T-J, Zhang J-C, Gao X-Z, Tan Z-B, Wang J-J, Zhang P-P, et al. Clinical research of goal-directed fluid therapy in elderly patients with radical resection of bladder cancer. J Cancer Res Ther. 2018;14(8):173.

83 Lobo SM, Lobo FR, Polachini CA, Patini DS, Yamamoto AE, de Oliveira NE, et al. Prospective, randomized trial comparing fluids and dobutamine optimization of oxygen delivery in high-risk surgical patients [ISRCTN42445141]. Crit Care. 2006;10(3):R72.

84 Lobo SMA, Salgado PF, Castillo VGT, Borim AA, Polachini CA, Palchetti JC, et al. Effects of maximizing oxygen delivery on morbidity and mortality in high-risk surgical patients. Crit Care Med. 2000 Oct;28(10):3396–404.

85 Lopes MR, Oliveira MA, Pereira VOS, Lemos IPB, Auler JOC, Michard F. Goal-directed fluid management based on pulse pressure variation monitoring during high-risk surgery: A pilot randomized controlled trial. Crit Care. 2007;11(5):R100.

86 Bartha E, Arfwedson C, Imnell A, Fernlund ME, Andersson LE, Kalman S. Randomized controlled trial of goal-directed haemodynamic treatment in patients with proximal femoral fracture. Br J Anaesth. 2013 Apr;110(4):545–53.

87 Luo J, Xue J, Liu J, Liu B, Liu L, Chen G. Goal-directed fluid restriction during brain surgery: a prospective randomized controlled trial. Ann Intensive Care. 2017;7(1):16.

88 Mayer J, Boldt J, Mengistu AM, Röhm KD, Suttner S. Goal-directed intraoperative therapy based on autocalibrated arterial pressure waveform analysis reduces hospital stay in high-risk surgical patients: a randomized, controlled trial. Crit Care. 2010;14(1):R18.

89 McKendry M, McGloin H, Saberi D, Caudwell L, Brady AR, Singer M. Randomised controlled trial assessing the impact of a nurse delivered, flow monitored protocol for optimisation of circulatory status after cardiac surgery. BMJ. 2004 Jul 31;329(7460):258.
90 McKenny M, Conroy P, Wong A, Farren M, Gleeson N, Walsh C, et al. A randomised prospective trial of intra-operative oesophageal Doppler-guided fluid administration in major gynaecological surgery. Anaesthesia. 2013 Dec;68(12):1224–31.
91 Mikor A, Trásy D, Németh MF, Osztroluczki A, Kocsi S, Kovács I, et al. Continuous central venous oxygen saturation assisted intraoperative hemodynamic management during major abdominal surgery: a randomized, controlled trial. BMC Anesthesiol. 2015 Jun 4;15(1):82.
92 Moppett IK, Rowlands M, Mannings A, Moran CG, Wiles MD, NOTTS Investigators. LiDCO-based fluid management in patients undergoing hip fracture surgery under spinal anaesthesia: a randomized trial and systematic review. Br J Anaesth. 2015 Mar;114(3):444–59.
93 Mythen MG, Webb AR. Perioperative plasma volume expansion reduces the incidence of gut mucosal hypoperfusion during cardiac surgery. Arch Surg. 1995 Apr;130(4):423–9.
94 Noblett SE, Snowden CP, Shenton BK, Horgan AF. Randomized clinical trial assessing the effect of Doppler-optimized fluid management on outcome after elective colorectal resection. Br J Surg. 2006 Sep;93(9):1069–76.
95 Osawa EA, Rhodes A, Landoni G, Galas FRBG, Fukushima JT, Park CHL, et al. Effect of Perioperative Goal-Directed Hemodynamic Resuscitation Therapy on Outcomes Following Cardiac Surgery: A Randomized Clinical Trial and Systematic Review. Crit Care Med. 2016 Apr;44(4):724–33.
96 Parke RL, McGuinness SP, Gilder E, McCarthy LW, Cowdrey K-AL. A randomised feasibility study to assess a novel strategy to rationalise fluid in patients after cardiac surgery. Br J Anaesth. 2015 Jul;115(1):45–52.
97 Bender JS, Smith-Meek MA, Jones CE. Routine pulmonary artery catheterization does not reduce morbidity and mortality of elective vascular surgery: results of a prospective, randomized trial. Ann Surg. 1997 Sep;226(3):229–36; discussion 236-7.
98 Pavlovic G, Diaper J, Ellenberger C, Frei A, Bendjelid K, Bonhomme F, et al. Impact of early haemodynamic goal-directed therapy in patients undergoing emergency surgery: an open prospective, randomised trial. J Clin Monit Comput. 2016 Feb;30(1):87–99.
99 Pearse RM, Harrison DA, MacDonald N, Gillies MA, Blunt M, Ackland G, et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. JAMA. 2014 Jun 4;311(21):2181–90.
100 Pearse R, Dawson D, Fawcett J, Rhodes A, Grounds RM, Bennett ED. Early goal-directed therapy after major surgery reduces complications and duration of hospital stay. A randomised, controlled trial [ISRCTN38797445]. Crit Care. 2005;9(6):R687-93.
101 Peng K, Li J, Cheng H, Ji F. Goal-Directed Fluid Therapy Based on Stroke Volume Variations Improves Fluid Management and Gastrointestinal Perfusion in Patients Undergoing Major Orthopedic Surgery. Med Princ Pract. 2014;23(5):413–20.
102 Pestaña D, Espinosa E, Eden A, Nájera D, Collar L, Aldecoa C, et al. Perioperative Goal-Directed Hemodynamic Optimization Using Noninvasive Cardiac Output Monitoring in Major Abdominal Surgery. Anesth Analg. 2014 Sep;119(3):579–87.
103 Phan TD, D'Souza B, Rattray MJ, Johnston MJ, Cowie BS. A randomised controlled trial of fluid restriction compared to oesophageal Doppler-guided goal-directed fluid therapy in elective major colorectal surgery within an Enhanced Recovery After Surgery program. Anaesth Intensive Care. 2014 Nov;42(6):752–60.

104 Picard J, Bedague D, Bouzat P, Ollinet C, Albaladejo P, Bosson J-L, et al. Oesophageal Doppler to optimize intraoperative haemodynamics during prone position. A randomized controlled trial. Anaesthesia, Crit care pain Med. 2016 Aug;35(4):255–60.

105 Pillai P, McEleavy I, Gaughan M, Snowden C, Nesbitt I, Durkan G, et al. A double-blind randomized controlled clinical trial to assess the effect of doppler optimized intraoperative fluid management on outcome following radical cystectomy. J Urol. 2011 Dec;186(6):2201–6.

106 Pölönen P, Ruokonen E, Hippeläinen M, Pöyhönen M, Takala J. A prospective, randomized study of goal-oriented hemodynamic therapy in cardiac surgical patients. Anesth Analg. 2000 May;90(5):1052–9.

107 Pösö T, Winsö O, Aroch R, Kesek D. Perioperative Fluid Guidance with Transthoracic Echocardiography and Pulse-Contour Device in Morbidly Obese Patients. Obes Surg. 2014 Dec;24(12):2117–25.

108 Benes J, Chytra I, Altmann P, Hluchy M, Kasal E, Svitak R, et al. Intraoperative fluid optimization using stroke volume variation in high risk surgical patients: results of prospective randomized study. Crit Care. 2010;14(3):R118.

109 Ramsingh DS, Sanghvi C, Gamboa J, Cannesson M, Applegate RL. Outcome impact of goal directed fluid therapy during high risk abdominal surgery in low to moderate risk patients: A randomized controlled trial. J Clin Monit Comput. 2013 Jun;27(3):249–57.

110 Reisinger KW, Willigers HM, Jansen J, Buurman WA, Von Meyenfeldt MF, Beets GL, et al. Doppler-guided goal-directed fluid therapy does not affect intestinal cell damage but increases global gastrointestinal perfusion in colorectal surgery: a randomized controlled trial. Color Dis. 2017;19(12):1081–91.

111 Salzwedel C, Puig J, Carstens A, Bein B, Molnar Z, Kiss K, et al. Perioperative goal-directed hemodynamic therapy based on radial arterial pulse pressure variation and continuous cardiac index trending reduces postoperative complications after major abdominal surgery: A multi-center, prospective, randomized study. Crit Care. 2013 Sep 8;17(5):R191.

112 Scheeren TWL, Wiesenack C, Gerlach H, Marx G. Goal-directed intraoperative fluid therapy guided by stroke volume and its variation in high-risk surgical patients: A prospective randomized multicentre study. J Clin Monit Comput. 2013 Jun;27(3):225–33.

113 1Schmid S, Kapfer B, Heim M, Bogdanski R, Anetsberger A, Blobner M, et al. Algorithm-guided goal-directed haemodynamic therapy does not improve renal function after major abdominal surgery compared to good standard clinical care: A prospective randomised trial. Crit Care. 2016 Mar 8;20(1):50.

114 Schultz RJ, Whitfield GF, LaMura JJ, Raciti A, Krishnamurthy S. The role of physiologic monitoring in patients with fractures of the hip. J Trauma. 1985 Apr;25(4):309–16.

115 Senagore AJ, Emery T, Luchtefeld M, Kim D, Dujovny N, Hoedema R. Fluid management for laparoscopic colectomy: A prospective, randomized assessment of goal-directed administration of balanced salt solution or hetastarch coupled with an enhanced recovery program. Dis Colon Rectum. 2009 Dec;52(12):1935–40.

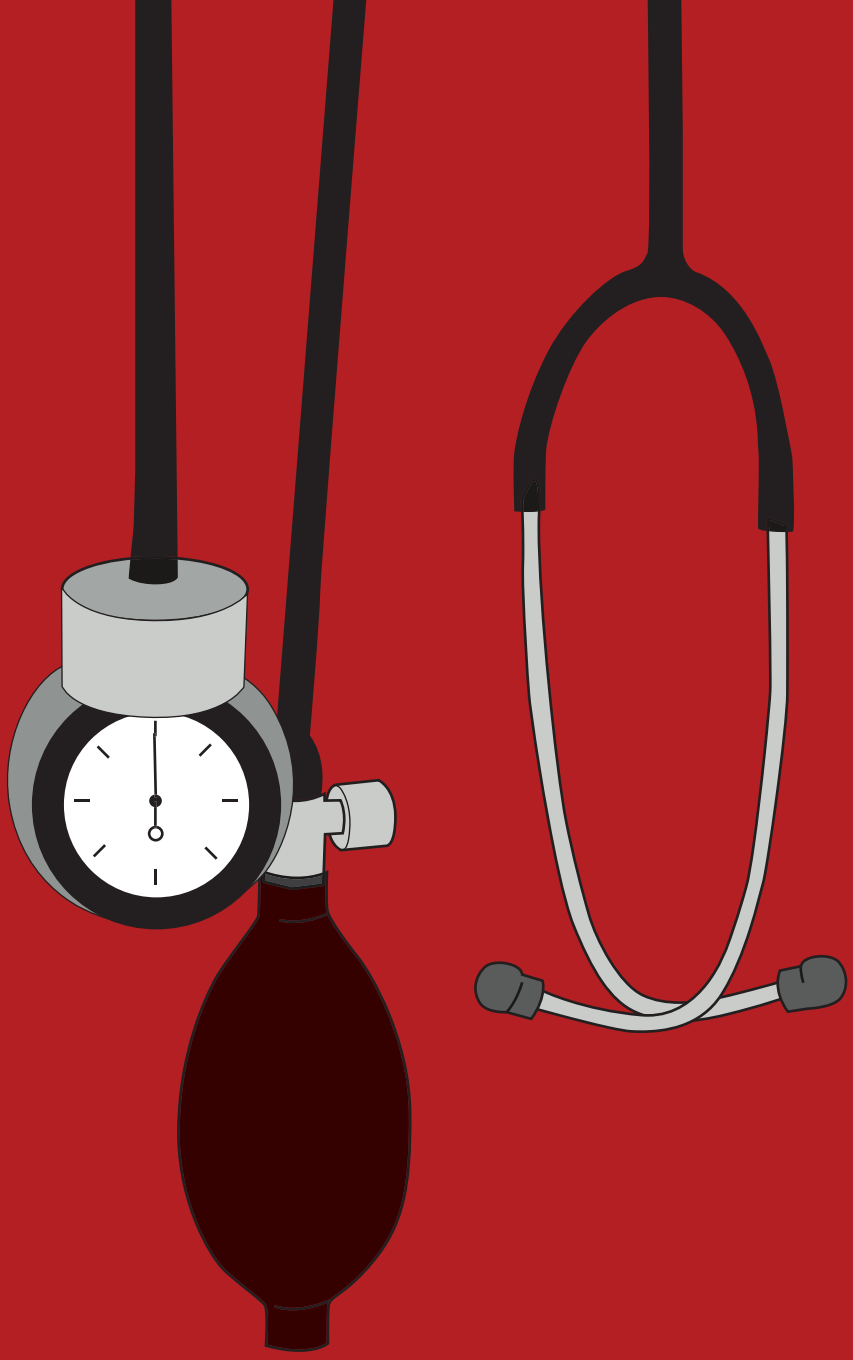
116 Sethi A, Debbarma M, Narang N, Saxena A, Mahobia M, Tomar GS. Impact of Targeted Preoperative Optimization on Clinical Outcome in Emergency Abdominal Surgeries: A Prospective Randomized Trial. Anesth essays Res. 2017;12(1):149–54.

117 Sinclair S, James S, Singer M. Intraoperative intravascular volume optimisation and length of hospital stay after repair of proximal femoral fracture: randomised controlled trial. BMJ. 1997 Oct 11;315(7113):909–12.

118 Smetkin AA, Kirov MY, Kuzkov V V., Lenkin AI, Eremeev A V., Slastilin VY, et al. Single transpulmonary thermodilution and continuous monitoring of central venous oxygen saturation during off-pump coronary surgery. Acta Anaesthesiol Scand. 2009 Apr;53(4):505–14.

119 Benes J, Haidingerova L, Pouska J, Stepanik J, Stenglova A, Zatloukal J, et al. Fluid management guided by a continuous non-invasive arterial pressure device is associated with decreased postoperative morbidity after total knee and hip replacement. BMC Anesthesiol. 2015 Oct 15;15(1):148.

120 Stens J, Hering JP, van der Hoeven CWP, Boom A, Traast HS, Garmers LE, et al. The added value of cardiac index and pulse pressure variation monitoring to mean arterial pressure-guided volume therapy in moderate-risk abdominal surgery (COGUIDE): a pragmatic multicentre randomised controlled trial. Anaesthesia. 2017 Sep;72(9):1078–87.
121 Szakmany T, Toth I, Kovacs Z, Leiner T, Mikor A, Koszegi T, et al. Effects of volumetric vs. pressure-guided fluid therapy on postoperative inflammatory response: A prospective, randomized clinical trial. Intensive Care Med. 2005 May;31(5):656–63.
122 Thomson R, Meeran H, Valencia O, Al-Subaie N. Goal-directed therapy after cardiac surgery and the incidence of acute kidney injury. J Crit Care. 2014 Dec;29(6):997–1000.
123 Ueno S, Tanabe G, Yamada H, Kusano C, Yoshidome S, Nuruki K, et al. Response of patients with cirrhosis who have undergone partial hepatectomy to treatment aimed at achieving supranormal oxygen delivery and consumption. Surgery. 1998 Mar;123(3):278–86.
124 Valentine RJ, Duke ML, Inman MH, Grayburn PA, Hagino RT, Kakish HB, et al. Effectiveness of pulmonary artery catheters in aortic surgery: A randomized trial. J Vasc Surg. 1998 Feb;27(2):203–12.
125 Van Der Linden PJ, Dierick A, Wilmin S, Bellens B, De Hert SG. A randomized controlled trial comparing an intraoperative goal-directed strategy with routine clinical practice in patients undergoing peripheral arterial surgery. Eur J Anaesthesiol. 2010 Sep;27(9):788–93.
126 Velmahos GC, Demetriades D, Shoemaker WC, Chan LS, Tatevossian R, Wo CC, et al. Endpoints of resuscitation of critically injured patients: normal or supranormal? A prospective randomized trial. Ann Surg. 2000 Sep;232(3):409–18.
127 Venn R, Steele A, Richardson P, Poloniecki J, Grounds M, Newman P. Randomized controlled trial to investigate influence of the fluid challenge on duration of hospital stay and perioperative morbidity in patients with hip fractures. Br J Anaesth. 2002 Jan;88(1):65–71.
128 Wakeling HG, McFall MR, Jenkins CS, Woods WGA, Miles WFA, Barclay GR, et al. Intraoperative oesophageal Doppler guided fluid management shortens postoperative hospital stay after major bowel surgery. Br J Anaesth. 2005 Nov;95(5):634–42.
129 WenKui Y, Ning L, JianFeng G, WeiQin L, ShaoQiu T, Zhihui T, et al. Restricted peri-operative fluid administration adjusted by serum lactate level improved outcome after major elective surgery for gastrointestinal malignancy. Surgery. 2010 Apr;147(4):542–52.
130 Bennett VA, Cecconi M. Perioperative fluid management: From physiology to improving clinical outcomes. Indian J Anaesth. 2017 Aug;61(8):614–21.
131 Meng L, Heerdt PM. Perioperative goal-directed haemodynamic therapy based on flow parameters: a concept in evolution. Br J Anaesth. 2016 Dec;117(suppl 3):iii3–17.
132 Saugel B, Michard F, Scheeren TWL. Goal-directed therapy: hit early and personalize! J Clin Monit Comput. 2018 Jun 26;32(3):375–7.
133 Uhlig C, Spieth PM. Perioperative hemodynamic therapy: Goal-directed or meta-directed? Minerva Anestesiol. 2016 Nov;82(11):1135–7.
134 Giglio M, Manca F, Dalfino L, Brienza N. Perioperative hemodynamic goal-directed therapy and mortality: A systematic review and meta-analysis with meta-regression. Minerva Anestesiol. 2016 Nov;82(11):1199–213.
135 Cronhjort M, Wall O, Nyberg E, Zeng R, Svensen C, Mårtensson J, et al. Impact of hemodynamic goal-directed resuscitation on mortality in adult critically ill patients: a systematic review and meta-analysis. J Clin Monit Comput. 2018 Jun 8;0(3):403–14.
136 Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: Version 1.0. Trials. 2017;18(S3):1–50.



3

Perioperative goal-directed therapy – What is the evidence?

Thomas Kaufmann
Bernd Saugel
Thomas W.L. Scheeren

Best Practice & Research: Clinical Anaesthesiology

Abstract

Perioperative goal-directed therapy aims at optimizing global hemodynamics during the perioperative period by titrating fluids, vasopressors, and/or inotropes to predefined hemodynamic goals. There is evidence on the benefit of perioperative goal-directed therapy, but its adoption into clinical practice is slow and incomprehensive. Current evidence indicates that treating patients according to perioperative goal-directed therapy protocols reduces morbidity and mortality, particularly in patients having high-risk surgery. Perioperative goal-directed therapy protocols need to be started early, should include vasoactive agents in addition to fluids, and should target blood flow related variables. Future promising developments in the field of perioperative goal-directed therapy include personalized hemodynamic management and closed-loop system management.

Introduction

Optimal hemodynamic management in patients undergoing surgery remains a challenge for anesthesiologists, and there is large variability in care between health care providers, even for the same type of surgery and within the same hospital (1,2). Insufficient hemodynamic management may lead to tissue hypoperfusion or tissue edema formation, both resulting in inadequate oxygen delivery to the tissue associated with organ dysfunction and adverse postoperative patient outcome (3). To reduce the uncertainties regarding the optimal hemodynamic management, various protocols for perioperative goal-directed therapy (PGDT) have been proposed (4). PGDT aims at optimizing global hemodynamics during the perioperative period by titrating fluids, vasopressors, and/or inotropes to predefined hemodynamic goals (5).

Since the first concepts of PGDT were described in the 1980s, more than 100 randomized controlled trials (RCTs) have been published, followed by an ever-increasing number of systematic reviews with meta-analyses. The cumulative evidence that PGDT is capable of improving patient outcome has led to its adoption in clinical practice guidelines, e.g., in the United Kingdom (6), France (7), and by the European Society of Anaesthesiology (8). Despite this, the actual implementation and adoption of PGDT protocols in clinical practice has been rather slow and incomprehensive (9). One reason for the poor adoption rate might be that PGDT is a vague, poorly defined, and confusing term for physicians and researchers, as it is used in multiple different clinical scenarios. In general, PGDT can involve optimization of cardiac preload, afterload, and contractility to achieve a balance between systematic oxygen delivery and oxygen demand.

In this review, we summarize the currently available evidence for PGDT. We discuss early trials employing a protocol guided by hemodynamic variables assessed with the pulmonary artery catheter (PAC) as well as trials using modern PGDT protocols with less invasive monitoring. In addition, we also critically discuss why studies on PGDT may have led to conflicting results. Finally, we consider future directions of PGDT, which may help to further evolve this research field

Different concepts of perioperative goal-directed therapy

Early concepts of perioperative goal-directed therapy

The invention of the flow-directed balloon-tipped PAC in the 1970s has led to an increase in the use of invasive hemodynamic monitoring in critically ill patients (10). However, it took some time for the first prospective RCT to be performed, which analyzed improvement of clinical outcome when using interventions guided by hemodynamic variables assessed with the PAC.

One of the first RCTs evaluated elderly patients undergoing hip fracture surgery, with the intervention group undergoing perioperative optimization of their hemodynamic status guided by PAC-derived variables (11). In these patients, certain hemodynamic variables were monitored and optimized both pre- and postoperatively (among others, a cardiac index (CI) between 3.0

and 3.5 L/min/m^2 and oxygen consumption (VO_2) between 110 mL/min/m^2 and 165 mL/min/m^2) (11). This was one of the first treatment protocols that can be called PGDT. Patients treated according to this PAC-based protocol showed markedly lower postoperative mortality rates than control patients (1 of 35 (2.9%) versus 10 of 35 (29%)). Another RCT studied patients having surgery for peripheral vascular disease, who are at high risk of cardiovascular complications (12). A preoperative optimization protocol was used targeting a pulmonary artery wedge pressure between 8 and 15 mmHg, a CI of more than 2.8 L/min/m^2, and a systemic vascular resistance of less than 1100 dyne-sec/cm^5 (12). Fluids, inotropes, and vasodilators could be used to achieve these goals, and in intervention group patients, the postoperative mortality was 1 of 68 (1.5%) compared to 1 of 21 (9.5%) in control group patients (12).

Supranormal oxygen delivery

Measurements of hemodynamic variables in survivors and nonsurvivors of shock after major trauma surgery suggested that higher – so-called supranormal – values of hemodynamic variables were associated with improved survival in these patients (13). It was hypothesized that the survival benefit might be the consequence of less shock-related complications and organ failure in patients having supranormal values, e.g., CI of more than 4.5 L/min/m^2, oxygen delivery (DO_2) of more than 600 mL/min/m^2, and VO_2 of more than 170 mL/min/m^2. These values, which are higher than normal resting hemodynamic values, were then used in prospective trials as target values for high-risk patients having major noncardiac surgery. The results of the first trial in high-risk patients showed a reduction in mortality, postoperative complications, and length of hospital stay after using a treatment protocol based on the supranormal target values (14). These results led to the hypothesis that high-risk patients having surgery should be hemodynamically optimized to these targets already before surgery, i.e., before the surgical trauma and the development of organ failure. In addition, it was suggested to maintain these target values throughout surgery and also for the first 24 h after surgery in the ICU. Several trials used similar protocols with varying results (11,15-21). These early trials were summarized in a systematic review with meta-analysis by the original author of the first trial, who concluded that preoperative initiation of the protocol was associated with a decrease in mortality (22). It was emphasized that PGDT protocols have to be initiated early and aggressively to be most effective (22).

Investigators in these initial studies on PGDT used the PAC to monitor patient hemodynamics and guide interventions (10). Consequently, the use of the PAC became widespread and the PAC-derived hemodynamic data were used by physicians to guide hemodynamic therapy in the ICU and the operating room. At its peak, more than 40% of all critically ill patients received a PAC as part of their care in the ICU, but evidence regarding its benefit was never objectified (23). The invasiveness of the PAC also meant that placement was associated with complications such as catheter-related infections and thromboembolic events. In one of the largest trials comparing a PGDT protocol based on PAC-derived hemodynamic variables with standard care without the use of a PAC, no survival benefit was found for PAC-guided therapy in high-risk surgical patients (24). The results of this RCT and numerous other trials combined in a Cochrane review led to a decrease in the use of the PAC (25,26). Also, routine PAC placement is not recommended by the

European Society of Cardiology/European Society of Anaesthesiology in their current guidelines on the management of patients undergoing non-cardiac surgery (27). In addition, optimizing patients preoperatively with a PAC requires admission to an ICU, which consumes many resources and is therefore not applicable to patients undergoing most types of surgery. The decline of the use of the PAC was further facilitated by the emergence of minimally invasive and noninvasive monitoring techniques to estimate cardiac output (CO) and other hemodynamic variables.

Minimally invasive and noninvasive monitoring in PGDT protocols

Minimally invasive monitoring refers to devices that measure hemodynamic variables by only using an arterial catheter or esophageal Doppler (28). Several of these devices have been developed, and the most important ones are discussed here.

Pulse wave analysis

Pulse wave analysis (PWA) allows estimation of stroke volume (SV) or CO and of dynamic preload variables. The method is based on the principle that aortic pulse pressure is proportional and inversely proportional to aortic compliance. Although static variables are single snapshots taken at specific points in the cardiac cycle (e.g., CO measured by thermodilution or central venous pressure (CVP)), dynamic preload variables express rapid changes in the cardiovascular status and can be monitored continuously. In addition, an increase in CO induced by volume expansion can be predicted by dynamic preload variables before volume expansion is actually performed, which is helpful in PGDT algorithms and helps to avoid unnecessary fluid administration. Examples of these dynamic preload variables include systolic pressure variation (SPV), pulse pressure variation (PPV), and stroke volume variation (SVV). These variables are induced by heart-lung interactions during a respiratory cycle in mechanically ventilated patients and are an indicator of the position on the Frank-Starling curve, which is proportional to the degree of preload dependency (29,30). Of these dynamic variables, PPV is considered to have the best predictive ability for fluid responsiveness (31). A disadvantage of these dynamic preload variables is that they cannot be used in a number of concomitant conditions including cardiac arrhythmias and spontaneous breathing (32). Some PWA monitors can be calibrated with an independent measurement of CO done by transpulmonary thermodilution. This calibration is done similar to the PAC with injection of a small fluid bolus. To do this, transpulmonary thermodilution monitors require both a central venous catheter and a femoral arterial catheter and are therefore considered invasive, not minimally invasive (28). Transpulmonary thermodilution monitors can also estimate extravascular lung water, which is a measure of pulmonary edema, and pulmonary vascular permeability, which is a measure of pulmonary capillary leakage. These variables can help to guide fluid strategies, for example, as safety measures to avoid fluid overload in patients with acute respiratory distress syndrome. Transpulmonary thermodilution monitors are therefore mainly used for complex patients in the ICU (33). Uncalibrated PWA monitors estimate the CO only from arterial pressure waveform characteristics and biometric data. A proprietary algorithm uses the mean, standard deviation, skewness, and kurtosis of arterial pressure and arterial compliance

estimated from sex, age, weight, and height (34). The waveform characteristics become less reliable in pathophysiological conditions with low vascular resistance such as liver disease, during liver surgery, or septic shock. In these conditions, use of uncalibrated PWA monitors is not recommended (35).

Oesophageal Doppler Monitor

The Oesophageal Doppler Monitor (ODM) probe is placed in the patient's esophagus and uses Doppler ultrasound to measure the velocity of blood flow in the adjacent descending aorta. The blood flow in the descending aorta is correlated to CO, assuming a fixed proportion of blood flow going to the upper and lower part of the body. Estimation of CO using the ODM was originally thought to have agreement with invasive CO measurements using the PAC (36). However, results of a more recent systematic review with meta-analysis have shown that agreement between ODM and PAC derived CO measurements is moderate at best (37). Several PGDT protocols using ODM-derived hemodynamic variables have been developed and tested. The first trials evaluated a PGDT protocol using an ODM to titrate fluids in patients undergoing hip fracture surgery (38,39). Both studies concluded that intraoperative volume loading to optimize stroke volume using an ODM resulted in more rapid postoperative recovery and a reduced length of hospital stay. Use of ODM was not limited to hip fracture surgery patients but ODM was also used to titrate fluids and vasoactive medication in patients undergoing abdominal surgery (40). This was also associated with a reduction in length of hospital stay, but also in postoperative complications, number of patients requiring ICU admission, and time to return of bowel function (40). Eventually, use of PGDT protocols employing an ODM was implemented in national guidelines in the United Kingdom to enhance recovery after surgery (6). In general, a limitation of the ODM is that it can only be placed intraoperatively under general anesthesia due to patient convenience.

Perioperative goal-directed therapy – what is the evidence?

Summarizing the evidence on PGDT is challenging because many different protocols have been developed and studied over the years. Nevertheless, a multitude of meta-analyses has been conducted on this subject. Several meta-analyses have suggested a beneficial effect of PGDT in terms of reduction in postoperative complications and length of stay (41-43). However, the studies included in these meta-analyses were heterogeneous and varied with regard to not only types of surgery but also all individual components of the PGDT intervention, including the timing of the intervention; the type of monitoring device; the hemodynamic variables assessed; the hemodynamic values targeted; and the types and amounts of fluids, vasopressors, and/or inotropes used (44). This clinical heterogeneity needs to be considered when pooling results of individual RCTs (45).

We previously performed a systematic review on PGDT where the observed clinical heterogeneity made us conclude that it was inappropriate to pool all the data to estimate an intervention effect of the PGDT intervention (44). Therefore, we believe a definite conclusion on the effect of PGDT remains to be elucidated (44).

Recently, a systematic review with meta-analysis was published with more rigorous methodological conditions (4). Benefit for PGDT was analyzed in various predefined subgroups, such as timing of initiation of PGDT protocol or whether a CO monitoring device was used. Subgroup analysis can be a way of reducing clinical heterogeneity within a meta-analysis, particularly if different patient populations, technologies, and interventions exist within the literature. This approach also takes into account that many aspects of the management of high-risk surgical patients and trial methodology may have changed over the last decades (46). The authors of this systematic review discuss a number of results. First, PGDT reduced the risk of mortality compared with standard care only in high-risk patients (OR 0.60 (95% CI 0.42-0.85)) but not in low-risk patients (OR 0.79 (95% CI 0.50-1.24)) (4). High risk was defined as patients undergoing cardiac surgery, critically ill patients, or studies with >50% ASA III physical class patients. Second, PGDT reduced the risk of mortality compared with standard care only if PGDT was started intraoperatively (OR 0.65 (95% CI 0.47-0.89)) and not when PGDT was started postoperatively (OR 0.72 (95% CI 0.38-1.33)) (4). Third, PGDT reduced mortality only if vasoactive agents were used per protocol in addition to fluids (OR 0.59 (95% CI 0.40-0.89)) and not if only a fluid-based intervention was performed (OR 0.65 (95% CI 0.41-1.05)) (4). Last, PGDT reduced mortality only if a CO monitor was used to guide therapy (OR 0.68 (95% CI 0.49-0.95)) (4).

Based on the abovementioned evidence, we believe that PGDT should be employed in patients who will benefit the most: high-risk patients undergoing high-risk surgery. The PGDT protocol should be started early during the perioperative period (47). In addition, an algorithm which mainly targets flow optimization (i.e., CO or SV) should be used, and dynamic preload variables might also be included to assess fluid responsiveness.

Two large RCTs on PGDT are currently being conducted. The first is the FLuid Optimisation in Emergency LAparotomy (FLO-ELA) trial (www.floela.org), which aims to include 7646 patients undergoing emergency bowel surgery using a PGDT protocol. The other trial is the OPtimisation of Peri-operaTive Cardlovascular Management to Improve Surgical outcomE II (OPTIMISE II) trial (www.optimiseii.org), which aims to include 2502 patients undergoing elective major abdominal surgery. Currently, a third large multicenter trial is being analyzed, and results should soon be available (48).

Perioperative goal-directed therapy – future directions

Personalized hemodynamic management

The above-mentioned methods and PGDT protocols all use a general strategy of hemodynamic optimization employing predefined "normal" values as hemodynamic targets. However, it is now known that many hemodynamic variables have marked inter-individual variability and depend on biometric factors (49). For example, CO measured by transpulmonary thermodilution in critically ill patients was shown to be independently associated with age, height, and body weight (50). In addition, left ventricular volume and stroke volume varies by gender and decreases with age

(51,52). Consequently, the term personalized hemodynamic management has been suggested as a means to optimize cardiovascular dynamics based on the patient's personal hemodynamic profile (49). For PGDT, this could be realized by implementing personalized concepts of hemodynamic management based on individual baseline values and functional assessment of fluid responsiveness in the operating room.

Closed-loop system management

Other developments such as closed-loop system management are suggested to further help implementing PGDT protocols in clinical practice. One of the largest trials on PGDT showed that there is a learning curve for clinicians and that compliance to PGDT protocols is suboptimal (53). A possible way of integrating the monitoring of the variables in one place and increasing the compliance to the protocols is by implementing a closed-loop system for hemodynamic management. A closed-loop system is a system where a controller monitors multiple variables and adjusts interventions using a feedback process (54). In the context of PGDT, a clinical example of a closed-loop system uses dynamic variables (e.g., PPV and SVV) collected from an uncalibrated PCA monitor to automatically titrate fluid application. This system has been tested in simulation (55,56), engineering studies (57), and animal studies (58). Clinical studies in patients undergoing moderate- and high-risk surgery have established feasibility of implementing these systems and show that patients consequently spent a great portion of time in a preload-independent state throughout surgery (59,60). At the moment, the feasibility of a closed-loop system for vasopressor infusion is being developed, with initial promising results in simulation and animal studies (61,62).

Summary

Perioperative goal-directed therapy aims at optimizing global hemodynamics during the perioperative period by titrating fluids, vasopressors, and/or inotropes to predefined hemodynamic goals. There is evidence on the benefit of perioperative goal-directed therapy, but its adoption into clinical practice is slow and incomprehensive. Current evidence indicates that treating patients according to perioperative goal-directed therapy protocols reduces morbidity and mortality, particularly in patients having high-risk surgery. Perioperative goal-directed therapy protocols need to be started early, should include vasoactive agents in addition to fluids, and should target blood flow related variables. Future promising developments in the field of perioperative goal-directed therapy include personalized hemodynamic management and closed-loop system management.

Practice points

- Current evidence indicates that treating patients according to PGDT protocols reduces morbidity and mortality, particularly in patients undergoing high-risk surgery. The PGDT protocol needs to be started early (i.e., intraoperatively, not postoperatively), should include vasoactive agents in addition to fluids, and should target blood flow related variables.

- Clinical heterogeneity between different studies makes drawing definite conclusions on the benefit of PGDT on outcome difficult so that multicenter trials on this topic are still needed.

Research agenda

- Large multicenter trials (OPTIMISE-II and FLO-ELA) are ongoing to further elucidate the benefit of PGDT protocols.

- Future PGDT protocols will consider individual variability in optimal hemodynamic targets and move toward personalized hemodynamic management.

- Further research on closed-loop systems to administer fluids and vasoactive agents will help with improving protocol compliance.

References

1 Minto G, Mythen MG. Perioperative fluid management: science, art or random chaos? Br J Anaesth. 2015;114(5):717–21.

2 Ehrenfeld JM, Cannesson M, Lilot M, Harrington B, Rinehart J, Lee C. Variability in practice and factors predictive of total crystalloid administration during abdominal surgery: retrospective two-centre analysis †. Br J Anaesth. 2015;114(5):767–76.

3 Shin CH, Long DR, McLean D, Grabitz SD, Ladha K, Timm FP, et al. Effects of Intraoperative Fluid Management on Postoperative Outcomes: A Hospital Registry Study. Ann Surg. 2018 Jun 10;267(6):1084–92.

4 Chong MA, Wang Y, Berbenetz NM, McConachie I. Does goal-directed haemodynamic and fluid therapy improve peri-operative outcomes? Eur J Anaesthesiol. 2018;35(7):469–83.

5 Cove ME, Pinsky MR. Perioperative hemodynamic monitoring. Best Pract Res Clin Anaesthesiol. 2012;26(4):453–62.

6 National Institute for Health and Clinical Excellence. Medical technologies guidance MTG3: CardioQ-ODM oesophageal doppler monitor. 2011.

7 Vallet B, Blanloeil Y, Cholley B, Orliaguet G, Pierre S, Tavernier B. Guidelines for perioperative haemodynamic optimization. Ann Fr Anesth Reanim. 2013 Oct;32(10):e151–8.

8 European Society of Anaesthesiology. Perioperative Goal-Directed Fluid Therapy Protocol Summary. Available from: http://html.esahq.org/patientsafetykit/resources/downloads/05_Checklists/Various_Checklists/Perioperative_Goal_Directed_Therapy_Protocols.pdf

9 Cannesson M, Pestel G, Ricks C, Hoeft A, Perel A. Hemodynamic monitoring and management in patients undergoing high risk surgery: A survey among North American and European anesthesiologists. Crit Care. 2011;15(4):R197.

10 Swan HJ, Ganz W, Forrester J, Marcus H, Diamond G, Chonette D. Catheterization of the heart in man with use of a flow-directed balloon-tipped catheter. N Engl J Med. 1970 Aug 27;283(9):447–51.

11 Schultz RJ, Whitfield GF, LaMura JJ, Raciti A, Krishnamurthy S. The role of physiologic monitoring in patients with fractures of the hip. J Trauma. 1985 Apr;25(4):309–16.

12 Berlauk JF, Abrams JH, Gilmour IJ, O'Connor SR, Knighton DR, Cerra FB. Preoperative optimization of cardiovascular hemodynamics improves outcome in peripheral vascular surgery. A prospective, randomized clinical trial. Ann Surg. 1991 Sep;214(3):289–97; discussion 298-9.

13 Shoemaker WC, Montgomery ES, Kaplan E, Elwyn DH. Physiologic Patterns in Surviving and Nonsurviving Shock Patients: Use of Sequential Cardiorespiratory Variables in Defining Criteria for Therapeutic Goals and Early Warning of Death. Arch Surg. 1973 May;106(5):630–6.

14 Shoemaker WC, Appel PL, Kram HB, Waxman K, Lee TS. Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients. Chest. 1988 Dec;94(6):1176–86.

15 Fleming A, Bishop M, Shoemaker W, Appel P, Sufficool W, Kuvhenguwha A, et al. Prospective trial of supranormal values as goals of resuscitation in severe trauma. Arch Surg. 1992 Oct;127(10):1175–9; discussion 1179-81.

16 Boyd O, Grounds RM, Bennett ED. A randomized clinical trial of the effect of deliberate perioperative increase of oxygen delivery on mortality in high-risk surgical patients. JAMA. 1993 Dec 8;270(22):2699–707.

17 Bishop MH, Shoemaker WC, Appel PL, Meade P, Ordog GJ, Wasserberger J, et al. Prospective, Randomized Trial of Survivor Values of Cardiac Index, Oxygen Delivery, and Oxygen Consumption as Resuscitation Endpoints in Severe Trauma. J Trauma Inj Infect Crit Care. 1995 May;38(5):780–7.
18 Ueno S, Tanabe G, Yamada H, Kusano C, Yoshidome S, Nuruki K, et al. Response of patients with cirrhosis who have undergone partial hepatectomy to treatment aimed at achieving supranormal oxygen delivery and consumption. Surgery. 1998 Mar;123(3):278–86.
19 Wilson J, Woods I, Fawcett J, Whall R, Dibb W, Morris C, et al. Reducing the risk of major elective surgery: randomised controlled trial of preoperative optimisation of oxygen delivery. BMJ. 1999 Apr 24;318(7191):1099–103.
20 Velmahos GC, Demetriades D, Shoemaker WC, Chan LS, Tatevossian R, Wo CC, et al. Endpoints of resuscitation of critically injured patients: normal or supranormal? A prospective randomized trial. Ann Surg. 2000 Sep;232(3):409–18.
21 Lobo SMA, Salgado PF, Castillo VGT, Borim AA, Polachini CA, Palchetti JC, et al. Effects of maximizing oxygen delivery on morbidity and mortality in high-risk surgical patients. Crit Care Med. 2000 Oct;28(10):3396–404.
22 Kern JW, Shoemaker WC. Meta-analysis of hemodynamic optimization in high-risk patients. Crit Care Med. 2002;30(8):1686–92.
23 Connors AF, Speroff T;, Dawson N V, Thomas C. The effectiveness of right heart catheterization in the initial care of criti. JAMA Res Libr pg. 1996;276(11).
24 Sandham JD, Hull RD, Brant RF, Knox L, Pineo GF, Doig CJ, et al. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. N Engl J Med. 2003 Jan 2;348(1):5–14.
25 Wiener RS, Welch HG. Trends in the Use of the Pulmonary Artery Catheter in the United States, 1993-2004. JAMA. 2007 Jul 25;298(4):1993–2004.
26 Rajaram SS, Desai NK, Kalra A, Gajera M, Cavanaugh SK, Brampton W, et al. Pulmonary artery catheters for adult patients in intensive care. Cochrane database Syst Rev. 2013 Feb 28;373(2):CD003408.
27 Kristensen SD, Knuuti J, Saraste A, Anker S, Bøtker HE, De Hert S, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: Cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: Cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesth. Eur Heart J. 2014;35(35):2383–431.
28 Saugel B, Vincent J-L. Cardiac output monitoring: how to choose the optimal method for the individual patient. Curr Opin Crit Care. 2018;24(3):165–72.
29 Reuter DA, Felbinger TW, Schmidt C, Kilger E, Goedje O, Lamm P, et al. Stroke volume variations for assessment of cardiac responsiveness to volume loading in mechanically ventilated patients after cardiac surgery. Intensive Care Med. 2002;28(4):392–8.
30 Michard F. Changes in arterial pressure during mechanical ventilation. Anesthesiology. 2005 Aug;103(2):419–28; quiz 449–5.
31 Marik PE, Cavallazzi R, Vasu T, Hirani A. Dynamic changes in arterial waveform derived variables and fluid responsiveness in mechanically ventilated patients: A systematic review of the literature. Crit Care Med. 2009;37(9):2642–7.
32 Monnet X, Marik PE, Teboul JL. Prediction of fluid responsiveness: an update. Ann Intensive Care 2016;6:1-11.
33 Monnet X, Teboul JL. Transpulmonary thermodilution: Advantages and limits. Crit Care. 2017;21(1):1–12.
34 Langewouters GJ, Wesseling KH, Goedhard WJ. The pressure dependent dynamic elasticity of 35 thoracic and 16 abdominal human aortas in vitro described by a five component model. J Biomech. 1985;18(8):613–20.

35 Slagt C, Malagon I, Groeneveld ABJ. Systematic review of uncalibrated arterial pressure waveform analysis to determine cardiac output and stroke volume variation. Br J Anaesth. 2014 Apr;112(4):626–37.

36 Singer M, Bennett ED. Noninvasive optimization of left ventricular filling using esophageal Doppler. Crit Care Med. 1991 Sep;19(9):1132–7.

37 Peyton PJ, Chong SW. Minimally Invasive Measurement of Cardiac Output during Surgery and Critical Care. Anesthesiology. 2010 Nov;113(5):1220–35.

38 Sinclair S, James S, Singer M. Intraoperative intravascular volume optimisation and length of hospital stay after repair of proximal femoral fracture: randomised controlled trial. BMJ. 1997 Oct 11;315(7113):909–12.

39 Venn R, Steele A, Richardson P, Poloniecki J, Grounds M, Newman P. Randomized controlled trial to investigate influence of the fluid challenge on duration of hospital stay and perioperative morbidity in patients with hip fractures. Br J Anaesth. 2002 Jan;88(1):65–71.

40 Abbas SM, Hill AG. Systematic review of the literature for the use of oesophageal Doppler monitor for fluid replacement in major abdominal surgery. Anaesthesia. 2008 Jan;63(1):44–51.

41 Benes J, Giglio M, Brienza N, Michard F. The effects of goal-directed fluid therapy based on dynamic parameters on post-surgical outcome: A meta-analysis of randomized controlled trials. Crit Care. 2014;18(5):1–11.

42 Cecconi M, Corredor C, Arulkumaran N, Abuella G, Ball J, Grounds RM, et al. Clinical review: Goal-directed therapy-what is the evidence in surgical patients? The effect on different risk groups. Crit Care. 2012 Mar 5;17(2):209.

43 Michard F, Giglio MT, Brienza N. Perioperative goal-directed therapy with uncalibrated pulse contour methods: Impact on fluid management and postoperative outcome. Br J Anaesth. 2017 Jun 11;119(1):22–30.

44 Kaufmann T, Clement RP, Scheeren TWL, Saugel B, Keus F, van der Horst ICC. Perioperative goal-directed therapy: A systematic review without meta-analysis. Acta Anaesthesiol Scand. 2018 Nov 6;62(10):1340–55.

45 Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011.

46 Gillies MA, Pearse R, Chew MS. Peri-operative goal-directed therapy. Eur J Anaesthesiol. 2018;35(7):467–8.

47 Saugel B, Michard F, Scheeren TWL. Goal-directed therapy: hit early and personalize! J Clin Monit Comput. 2018 Jun 26;32(3):375–7.

48 Montenij L, de Waal E, Frank M, van Beest P, de Wit A, Kruitwagen C, et al. Influence of early goal-directed therapy using arterial waveform analysis on major complications after high-risk abdominal surgery: study protocol for a multicenter randomized controlled superiority trial. Trials. 2014 Sep 16;15:360.

49 Saugel B, Vincent JL, Wagner JY. Personalized hemodynamic management. Curr Opin Crit Care. 2017;23(4):334–41.

50 Saugel B, Mair S, Götz SQ, Tschirdewahn J, Frank J, Höllthaler J, et al. Indexation of cardiac output to biometric parameters in critically ill patients: A systematic analysis of a transpulmonary thermodilution-derived database. J Crit Care. 2015 Oct;30(5):957–62.

51 Cain PA, Ahl R, Hedstrom E, Ugander M, Allansdotter-Johnsson A, Friberg P, et al. Age and gender specific normal values of left ventricular mass, volume and function for gradient echo magnetic resonance imaging: A cross sectional study. BMC Med Imaging. 2009;9(Lvm).

52 Kou S, Caballero L, Dulgheru R, Voilliot D, De Sousa C, Kacharava G, et al. Echocardiographic reference ranges for normal cardiac chamber size: Results from the NORRE study. Eur Heart J Cardiovasc Imaging. 2014;15(6):680–90.

53 Pearse RM, Harrison DA, MacDonald N, Gillies MA, Blunt M, Ackland G, et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. JAMA. 2014 Jun 4;311(21):2181–90.

54 Rinehart J, Liu N, Alexander B, Cannesson M. Closed-loop systems in anesthesia: Is there a potential for closed-loop fluid management and hemodynamic optimization? Anesth Analg. 2012 Jan;114(1):130–43.

55 Rinehart J, Alexander B, Le Manach Y, Hofer C, Tavernier B, Kain ZN, et al. Evaluation of a novel closed-loop fluid-administration system based on dynamic predictors of fluid responsiveness: an in silico simulation study. Crit Care. 2011;15(6):R278.

56 Rinehart J, Chung E, Canales C, Cannesson M. Intraoperative stroke volume optimization using stroke volume, arterial pressure, and heart rate: closed-loop (learning intravenous resuscitator) versus anesthesiologists. J Cardiothorac Vasc Anesth. 2012 Oct;26(5):933–9.

57 Rinehart J, Lee C, Cannesson M, Dumont G. Closed-Loop Fluid Resuscitation. Anesth Analg. 2013 Nov;117(5):1110–8.

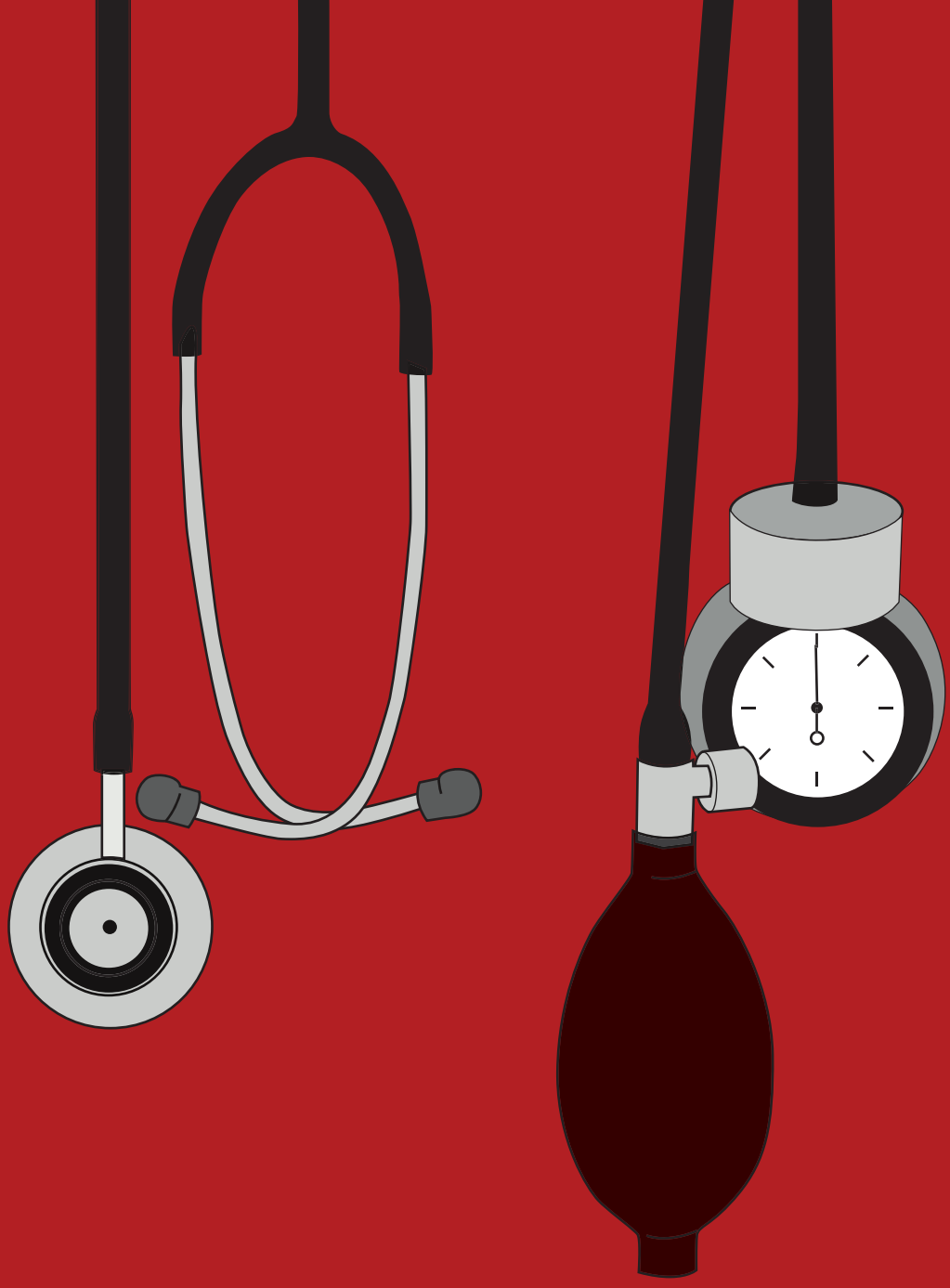
58 Rinehart J, Lee C, Canales C, Kong A, Kain Z, Cannesson M. Closed-Loop Fluid Administration Compared to Anesthesiologist Management for Hemodynamic Optimization and Resuscitation During Surgery. Anesth Analg. 2013 Nov;117(5):1119–29.

59 Joosten A, Huynh T, Suehiro K, Canales C, Cannesson M, Rinehart J. Goal-Directed fluid therapy with closed-loop assistance during moderate risk surgery using noninvasive cardiac output monitoring: A pilot study. Br J Anaesth. 2015 Jun;114(6):886–92.

60 Rinehart J, Lilot M, Lee C, Joosten A, Huynh T, Canales C, et al. Closed-loop assisted versus manual goal-directed fluid therapy during high-risk abdominal surgery: A case-control study with propensity matching. Crit Care. 2015 Dec 19;19(1):1–11.

61 Rinehart J, Joosten A, Ma M, Calderon M-D, Cannesson M. Closed-loop vasopressor control: in-silico study of robustness against pharmacodynamic variability. J Clin Monit Comput. 2018 Dec 11;

62 Joosten A, Delaporte A, Alexander B, Su F, Creteur J, Vincent J-L, et al. Automated Titration of Vasopressor Infusion Using a Closed-loop Controller: In Vivo Feasibility Study Using a Swine Model. Anesthesiology. 2019 Mar;130(3):394–403.



4

A Bayesian network analysis of the diagnostic process and its accuracy to determine how clinicians estimate cardiac function in critically ill patients: Prospective observational cohort study

Thomas Kaufmann*
José Castela Forte*
Bart Hiemstra
Marco A. Wiering
Marco Grzegorczyk
Anne H. Epema
Iwan C.C. van der Horst
SICS Study Group

** Both authors contributed equally to this work*

Journal of Medical Internet Research: Medical Informatics

Abstract

Background
Hemodynamic assessment of critically ill patients is a challenging endeavor, and advanced monitoring techniques are often required to guide treatment choices. Given the technical complexity and occasional unavailability of these techniques, estimation of cardiac function based on clinical examination is valuable for critical care physicians to diagnose circulatory shock. Yet, the lack of knowledge on how to best conduct and teach the clinical examination to estimate cardiac function has reduced its accuracy to almost that of "flipping a coin."

Objective
The aim of this study was to investigate the decision-making process underlying estimates of cardiac function of patients acutely admitted to the intensive care unit (ICU) based on current standardized clinical examination using Bayesian methods.

Methods
Patient data were collected as part of the Simple Intensive Care Studies-I (SICS-I) prospective cohort study. All adult patients consecutively admitted to the ICU with an expected stay longer than 24 hours were included, for whom clinical examination was conducted and cardiac function was estimated. Using these data, first, the probabilistic dependencies between the examiners' estimates and the set of clinically measured variables upon which these rely were analyzed using a Bayesian network. Second, the accuracy of cardiac function estimates was assessed by comparison to the cardiac index values measured by critical care ultrasonography.

Results
A total of 1075 patients were included, of which 783 patients had validated cardiac index measurements. A Bayesian network analysis identified two clinical variables upon which cardiac function estimate is conditionally dependent, namely, noradrenaline administration and presence of delayed capillary refill time or mottling. When the patient received noradrenaline, the probability of cardiac function being estimated as reasonable or good $P(E_{R,G})$ was lower, irrespective of whether the patient was mechanically ventilated ($P[_{ER},G|$ventilation, noradrenaline]=0.63, $P[E_{R,G}|$ventilation, no noradrenaline]=0.91, $P[E_{R,G}|$no ventilation, noradrenaline]=0.67, $P[E_{R,G}|$no ventilation, no noradrenaline]=0.93). The same trend was found for capillary refill time or mottling. Sensitivity of estimating a low cardiac index was 26% and 39% and specificity was 83% and 74% for students and physicians, respectively. Positive and negative likelihood ratios were 1.53 (95% CI 1.19-1.97) and 0.87 (95% CI 0.80-0.95), respectively, overall.

Conclusions
The conditional dependencies between clinical variables and the cardiac function estimates resulted in a network consistent with known physiological relations. Conditional probability queries allow for multiple clinical scenarios to be recreated, which provide insight into the possible thought process underlying the examiners' cardiac function estimates. This information can help develop interactive digital training tools for students and physicians and contribute toward the goal of further improving the diagnostic accuracy of clinical examination in ICU patients.

Introduction

Background

In hemodynamically unstable patients admitted to the intensive care unit (ICU) for circulatory shock, the diagnosis and treatment decisions initially rely on accurate assessment of clinical examination (1,2). Shock is the clinical expression of circulatory failure that results in inadequate cellular oxygen utilization and is often accompanied by systemic arterial hypotension, clinical signs of tissue hypoperfusion, and hyperlactatemia (3). About one-third of critically ill patients experience circulatory shock, which is associated with increased morbidity and mortality (4).

Hemodynamic assessment of critically ill patients is challenging; depending on the type of shock, patients present with highly variable states of circulating blood volume, cardiac contractility, sympathetic nervous activity, vascular tone, and microcirculatory dysfunction. In addition, assessment is even more difficult if comorbidities are present (5). Currently, hemodynamic estimates based on clinical examination show poor association with cardiac index in both univariate and multivariate analyses, and these estimates are no better than flipping a coin (6). Due to this limited ability to assess a patient's hemodynamic status using clinical examination, physicians often base changes in treatment primarily on information obtained through advanced monitoring techniques (7). However, advanced monitoring techniques are currently advised and desired when clinical examination does not lead to a clear diagnosis, or when a patient does not respond to initial therapy (2,8). Therefore, it is important to place emphasis on improving hemodynamic estimates made with clinical examination, to avoid inappropriate overuse of technological aid (9).

The first step in developing improved clinical examination structures for hemodynamic estimates is to study the current clinical practice. To understand how students and physicians diagnosed low cardiac index, Bayesian networks can be used to gain insight into the thought process behind the educated guess on hemodynamic status.

Bayesian networks have been frequently used to model domain knowledge in the context of decision support in other fields of medicine, given their ability to be interpreted as causal networks when no confounders are present (10-13). By combining prior knowledge and the uncertainty in data, Bayesian networks allow for inference tasks to be performed, which establish conditional, possibly causal, dependencies between variables (14). Conditional probabilities queries are interesting tools to study clinical reasoning, which are seen as an additive thought process where, at every step, information is interpreted conditioned on previously acquired information.

Objectives

The aim of this study was to use Bayesian networks to investigate the decision-making process underlying estimates of cardiac function of patients acutely admitted to the ICU, based on current standardized clinical examination using Bayesian methods. Additionally, we aimed to determine the diagnostic accuracy of the current standardized clinical examination for estimating cardiac function in patients acutely admitted to the ICU.

Methods

Design, Setting, and Participants

This study was a predefined substudy of the prospective observational cohort Simple Intensive Care Studies-I (SICS-I) (ClinicalTrial.gov trial registration: NCT02912624) (15). The study was approved by the local institutional review board (METc M15.168207). In SICS-I, all consecutive, acutely admitted adults expected to stay beyond 24 hours were included on their first day of admission to the ICU. Written informed consent was obtained from all patients or their relatives. This study is reported following the Standards for the Reporting of Diagnostic Accuracy Studies guidelines (16).

Aims

The primary aim was to determine the conditional probabilities relating the variables measured during clinical examination to the cardiac function estimate made by the examiners.

The secondary aim of this study was to assess the diagnostic accuracy of cardiac function estimates made by the examiners and compare them to the cardiac index measured by critical care ultrasonography (CCUS).

Bayesian Network Analysis

Bayesian networks are probabilistic models that represent the conditional (in)dependence relations between a set of variables in the form of a directed acyclic graph. In the graph, each variable is represented as a node and the directed edges (arcs) connecting the nodes represent the conditional dependency relations among the variables. Given the conditional (in)dependencies implied by the directed acyclic graph, the joint probability distribution of all variables can be factorized into a product of simpler local probability distributions.

From the initial set of variables registered during clinical examination, 14 clinical variables available from bedside monitors and patient record files, perfusors, physical examination, and the cardiac function estimate were included for modeling (Multimedia Appendix 1). All continuous variables were discretized according to the definitions provided in the study protocol. The correlation coefficients between variables after discretization were calculated with the Cramér V test for correlation strength.

The network structure was learned using the Max-Min Hill-Climbing algorithm with the Bayesian-Dirichlet equivalent scoring metric, as implemented in the R package "bnlearn" (17). The Max-Min Hill-Climbing algorithm searches for the best network structure (ie, the best directed acyclic graph) that maximizes the Bayesian-Dirichlet equivalent scoring metric. To this end, the algorithm starts with an initial directed acyclic graph and then improves the Bayesian-Dirichlet equivalent score by iteratively adding, deleting, and reversing individual edges until the Bayesian-Dirichlet equivalent score does not improve further (18).

A set of restrictions can be applied to enforce certain connections between arcs in the network, so that prior knowledge is implemented *a priori* (13). Arcs representing known dependencies can be whitelisted (ie, forced to appear in the directed acyclic graph), while arcs that represent impossible dependencies can be blacklisted (i.e., excluded from the directed acyclic graph). In this network, *age* and *gender* are not determined by any other variables, so all arcs from other variables to these two were blacklisted. Similarly, as *estimate* does not influence any clinical variable, any arc from *estimate* to other variables was also blacklisted.

After the restrictions are defined, to obtain a confidence measure for the presence and directionality of the individual network edges, the bootstrap technique was applied. R=2000 bootstrap samples were generated from the original data, and the Max-Min Hill-Climbing algorithm was used to search for the best network for each bootstrap data set. This gives R=2000 best networks, and the confidence on the presence of an edge ranges from 0 (learned from 0 bootstrap samples) to 1 (learned from all bootstrap samples) (13). To further increase the robustness of the final or consensus network, we defined the minimum significance threshold for arc strength as 0.700 if the calculated significance threshold was lower and accepted the calculated threshold otherwise. Regarding directionality, arcs with a direction coefficient below 0.666 after bootstrapping were considered undirected.

To determine the distributions of the variables and calculate the associated probabilities of the network, the adjacency matrix of the average bootstrapped directed acyclic graph was reproduced using the Bayesian network function, and belief propagation was carried out using the *gRain* package (13,19).

Belief propagation allows for inference tasks (probability queries) to be performed on the learned Bayesian networks, thereby providing a calculation of the distribution of values of a certain variable and the marginal and conditional probabilities of these values occurring based on the known value of an observed variable. Given a certain distribution, the marginal probability of a certain value occurring is calculated by integrating out all other variables, while the conditional probability is the probability of a value occurring for one variable, given a known, fixed value for at least one other variable (20). These probability queries will allow for multiple relevant clinical scenarios to be recreated, based on the consensus network and the properties of the Markov blanket. When carrying out a query for *estimate*, if the values of its parent nodes are known, no other node can influence the conditional distribution of *estimate* (21). If only some of its parent nodes are known, however, then some of the ancestors upstream of the undefined parent nodes can still influence the conditional probability of *estimate* (21). To validate the structure learning process beyond the bootstrapping strategy used in learning a consensus network, two steps were taken. First, an ad hoc expert analysis was conducted to assess the plausibility and accuracy of the physiological relationships identified in the network. Second, 10-fold cross-validation was used to determine its predictive accuracy. Using the consensus network, the accuracy of the cross-validated predictions was determined by dichotomizing the estimates as described below and by calculating the area under the receiver operating curve, specificity, and sensitivity of the predictions made for patients, from which a validated cardiac index measurement was available.

Definitions and Bias

Patients underwent a protocolized and standardized clinical examination and subsequent CCUS, as described in the SICS-I protocol (15). The main variable of interest was cardiac function estimation made by the student or physician after clinical examination was performed but before CCUS was performed. Examiners could score cardiac function as "poor," "moderate," "reasonable," or "good." For diagnostic test analyses and the validation step of the network structure, the "poor" and "moderate" estimates were grouped as "low," and the "reasonable" or "good" estimates were grouped as "high." Quality of the CCUS images and measurements of cardiac index were validated by core laboratory technicians (Groningen Image Core Lab, Groningen, The Netherlands) who were blinded for the rest of the measurements. Cardiac index measurements were categorized in two groups: "low" for cardiac index≤2.2 L/min/m^2 and "high" for cardiac index>2.2 L/min/m^2 (22). All patients for whom a validated cardiac index measurement and estimate of cardiac function were available were included in the Bayesian network analysis. Patients for whom CCUS images were of insufficient quality or cardiac index measurements were not available, were excluded from the diagnostic accuracy analysis.

Statistical Analysis

Due to the observational nature of the study, a formal sample size calculation was not possible. Statistical analyses were performed in STATA 15.0 (StataCorp, College Station, Texas) and R version 3.5.1 (R Core Team, Vienna, Austria). Data are presented as mean with SD when normally distributed, or as median with interquartile range in case of skewed data. Dichotomous and categorical data are presented in proportions. Sensitivity and specificity for both the network's and the examiners' estimated guess were calculated by cross-tabulation of the respective predictions and the validated cardiac index measurements. Additionally, positive predictive values (PPV) and negative predictive values (NPV) and positive likelihood ratios (LR+) and negative likelihood ratios (LR-) were calculated with 95% CIs for the examiners' estimates. For these, the overall accuracy was further expressed as a proportion of correctly classified cardiac index measurements (true negative and true positive measures) among all measures.

Results

Participants

A total of 1075 patients fulfilled our inclusion criteria, of which 1073 patients had available cardiac function estimates and were therefore included in the Bayesian network analysis. Of the included patients, 783 (73%) had validated cardiac index measurements and were included in the diagnostic accuracy tests. Further, 569 patients (73%) were included by students and 214 patients (27%) were included by physicians.

Descriptive Measures

Characteristics of included patients according to availability of cardiac index measurements are shown in Table 1. Body mass index and Simplified Acute Physiology Score (SAPS) II score were significantly different between patients (Table 1).

Bayesian Network Analysis

The structure learned for the network identified two clinical variables, namely, noradrenaline administration and the presence of delayed capillary refill time or mottling (dCRT-M), upon which the estimates of cardiac function are directly conditionally dependent (Table 2).

Table 1. *Patient characteristics Variable*

Variable		No cardiac index measurement (N=292)	Cardiac index measurement (N=783)	Total (N=1075)	P value
Age (years)		62 ± 14	62 ± 15	62 ± 15	0.75
Male gender		188 (64%)	486 (62%)	674 (63%)	0.49
Body mass index (kg/m^2)		27.5 (5.4)	26.7 (5.6)	26.9 (5.5)	0.043
Mean Arterial Pressure (mmHg)		78 ± 14	79 ± 14	78 ± 14	0.30
Heart rate (bpm)		87 ± 22	88 ± 21	88 ± 21	0.35
Irregular heart rhythm		28 (10%)	88 (11%)	116 (11%)	0.44
Central venous pressure (mmHg)		9 (5, 12)	9 (5, 13)	9 (5, 13)	0.74
Patients administered noradrenaline		142 (49%)	386 (49%)	528 (49%)	0.85
Urine output (mL/kg/h)		0.6 (0.3, 1.2)	0.7 (0.4, 1.2)	0.6 (0.4, 1.2)	0.22
Respiratory rate (bpm)		18 ± 5	18 ± 6	18 ± 6	0.50
Mechanical ventilation		179 (61%)	452 (58%)	631 (59%)	0.29
Positive End-Expiratory Pressure (cm H_2O)		7 (5, 8)	7 (5, 8)	7 (5, 8)	0.41
Central temperature (°C)		37.0 ± 0.9	36.9 ± 0.9	36.9 ± 0.9	0.84
Difference between central temperature and temperature on the dorsum of the foot (°C)		7.7 ± 3.2	7.8 ± 3.2	7.8 ± 3.2	0.66
Subjective "cold" temperature		109 (37.6%)	289 (37.1%)	398 (37.2%)	0.88
Capillary refill time knee (sec)		3.0 (2.0, 4.5)	3.0 (2.0, 4.5)	3.0 (2.0, 4.5)	0.48
Capillary refill time sternum (sec)		2.8 (2.0, 3.0)	3.0 (2.0, 3.0)	3.0 (2.0, 3.0)	0.84
Capillary refill time finger (sec)		3.0 (2.0, 4.0)	2.5 (2.0, 4.0)	2.5 (2.0, 4.0)	0.37
Mottling rate	None	157 (58.8%)	397 (56.8%)	554 (57.3%)	0.64
	Mild	24 (9.0%)	79 (11.3%)	103 (10.7%)	
	Moderate	75 (28.1%)	201 (28.8%)	276 (28.6%)	
	Severe	11 (4.1%)	22 (3.1%)	33 (3.4%)	
Hemoglobin (mmol/L)		6.8 ± 1.5	6.8 ± 1.4	6.8 ± 1.4	0.90
Lactate (mmol/L)		1.4 (0.9, 2.4)	1.4 (0.9, 2.2)	1.4 (0.9, 2.2)	0.79
ICU Length of Stay (days)		3.5 (1.9, 6.9)	3.1 (1.9, 6.5)	3.2 (1.9, 6.6)	0.29
SAPS II (points)		47 (37, 58)	44 (34, 56)	45 (35, 57)	0.037
APACHE IV score (points)		77 (56, 92)	73 (55, 91)	74 (56, 92)	0.14
90-day Mortality (n, %)		81 (27.7%)	217 (27.7%)	298 (27.7%)	0.99
Cardiac function estimate	Poor	8 (2.8%)	18 (2.3%)	26 (2.4%)	0.004
	Moderate	46 (15.9%)	165 (21.1%)	211 (19.7%)	
	Reasonable	164 (56.6%)	349 (44.6%)	513 (47.8%)	
	Good	72 (24.8%)	251 (32.1%)	323 (30.1%)	

Abbreviations: SAPS, Simplified Acute Physiology Score; APACHE, Acute Physiology And Chronic Health Evaluation

Table 2. Strength and direction coefficients of the consensus directed acyclic graph

From	To	Strength	Direction
Age	Irregular rhythm	0.983	1.00
Mechanically ventilated	High respiratory rate	0.994	0.504
Mechanically ventilated	dCRT-M	0.875	0.884
Irregular rhythm	Tachycardia	0.848	0.954
Tachycardia	High respiratory rate	0.999	0.931
Tachycardia	Low SBP	0.821	0.883
Tachycardia	Elevated lactate	0.832	0.821
Low SBP	Low MAP	1	1
Low DBP	Low MAP	1	1
Elevated lactate	Oliguria	0.728	0.803
Elevated lactate	Noradrenaline administration	1	1
Noradrenaline administration	Mechanically ventilated	1	0.957
Noradrenaline administration	Estimate	0.999	1
dCRT-M	Estimate	0.876	1

Abbreviations: dCRT-M, delayed capillary refill time or mottling; SBP, systolic blood pressure; MAP, mean arterial pressure; DBP, diastolic blood pressure

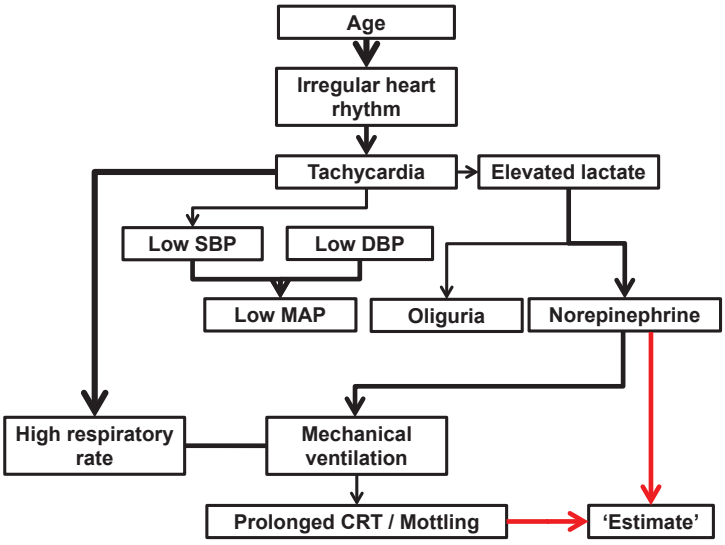


Figure 1. Consensus directed acyclic graph. Red lines represent direct conditional dependencies to estimate. Black lines represent direct conditional dependencies to other variables. Width of the line represents strength coefficient. The dotted line represents the weakest strength coefficient. DBP, diastolic blood pressure; SBP, systolic blood pressure; MAP, mean arterial pressure; CRT, capillary refill time.

As denoted in Figure 1 by the dotted line, the arc from elevated lactate to oliguria had the lowest strength coefficient (0.728). The average directionality coefficient was 0.909, indicating well-defined directionality. Only one edge (between mechanical ventilation and high respiratory rate) did not meet the threshold for directionality and was thereby left undirected in the consensus directed acyclic graph (for querying, however, a direction from high respiratory rate to mechanical ventilation was defined based on expert knowledge to comply with the formal computational requirements) (15). Additionally, there was no difference in network structure when including only students (n=801) or only physicians (n=271) compared to the network obtained with all the participants' estimates.

The probability queries conducted with the conditional probabilities for estimate are presented in a tree diagram in Figure 2. Each of the pathways in the diagram represents a scenario that could occur during clinical examination. Since one of the main focuses of SICS-I was the collection and interpretation of information available at bedside during physical examination, we expanded the conditional probability queries to also include respiratory rate and mechanical ventilation. Tachypnea virtually did not influence the probability of cardiac pump function being estimated as reasonable or good $P(E_{R,G})$, whereas ventilation status did ($P[E_{R,G}|$not ventilated, no tachypnea$]=P[E_{R,G}|$not ventilated, tachypnea$)=0.85$; $P[E_{R,G}|$ventilated, tachypnea$]=0.69$ and $P[E_{R,G}|$ventilated, no tachypnea$]=0.63$). When the patient received noradrenaline, $P(E_{R,G})$ was lower irrespective of whether they were mechanically ventilated ($P[E_{R,G}|$ventilation, noradrenaline$]=0.63$, $P[E_{R,G}|$ventilation, no noradrenaline$]=0.91$, $P[E_{R,G}|$no ventilation, noradrenaline$]=0.67$, $P[E_{R,G}|$no ventilation, no noradrenaline$]=0.93$). The same trend was found for dCRT-M, with reasonable or good estimates being more likely in the absence of dCRT-M.

Finally, an area under the receiver operating characteristic curve of 0.58 was obtained for the 10-fold cross-validated predictions of cardiac function made by the consensus network, with a specificity of 36% and a sensitivity of 79% (23).

Diagnostic Accuracy

Diagnostic accuracy tests for estimating of a low cardiac index showed a sensitivity of 26% and 39%, a specificity of 83% and 74%, PPV of 45% and 48%, NPV of 67% and 66%, LR+ of 1.52 and 1.52, and LR- of 0.89 and 0.82 for students and physicians, respectively. The overall accuracy of cardiac index estimates was 63% and 61% for students and physicians, respectively. For all patients combined, sensitivity was 30%, specificity was 80%, PPV was 46%, NPV was 67%, LR+ was 1.53, LR- was 0.87, and the overall accuracy of diagnostic tests was 62% (Table 3).

Table 3. Accuracy, sensitivity, specificity, predictive values, and likelihood ratios for students' and physicians' estimates

Variable	Students (N=569)	Physicians (N=214)	All Groups (N=783)
Sensitivity (%)	26 (20 – 33)	39 (28 – 50)	30 (25 – 36)
Specificity (%)	83 (78 - 86)	74 (66 – 82)	80 (77 – 84)
Positive predictive value (%)	45 (38 - 53)	48 (39 – 58)	46 (40 – 53)
Negative predictive value (%)	67 (65 - 69)	66 (61 – 71)	67 (65 – 69)
Positive likelihood ratio	1.52 (1.10 – 2.09)	1.52 (1.02 – 2.25)	1.53 (1.19 – 1.97)
Negative likelihood ratio	0.89 (0.81 – 0.98)	0.82 (0.67 – 1.00)	0.87 (0.80 – 0.95)
Overall accuracy (%)	63 (59 – 67)	61 (54 – 67)	62 (59 – 66)

All values are presented with 95% confidence interval.

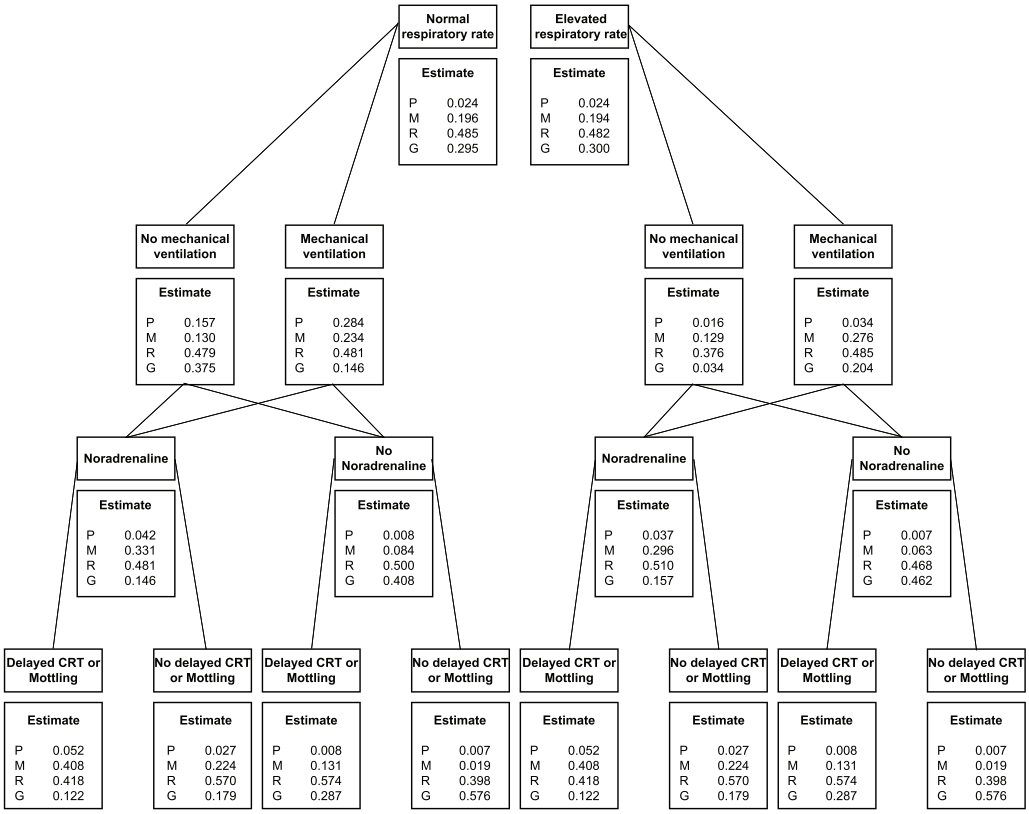


Figure 2. Tree diagram showing the conditional probabilities queries for estimate associated with multiple scenarios during clinical examination. At each step, only the variables above the split are known and as more information becomes available, the conditional probabilities change. P, Poor; M, Moderate; R, Reasonable; G, Good; CRT, capillary refill time

Discussion

Principal Findings

Clinical examination is used daily by physicians as an easy, cheap, and noninvasive way of gathering information to guide interventions and further diagnostic testing. Clinical signs such as oliguria; altered consciousness; and cold, clammy skin are known possible indicators of organ hypoperfusion and are used to diagnose shock in critically ill patients (2). However, the value of clinical examination has been questioned, and previous studies have shown physicians to perform poorly in diagnosing a low cardiac index based on physical signs alone (8,9). In this study, we confirmed that the accuracy of these estimates remains low for both students and physicians. Surprisingly, we identified noradrenaline administration and delayed CRT or mottling as seemingly the major factors influencing cardiac function estimates using Bayesian network analysis. These findings may serve as the basis for improving the value of clinical examination (1) by identifying some of the biases clinicians may be subjected to, which causes them to overdiagnose compared to students, and (2) by clarifying some of the thought process behind the clinical examination. This allows the examiner to "think about how they think" when performing clinical examination and can help clinicians be trained to prioritize or leave out certain variables when making their assessment.

Bayesian Network Analysis

Validation and Limitations

Validation of the network structure was a crucial yet challenging step toward our goal of trying to obtain a plausible representation of the examiners' knowledge network and thought process at bedside. We believe to have tackled this challenge in the best way possible by validating it in three different ways: using the bootstrapping process to generate a consensus network; conducting expert validation of the plausibility of the arcs; and using the network as a predictor, as previously suggested (13). We believe that the similarity in accuracy, sensitivity, and specificity between the network's predictions and the examiners' own estimates is further proof of the validity of its structure. It must be restated that the goal of this study was not to build and optimize a predictive model, in which case the predictive accuracy, sensitivity, and specificity we obtained would be subpar. In fact, had the network been able to make the estimates with a substantially higher accuracy than the examiners' estimates, we would be more reluctant to affirm that is parallel with the examiner's thought process.

As any exploratory study, however, we faced several limitations. The first was practical, as not all included patients had cardiac index measurements, since CCUS is not applicable for every ICU patient and views obtained by CCUS can be obstructed due to lines, wounds, or excess adiposity (24). This prevented us from using the complete cohort and likely accounted for the difference in SAPS-II score and body mass index in the patients with and without CCUS measurements. Second, the discretization required by the parametric assumptions of Bayesian network algorithms comes with the inherent risk of useful information being discarded in the process, which does

not guarantee that the dependence relationships involving the original variables are preserved. Last, for causality to be derived from Bayesian networks, there must be no unobserved variables influencing the variables included in the network that may act as confounding factors. In SICS-I, the focus was on examining and improving students' and physicians' educated guess, resorting primarily to bedside information, such as vasopressor and fluid perfusors, vital signs, and physical examination. Therefore, to best replicate this scenario, we opted to include in the network only variables that are readily available during the protocolized examination. Although this increases the risk of introducing bias in the causal network, the accuracy of the physiological dependencies identified gives us reason to believe that no substantial bias is present.

Do Probability Queries Help Explain the Modest Diagnostic Accuracy?

Previous studies on the diagnostic accuracy of clinical examination have found the performance of experienced physicians and students to be comparable (6). Expert physicians are more often affected by multiple cognitive biases, such as confirmatory bias and premature closure, compared to students, who remain more open to new hypotheses and persist in collecting data (25,26). Interestingly, while the diagnostic accuracy for individual physicians can be as low as 62.5%, there is a visible increase as the number of physicians involved increases (up to 85.6% for groups of nine physicians) (27). Our results are in line with the literature, and we additionally showed that physicians had a higher sensitivity but lower specificity than students (39% and 26%, and 74% and 83%, respectively). These differences in sensitivity and specificity represent a tendency of physicians to overdiagnose, which has previously been related to confirmatory bias and premature closure. Indeed, two other findings support the idea already given by the direct dependence of *estimate* solely on noradrenaline and dCRT-M that premature closure was a common phenomenon. First, in the probability queries, while machine ventilation does not directly influence the *estimate*, considerable changes in the probability of the *estimate* are still observable, depending on whether the patient is ventilated, before noradrenaline use and dCRT-M are known. This could be due to the fact that mechanical ventilation is almost inevitably the first variable to be noted when the examiner approaches bedside. Second, a comparison of the change in the probabilities of *estimate* based on varying clinical evidence with the likelihood ratios calculated in another SICS-I substudy shows that variables further upstream of *estimate* such as respiratory should be taken more into account (15). For example, while the positive and negative likelihood ratios of a high respiratory rate are as suggestive as those of a delayed CRT, the query shows that the probability of being estimated to have low cardiac function was considerably lower in those without dCRT-M (0.25) than in those with dCRT-M (0.46) and the probability of a patient with tachypnea being estimated to have low or high cardiac function was virtually the same. This is despite tachypnea having a positive and negative likelihood ratio of 1.16 and 0.68, respectively.

Conclusion and Future Implications

This study confirms that the accuracy of cardiac function estimates remains low for both students and physicians, and it identifies noradrenaline administration and delayed CRT or mottling as seemingly the major factors influencing these estimates. Although it will remain challenging to try to replicate the thought process of the examiner, not only methodologically, but also because different individuals have different levels of knowledge and different examination routines, Bayesian networks seem like a promising tool to help break down and better understand the educated guessing process. The insight gained in studies such as this one, can help teach students think about how they think and, on a clinical level, provide much-needed guidance for prioritization of variables during clinical examination. In fact, our team is currently compiling the knowledge acquired in the SICS-I substudies to build an interactive game for medical students, residents, and specialists. This electronic learning tool will ask the player to estimate cardiac function using the same scale and data from variables such as bedside monitor hemodynamic variables, ventilator and pump settings, and urine output.

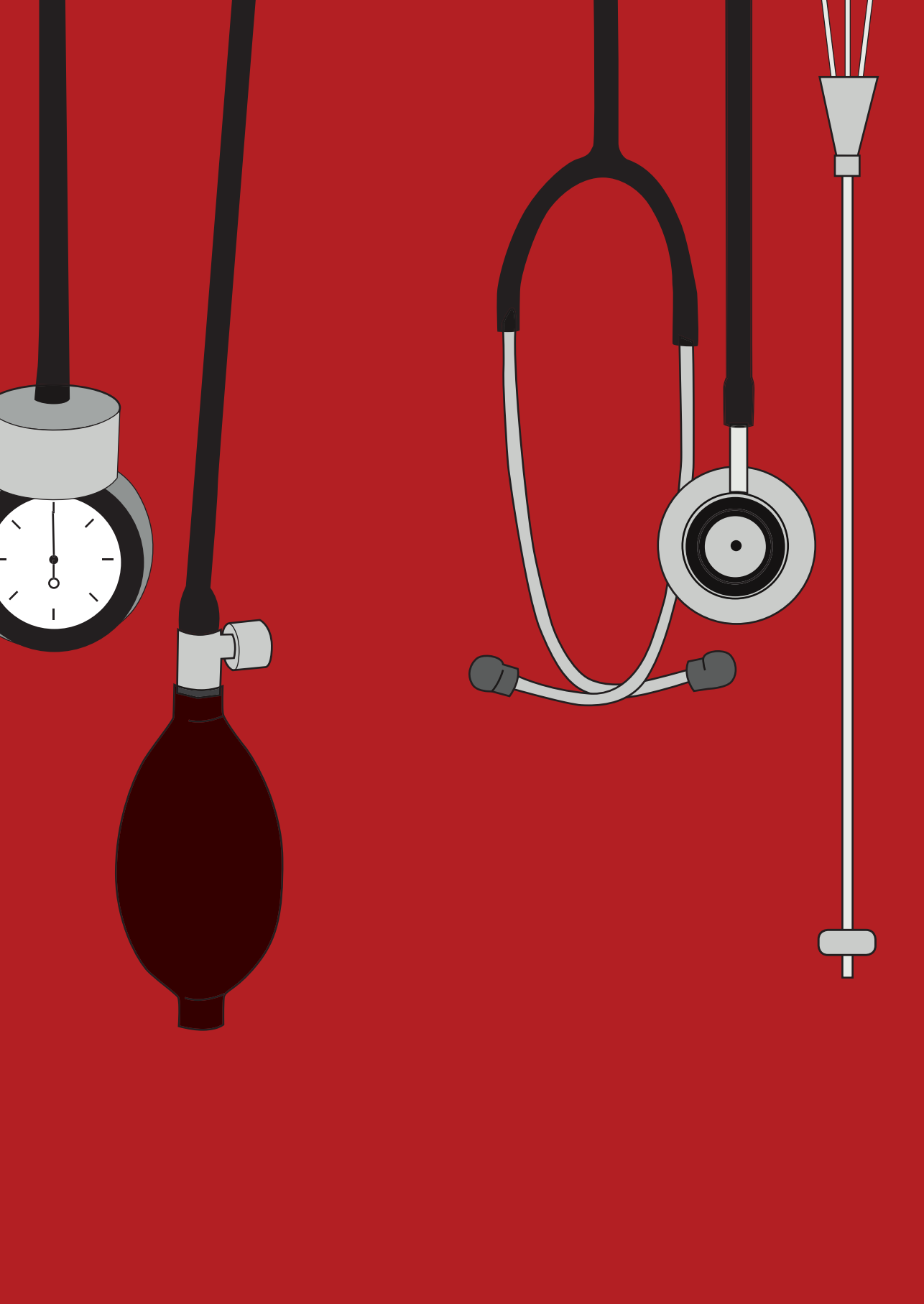
Supplementary material

Supplements are available online:
https://doi.org/10.2196/15358

References

1 Sevransky J. Clinical assessment of hemodynamically unstable patients. Curr Opin Crit Care. 2009 Jun;15(3):234–8.
2 Cecconi M, De Backer D, Antonelli M, Beale R, Bakker J, Hofer C, et al. Consensus on circulatory shock and hemodynamic monitoring. Task force of the European Society of Intensive Care Medicine. Intensive Care Med. 2014 Dec 13;40(12):1795–815.
3 Vincent J, de Backer D. Circulatory shock. N Engl J Med. 2013 Oct 31;369(18):1726-1734
4 Sakr Y, Reinhart K, Vincent J, Sprung CL, Moreno R, Ranieri VM, et al. Does dopamine administration in shock influence outcome? Results of the Sepsis Occurrence in Acutely Ill Patients (SOAP) Study. Crit Care Med. 2006 Mar;34(3):589-597.
5 Saugel B, Malbrain MLNG, Perel A. Hemodynamic monitoring in the era of evidence-based medicine. Crit Care. 2016 Dec 20;20(1):401.
6 Hiemstra B, Eck RJ, Keus F, van der Horst ICC. Clinical examination for diagnosing circulatory shock. Curr Opin Crit Care. 2017 Aug;23(4):293–301.
7 Perel A, Saugel B, Teboul J-L, Malbrain MLNG, Belda FJ, Fernández-Mondéjar E, et al. The effects of advanced monitoring on hemodynamic management in critically ill patients: a pre and post questionnaire study. J Clin Monit Comput. 2016 Oct 10;30(5):511–8.
8 Suess E, Pinsky M. Hemodynamic Monitoring for the Evaluation and Treatment of Shock: What Is the Current State of the Art? Semin Respir Crit Care Med. 2015 Nov 23;36(06):890–8.
9 Elder A, Japp A, Verghese A. How valuable is physical examination of the cardiovascular system? BMJ. 2016 Jul 27;354:i3309.
10 Soto-Ferrari M, Prieto D, Munene G. A Bayesian network and heuristic approach for systematic characterization of radiotherapy receipt after breast-conservation surgery. BMC Med Inform Decis Mak. 2017 Dec 28;17(1):93.
11 Stivaros SM, Gledson A, Nenadic G, Zeng X-J, Keane J, Jackson A. Decision support systems for clinical radiological practice — towards the next generation. Br J Radiol. 2010 Nov;83(995):904–14.
12 Peelen L, de Keizer NF, Jonge E de, Bosman R-J, Abu-Hanna A, Peek N. Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit. J Biomed Inform. 2010 Apr;43(2):273–86.
13 Scutari M, Auconi P, Caldarelli G, Franchi L. Bayesian Networks Analysis of Malocclusion Data. Sci Rep. 2017 Dec 10;7(1):15236.
14 Mukherjee S, Speed TP. Network inference using informative priors. Proc Natl Acad Sci. 2008 Sep 23;105(38):14313–8.
15 Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, et al. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med. 2019 Feb;45(2):190–200.
16 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015 Oct 28;351(3):h5527.
17 Scutari M. Learning Bayesian Networks with the bnlearn R Package. J Stat Softw. 2010;35(3):1–22.
18 Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn. 2006 Oct 28;65(1):31–78.
19 Hojsgaard S. Graphical Independence Networks with the gRain Package for R. J Stat Softw. 2012;46(10).
20 Annis C. Joint, Marginal, and Conditional Distributions. 2014 [cited 2019 Feb 13]. Available from: http://www.statisticalengineering.com/joint_marginal_conditional.htm

21 Barber D. Bayesian Reasoning and Machine Learning. New York: Cambridge University Press;2012.
22 Forrester JS, Diamond GA, Swan HJ. Correlative classification of clinical and hemodynamicfunction after acute myocardial infarction. Am J Cardiol. 1977 Feb;39(2):137–45.
23 Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17;12:77.
24 Koster G, Van Der Horst ICC. Critical care ultrasonography in circulatory shock. Curr Opin Crit Care. 2017 Aug;23(4):326–33.
25 Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medicaldecisions: a systematic review. BMC Med Inform Decis Mak. 2016 Dec 3;16(1):138.
26 Krupat E, Wormwood J, Schwartzstein RM, Richards JB. Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors. Med Educ. 2017 Nov;51(11):1127–37.
27 Barnett ML, Boddupalli D, Nundy S, Bates DW. Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians. JAMA Netw open. 2019 Mar 1;2(3):e190096.



5

Non-invasive oscillometric versus invasive arterial blood pressure measurements in critically ill patients: a post hoc analysis of a prospective observational study

Thomas Kaufmann
Eline G.M. Cox
Renske Wiersema
Bart Hiemstra
Ruben J. Eck
Geert Koster
Thomas W.L. Scheeren
Frederik Keus
Bernd Saugel
Iwan C.C. van der Horst
SICS Study Group

Journal of Critical Care

Abstract

Purpose
The aim was to compare non-invasive blood pressure measurements with invasive blood pressure measurements in critically ill patients.

Methods
Non-invasive blood pressure was measured via automated brachial cuff oscillometry, and simultaneously the radial arterial catheter-derived measurement was recorded as part of a prospective observational study. Measurements of systolic arterial pressure (SAP), diastolic arterial pressure (DAP), and mean arterial pressure (MAP) were compared using Bland-Altman and error grid analyses.

Results
Paired measurements of blood pressure were available for 736 patients. Observed mean difference (±SD, 95% limits of agreement) between oscillometrically and invasively measured blood pressure was 0.8 mmHg (±15.7 mmHg, -30.2 to 31.7 mmHg) for SAP, -2.9 mmHg (±11.0 mmHg, -24.5 to 18.6 mmHg) for DAP, and -1.0 mmHg (±10.2 mmHg, -21.0 to 18.9 mmHg) for MAP. Error grid analysis showed that the proportions of measurements in risk zones A to E were 78.3%, 20.7%, 1.0%, 0%, and 0.1% for MAP.

Conclusion
Non-invasive blood pressure measurements using brachial cuff oscillometry showed large limits of agreement compared to invasive measurements in critically ill patients. Error grid analysis showed that measurement differences between oscillometry and the arterial catheter would potentially have triggered at least low-risk treatment decisions in one in five patients.

Introduction

Blood pressure is one of the most frequently measured vital signs and can be measured non-invasively or invasively (1). Measurements of blood pressure are commonly used as triggers and targets to guide hemodynamic interventions, especially in critically ill patients treated in the intensive care unit (ICU) (2). Circulatory shock is a common condition in critically ill patients, affecting about one-third of all admitted ICU patients (3). Systemic arterial hypotension is typically present in circulatory shock with associated tachycardia and signs of altered tissue perfusion (4). Guidelines on hemodynamic monitoring and circulatory shock advocate initially targeting a mean arterial pressure (MAP) above 65 mmHg (2). Therefore, clinicians need to have quick, reliable, and accurate measurements of blood pressure available at the bedside.

Invasive arterial blood pressure monitoring using an arterial catheter is considered the clinical reference method in critically ill patients. In situ arterial catheters also facilitate drawing blood for laboratory testing and blood gas analysis. Even though the guidelines recommend the placement of an arterial catheter for invasive monitoring in patients with suspected circulatory shock (2), non-invasive oscillometric blood pressure measurements using an upper-arm cuff are a widely used alternative (5). This automated method allows for intermittent quick and convenient blood pressure measurements.

Several prospective studies showed an acceptable agreement of non-invasive oscillometric blood pressure measurements with invasive reference measurements in critically ill patients (6,7). Some say that intermittent non-invasive oscillometric blood pressure measurements may even safely replace invasive measurements (8). Other studies showed an unacceptable measurement performance of oscillometry in critically ill patients with circulatory shock, demonstrating a possible influence of the shock state and the use of vasoactive medication on the measurement performance (9,10).

This study aimed to compare blood pressure measurements obtained using upper-arm cuff oscillometry with arterial catheter-derived blood pressure measurements in a large prospective cohort of critically ill patients with and without receiving norepinephrine.

Material and Methods

Design and setting

This study was part of the Simple Intensive Care Studies-I (SICS-I), a prospective single-center observational cohort study designed to evaluate the diagnostic and prognostic value of combinations of clinical and hemodynamic variables in critically ill patients (11). We performed the study between 27 March 2015 and 22 July 2017. The local institutional review board approved the study (M15.168207).

Participants and study size
In SICS-I, all acutely admitted patients of 18 years and older with an expected ICU stay of at least 24 hours were eligible for inclusion. Exclusion criteria were planned admission, inability to acquire research data due to interference with clinical care, and absence of informed consent. Patients without either invasive blood pressure or non-invasive blood pressure measurement were excluded from this analysis.

Objectives
The primary objective of this study was to evaluate the agreement between blood pressure measurements obtained with non-invasive oscillometry (test method) and arterial catheter-derived measurements (reference method) using Bland-Altman and error-grid analyses. The agreement between measurements was defined according to the Association for the Advancement of Medical Instrumentation (AAMI) standards for non-invasive arterial pressure measurement (12). The AAMI definition of an acceptable agreement in adults between the test and reference method is a mean of the differences of ≤5 mmHg with a SD of ≤8 mmHg (12).

The secondary objective was to analyze the differences between the two methods separately in patients with and without receiving norepinephrine.

Variables
For the SICS-I study, clinical and hemodynamic variables were collected during a one-time clinical examination in the first 24 hours of patient admission. Study procedures were only performed when there was no interference with clinical care. Complete data management was described in the design paper of the SICS-I (13). Reference blood pressure was measured invasively using an indwelling arterial catheter placed in the radial artery. Non-invasive oscillometric blood pressure measurements were performed using an upper-arm cuff placed on the arm contralateral to the arm with the arterial catheter. The correct cuff size was estimated for each patient by the nurse, and the arterial catheter transducer was zeroed and leveled. Blood pressure data were recorded simultaneously from the display of the bedside monitor IntelliVue MP70 (Philips, Eindhoven, The Netherlands). The dose of norepinephrine infusion was documented at the time of the blood pressure measurements.

Statistical analysis
Data are presented as means with standard deviations (SD), medians with 25th and 75th percentile, or absolute numbers (with percentages). Student's T-test, Mann-Whitney U test, or the Chi-square tests were used as appropriate. Correlations between non-invasive and invasive blood pressure measurements are illustrated using scatter plots. The agreement was assessed using Bland-Altman plots, by plotting the mean of the two measurements against their difference and 95% limits of agreement (LOA) (= mean difference ± 1.96 × SD of the difference) (14).

Error grid analysis was used to assess the clinical relevance of differences between the two methods (15). Error grid analysis assigns a specific risk level value to each pair of measured arterial

pressures. Risk levels range from zones A to E with A representing no risk (i.e., no difference in clinical action between the reference and test method), B representing low-risk (i.e., test method values that deviate from the reference but would probably lead to benign or no treatment), C representing moderate risk (i.e., test method values that differ from the reference and would eventually lead to unnecessary treatment with potential moderate non-life-threatening consequences for the patient), D representing significant risk (i.e., test method values that deviate from the reference and would lead to unnecessary treatment with potential severe non-life-threatening consequences for the patient), and E representing dangerous risks (i.e., test method values deviate from the reference method and would lead to unnecessary treatment with potentially life-threatening consequences for the patient) (15). The clinical relevance of the difference between two methods is reflected by the proportion of measurements in each risk level; i.e., higher proportions in the low-risk level indicate a lower clinical relevance of the difference. The risk levels were quantified for systolic arterial pressure (SAP) and mean arterial pressure (MAP) by consensus among 25 international experts in anesthesiology and intensive care medicine (15). No consensus error grid for diastolic arterial pressure (DAP) was quantified by experts due to its limited use in anesthesiology and critical care as an isolated value. The proportions of measurements in the five risk levels were calculated and were visualized in a grid in which the consensus risk assessment is converted into a continuous risk level ranging from 0% to 100% (16). The Chi-square test was used to compare the proportions of measurements in risk zone A versus risk zone B-E between patients with and without norepinephrine.

For statistical analysis, Microsoft Office Excel 2010 (Microsoft Corp, Redmond, WA, USA) and Stata version 15 (StataCorp, College Station, TX, USA) were used. Continuous error grids were constructed using a computer program written for MATLAB version 2018b (The MathWorks Inc., Natick, MA, USA) (16).

Results

Overall, 1075 patients were included in the SICS-I study (11). Of these, 1052 patients (98%) had an invasive arterial pressure measurement, and 757 patients (70%) had a non-invasive arterial pressure measurement. Seven hundred thirty-six patients (68%) had paired blood pressure measurements, i.e., simultaneously measured using oscillometry and an arterial catheter (Figure 1). Of 736 patients with a paired measurement, 352 patients (48%) received norepinephrine during the blood pressure measurements. Table 1 presents the characteristics of the 736 included patients, and the 339 excluded patients for this study based on the availability of paired blood pressure measurement.

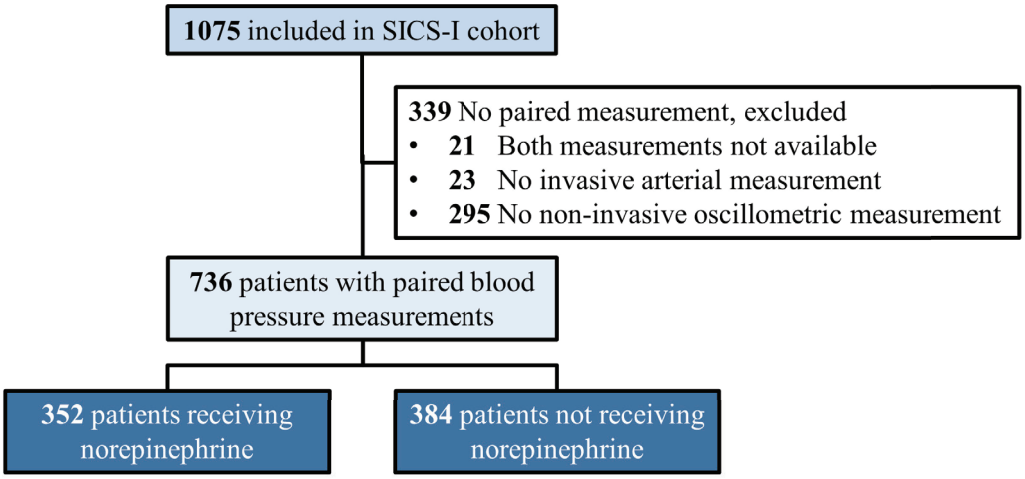


Figure 1. *Flowchart showing patient inclusion in this study.*

Patients without a paired measurement of invasive and non-invasive blood pressure had a higher body mass index (BMI), were more often admitted for a cardiovascular reason, and were more critically ill as reflected by a higher APACHE-IV score at admission and increased 90-day mortality (Table 1).

The distribution and relation of arterial pressure data obtained non-invasively and invasively is illustrated in scatter plots (Supplementary Fig 1 in Supplementary Material).

Bland-Altman analysis of all paired measurements revealed a mean difference (±SD, 95% limits of agreement) of 0.8 mmHg (±15.7 mmHg, -30.2 to 31.7 mmHg) for SAP, -2.9 mmHg (±11.0 mmHg, -24.5 to 18.6 mmHg) for DAP, and -1.0 mmHg (±10.2 mmHg, -21.0 to 18.9 mmHg) for MAP (Table 2; Figure 2).

Bland-Altman analysis of paired measurements in patients receiving norepinephrine and patients not receiving norepinephrine showed similar results. In patients receiving norepinephrine, Bland-Altman analysis showed a mean difference (±SD, 95% LOA) of -1.1 mmHg (±16.1 mmHg, -32.8 to 30.6 mmHg) for SAP, -1.9 mmHg (±10.0 mmHg, -21.5 to 17.7 mmHg) for DAP, and -1.0 mmHg (±9.9 mmHg, -20.4 to 18.5 mmHg) for MAP (Table 2; Supplementary Fig 2 in Supplementary Material).

In patients not receiving norepinephrine, Bland-Altman analysis showed a mean difference (±SD, 95% LOA) of 2.4 mmHg (±15.2 mmHg, -27.5 to 32.4 mmHg) for SAP, -3.9 mmHg (±11.8 mmHg, -27.0 to 19.3 mmHg) for DAP, and -1.1 mmHg (±10.4 mmHg, -21.5 to 19.4 mmHg) for MAP (Table 2; Supplementary Fig 3 in Supplementary Material).

In all patients with a paired measurement of blood pressure, error grid analysis showed that the proportions of measurements in risk zones A to E were 82.3%, 13.2%, 4.2%, 0.3%, and 0% for SAP and 78.3%, 20.7%, 1.0%, 0%, and 0.1% for MAP. Continuous error grids for SAP and MAP are shown in figure 3.

In patients receiving norepinephrine, error grid analysis showed that the proportions of measurements in risk zones A to E were 75.6%, 17.6%, 6.5%, 0.3%, and 0% for SAP and 74.7%, 23.6%, 1.7%, 0%, and 0% for MAP (Supplementary Fig 4 in Supplementary Material).

In patients not receiving norepinephrine, error grid analysis showed that the proportions of measurements in risk zones A to E were 88.5%, 9.1%, 2.1%, 0.3%, and 0% for SAP and 81.5%, 18.0%, 0.3%, 0%, and 0.3% for MAP (Supplementary Fig 5 in Supplementary Material).

Patients receiving norepinephrine had more paired measurements in risk zones B to E compared to patients not receiving norepinephrine ($p < 0.001$ for SAP, and $p = 0.03$ for MAP).

Table 1. Patient characteristics.

Variable	Type	SICS-I cohort N = 1075	Without paired measurement N = 339	With paired measurement N = 736	P value
Age (years)		62 (15)	62 (14)	62 (15)	0.30
Male gender		674 (63%)	206 (61%)	468 (64%)	0.37
Body Mass Index (kg/m^2)		26.9 (5.5)	27.4 (6.2)	26.7 (5.1)	0.038
Mechanical ventilation		632 (59%)	195 (58%)	437 (59%)	0.57
Sedation		430 (40%)	134 (40%)	296 (40%)	0.83
Heart rate (bpm)		87.7 (21.2)	86.8 (21.8)	88.1 (21.0)	0.35
Atrial fibrillation		78 (7%)	24 (7%)	54 (7%)	0.88
Norepinephrine use		529 (49%)	177 (52%)	352 (48%)	0.18
Norepinephrine dose (μg/kg/min)		0.13 [0.06, 0.27]	0.12 [0.07, 0.21]	0.14 [0.06, 0.29]	0.38
Admission type	Medical	713 (66%)	212 (63%)	501 (68%)	0.16
	Acute surgery	316 (29%)	113 (33%)	203 (28%)	
	Planned surgery	46 (4%)	14 (4%)	32 (4%)	
Admission diagnosis by organ system	Cardiovascular	318 (30%)	118 (35%)	200 (27%)	0.002
	Gastrointestinal	167 (16%)	49 (14%)	118 (16%)	
	Genito-urinary	23 (2%)	11 (3%)	12 (2%)	
	Hematological	19 (2%)	8 (2%)	11 (1%)	
	Metabolic	22 (2%)	3 (1%)	19 (3%)	
	Musculoskeletal/skin	13 (1%)	5 (1%)	8 (1%)	
	Neurological	143 (13%)	31 (9%)	112 (15%)	
	Respiratory	229 (21%)	60 (18%)	169 (23%)	
	Transplant	58 (5%)	19 (6%)	39 (5%)	
	Trauma	82 (8%)	35 (10%)	47 (6%)	
Time to inclusion (hours)		15 [8, 20]	15 [8, 19]	15 [8, 20]	0.39
Circulatory shock	Total[+]	540 (50%)	183 (54%)	357 (49%)	0.095
	Cardiogenic	140 (13%)	52 (15%)	88 (12%)	
	Distributive	327 (30%)	107 (32%)	220 (30%)	
	Hypovolemic	120 (11%)	48 (14%)	72 (10%)	
	Obstructive	25 (2%)	12 (4%)	13 (2%)	
APACHE-IV score		76.1 (29.3)	79.1 (31.0)	74.7 (28.4)	0.031
SAPS-II		46.4 (16.8)	47.3 (16.8)	46.0 (16.8)	0.22
90-day mortality		297 (28%)	108 (32%)	189 (26%)	0.035

[+]*Multiple types of circulatory shock may be diagnosed in a patient. Abbreviations: SICS-I, Simple Intensive Care Studies-I; APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score-II.*

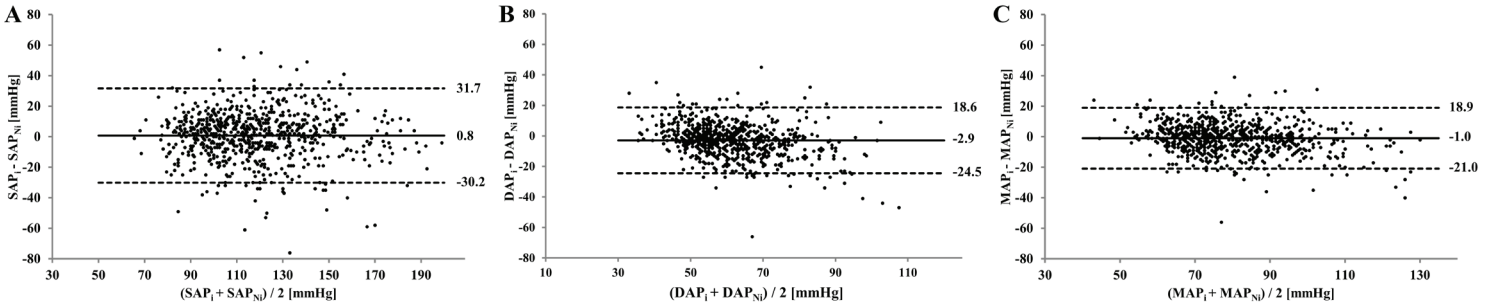


Figure 2.

A. Bland-Altman plot for systolic arterial pressure. Comparison between measurements of non-invasive oscillometric cuff (SAP_{NI}) and invasive arterial catheter (SAP_I) is illustrated.

B. Bland-Altman plot for diastolic arterial pressure. Comparison between measurements of non-invasive oscillometric cuff (DAP_{NI}) and invasive arterial catheter (DAP_I) is illustrated.

C. Bland-Altman plot for mean arterial pressure. Comparison between measurements of non-invasive oscillometric cuff (MAP_{NI}) and invasive arterial catheter (MAP_I) is illustrated. In each plot, the continuous horizontal line represents the mean difference, and the upper and lower dashed lines represent the 95% limits of agreement.

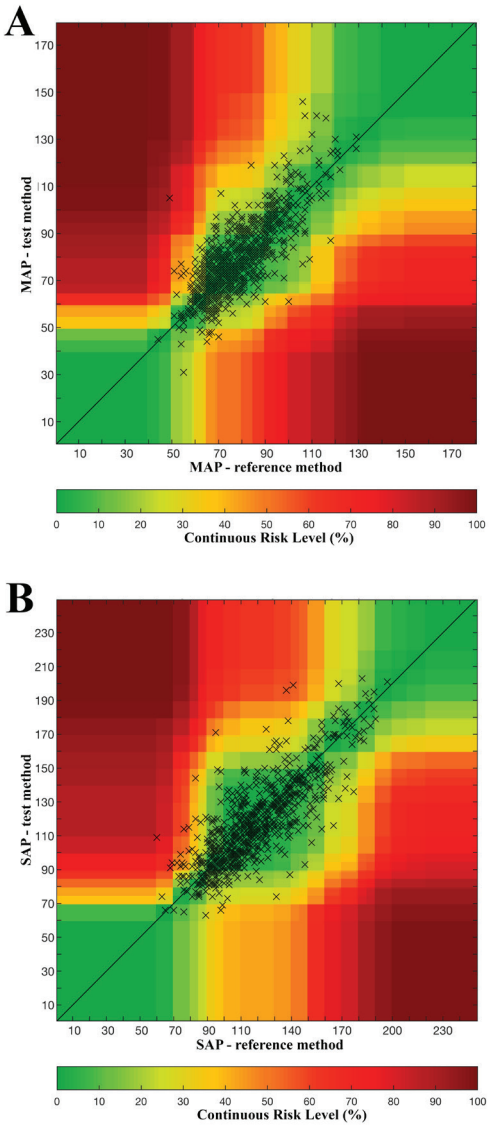


Figure 3.

A. Error-grid plot for mean arterial pressure. This figure illustrates the error grid for the test method (upper-arm cuff oscillometry) in comparison with the reference method (arterial catheter) for the 736 patients regarding mean arterial pressure.

B. Error-grid plot for systolic arterial pressure. This figure illustrates the error grid for the test method (upper-arm cuff oscillometry) in comparison with the reference method (arterial catheter) for the 736 patients regarding systolic arterial pressure. In each plot, the background colors correspond to the risk level for each pair of measurements. The risk ranges from 0% to 100%, as shown below the plots. A risk level between 0% and 100% is assigned to each combination of measurement and true value (test device versus gold standard).

Table 2. Arterial pressure values determined by non-invasive oscillometric cuff and arterial catheter

Variable	Invasive arterial catheter (mmHg)	Non-invasive oscillometric cuff (mmHg)	Mean difference (CI)	SD of the Mean difference (mmHg)	Lower 95% Limit of Agreement (CI)	Upper 95% Limit of Agreement (CI)	Percentage error (CI)
All patients (N = 736)							
Systolic arterial pressure	118.8 ± 24.8	117.9 ± 25.2	0.8 mmHg (-0.4 to 1.9)	± 15.7	-30.2 mmHg (-32.1 to -28.2)	31.7 mmHg (29.7 to 33.6)	26.1% (24.5 to 27.8)
Diastolic arterial pressure	59.9 ± 11.4	62.8 ± 14.6	-2.9 mmHg (-3.7 to -2.1)	± 11.0	-24.5 mmHg (-25.9 to -23.1)	18.6 mmHg (17.2 to 20.0)	35.1 % (32.9 to 37.4)
Mean arterial pressure	78.7 ± 14.3	79.7 ± 16.5	-1.0 mmHg (-1.7 to -0.3)	± 10.2	-21.0 mmHg (-22.2 to -19.7)	18.9 mmHg (17.7 to 20.2)	25.2% (23.6 to 26.8)
Patients on norepinephrine (N = 352)							
Systolic arterial pressure	107.8 ± 20.7	108.9 ± 21.3	-1.1 mmHg (-2.8 to 0.6)	± 16.1	-32.8 mmHg (-35.7 to -29.9)	30.6 mmHg (27.7 to 33.5)	29.3% (26.6 to 32.0)
Diastolic arterial pressure	56.5 ± 9.4	58.4 ± 11.9	-1.9 mmHg (-2.9 to -0.9)	± 10.0	-21.5 mmHg (-23.3 to -19.7)	17.7 mmHg (15.9 to 19.5)	34.1% (31.0 to 37.2)
Mean arterial pressure	72.7 ± 11.1	73.7 ± 12.9	-1.0 mmHg (-2.0 to 0.1)	± 9.9	-20.4 mmHg (-22.2 to -18.6)	18.5 mmHg (16.7 to 20.3)	26.6% (24.2 to 29.0)
Patients not on norepinephrine (N = 384)							
Systolic arterial pressure	128.6 ± 24.2	126.1 ± 25.8	2.4 mmHg (0.9 to 4.0)	± 15.2	-27.5 mmHg (-30.1 to -24.8)	32.4 mmHg (29.7 to 35.0)	23.5% (21.4 to 25.6)
Diastolic arterial pressure	63.0 ± 12.2	66.9 ± 15.7	-3.9 mmHg (-5.1 to -2.7)	± 11.8	-27.0 mmHg (-29.1 to -25.0)	19.3 mmHg (17.2 to 21.3)	35.6% (32.5 to 38.8)
Mean arterial pressure	84.2 ±14.6	85.3 ± 17.5	-1.1 mmHg (-2.1 to 0.0)	± 10.4	-21.5 mmHg (-23.3 to -19.7)	19.4 mmHg (17.6 to 21.2)	24.1% (22.0 to 26.2)

Discussion

In this study, non-invasive blood pressure measurements using upper-arm cuff oscillometry showed a low mean difference but large limits of agreement compared to direct invasive measurements in critically ill patients. Precision and accuracy of non-invasive SAP, DAP, and MAP measurements determined by Bland-Altman analysis failed the AAMI standards for non-invasive arterial pressure measurement (12).

Estimations of precision and accuracy obtained by Bland-Altman analysis provide information about the overall statistical agreement but do not offer the clinical relevance of differences between paired measurements. Error grid analysis enables quantification and illustration of the clinical significance of observed differences between two blood pressure measurement methods (15). This novel method is based on expert opinion and has yet to be externally validated using outcome data. In this study, error grid analysis showed that for all included patients, the majority of measurements were situated in the no-risk zone A, i.e., 82% for SAP and 78% for MAP. However, in approximately one in five patients, i.e., 18% for SAP and 22% for MAP, the paired measurements were positioned in risk zones B to E, which implies a potential risk of at least low-risk treatment decisions if treatment was based on the test method. In patients receiving norepinephrine, there were more paired measurements, 24% for SAP and 25% for MAP, in the risk zones B to E compared to patients not receiving norepinephrine, which implies a higher potential risk of at least low-risk treatment differences. These results suggest that in approximately one in four patients receiving norepinephrine in this cohort, the measured difference could potentially have triggered at least low-risk treatment decisions if treatment was based on the test method. The distribution difference in the error grid risk zones for patients receiving norepinephrine may be explained by the reduced performance of the oscillometric blood pressure method with lower blood pressures (17).

Data used in this study were obtained as part of the SICS-I prospective observational cohort study, which evaluated clinical and hemodynamic variables obtained as a one-time clinical examination within the first 24 hours of ICU admission to diagnose a low cardiac index (11) and to build a prognostic model for 90-day mortality (18). Not all variables could be obtained in case of life-threatening disease, which could explain why patients without a paired measurement of blood pressure had a higher APACHE-IV score and increased mortality. Furthermore, the adequate size oscillometric cuff was not always directly available when study procedures were performed, which could explain why patients with a higher BMI had fewer paired measurements.

Our results on measurement agreement are in line with other studies comparing non-invasive oscillometric and invasive arterial blood pressure in critically ill patients. A poor statistical agreement was found in two retrospective observational studies of large ICU databases (17,19). Similar to our study, MAP measurements were less inaccurate than SAP and DAP measurements.

Subgroup analysis according to the administration of norepinephrine showed similar results in both subgroups. Results we observed in patients treated with norepinephrine are comparable to results from previous studies. One study compared non-invasive oscillometric and invasive radial arterial blood pressure measurements in critically ill patients receiving norepinephrine (9). A comparison between measurements of MAP showed a mean difference of 6.6 mmHg (95% CI 5.3 to 7.9) (9) in this study. In another study, non-invasive and invasive blood pressure measurements were prospectively compared in adult patients with septic shock in the ICU (10). Similar to our study, non-invasive blood pressure monitoring showed poor statistical agreement with invasive measurements (10).

Currently, the invasive blood pressure measurement using an arterial catheter remains the clinical reference method in critically ill patients. Despite the widespread use of arterial catheters, there is no evidence suggesting that the outcomes of critically ill patients improve with continuous invasive compared to intermittent non-invasive blood pressure monitoring. Arterial catheters are associated with rare but serious complications such as infections, bleeding, thrombosis, and pseudoaneurysm formation (20), and placement of the catheters may be difficult and time-consuming in some patients. One retrospective cohort study showed an association between the use of arterial catheters and an increase in mortality in critically ill patients receiving vasopressors for shock (21). Furthermore, invasive arterial pressure monitoring can be inaccurate because of underdamping and resonance phenomena (22). Finally, a different blood pressure measurement may be obtained depending on the artery the catheter is placed in. Some evidence suggests that measurements obtained in the radial artery underestimates central blood pressure in septic shock patients receiving vasopressors and that femoral arterial pressure monitoring may be more appropriate in these patients (23).

Alternatively, frequent oscillometric cuff inflation is associated with patient discomfort and pressure bruises. Patient discomfort is increased if repeated percutaneous vascular punctures are needed for laboratory testing. Patient movement may also influence the performance of the oscillometric blood pressure method (24). In addition, clinically relevant hypotensive episodes might be missed or detected late by intermittent measurements. Multiple manufacturers have developed appropriate devices, and each makes use of proprietary algorithms that have not always been reliably validated against invasive direct blood pressure measurements with an arterial catheter (25). Although a structured method comparison study may better reveal the agreement between these two methods of blood pressure measurement, these data obtained as part of a prospective study may reflect clinical practice and therefore our results potentially have better generalizability. It is recommended to be cautious in clinical practice on the interchangeability of blood pressure measurement methods as the presented analyses do not provide information on individual measurement differences and preferences may vary for individual patients.

There are several limitations to this study. First, this study was a post hoc analysis of a single-center prospective observational study as this research question was not specified a priori (11). Due to the study design there was no formal power calculation and the results are exploratory.

These findings have to be used with caution and have to be validated in other cohorts. Second, not all patients included in the cohort had a paired measurement of blood pressure. Acquiring an oscillometric blood pressure measurement was not a primary aim of the SICS-I study, and the measurement was not always performed, most often if this would have led to interference with clinical care. Third, we only performed a single paired measurement during the first 24 hours of patient admission. Fourth, we did not compare the non-invasive blood pressure measurements on both arms before performing a measurement.

Conclusions

Non-invasive blood pressure measurements using upper-arm cuff oscillometry showed a low mean difference but large limits of agreement compared to direct invasive measurements in critically ill patients. Error grid analysis showed that measurement differences between oscillometry and the arterial catheter would potentially have triggered at least low-risk treatment decisions in one in five critically ill patients, and in one in four patients on norepinephrine if treatment was based on oscillometry. It is recommended to be cautious in clinical practice on the interchangeability of blood pressure measurement methods as these analyses do not provide information on individual measurement differences and preferences may vary for individual patients.

References

1 Saugel B, Dueck R, Wagner JY. Measurement of blood pressure. Best Pract Res Clin Anaesthesiol 2014;28:309–22. doi:10.1016/j.bpa.2014.08.001.

2 Cecconi M, De Backer D, Antonelli M, Beale R, Bakker J, Hofer C, et al. Consensus on circulatory shock and hemodynamic monitoring. Task force of the European Society of Intensive Care Medicine. Intensive Care Med 2014;40:1795–815. doi:10.1007/s00134-014-3525-z.

3 Sakr Y, Reinhart K, Vincent J-L, Sprung CL, Moreno R, Ranieri VM, et al. Does dopamine administration in shock influence outcome? Results of the Sepsis Occurrence in Acutely Ill Patients (SOAP) Study. Crit Care Med 2006;34:589–97. doi:10.1097/01.CCM.0000201896.45809.E3.

4 Vincent J-L, De Backer D. Circulatory Shock. N Engl J Med 2013;369:1726–34. doi:10.1056/NEJMra1208943.

5 Chatterjee A, Depriest K, Blair R, Bowton D, Chin R. Results of a survey of blood pressure monitoring by intensivists in critically ill patients: A preliminary study. Crit Care Med 2010;38:2335–8. doi:10.1097/CCM.0b013e3181fa057f.

6 Lakhal K, MacQ C, Ehrmann S, Boulain T, Capdevila X. Noninvasive monitoring of blood pressure in the critically ill: Reliability according to the cuff site (arm, thigh, or ankle). Crit Care Med 2012;40:1207–13. doi:10.1097/CCM.0b013e31823dae42.

7 Lakhal K, Ehrmann S, Martin M, Faiz S, Rignminiac F, Cinotti R, et al. Blood pressure monitoring during arrhythmia: Agreement between automated brachial cuff and intra-arterial measurements. Br J Anaesth 2015;115:540–9. doi:10.1093/bja/aev304.

8 Lakhal K, Ehrmann S, Boulain T. Noninvasive BP Monitoring in the Critically Ill: Time to Abandon the Arterial Catheter? Chest 2018;153:1023–39. doi:10.1016/j.chest.2017.10.030.

9 Pytte M, Dybwik K, Sexton J, Straume B, Waage Nielsen E. Oscillometric brachial mean artery pressures are higher than intra-radial mean artery pressures in intensive care unit patients receiving norepinephrine. Acta Anaesthesiol Scand 2006;50:718–21. doi:10.1111/j.1399-6576.2006.01045.x.

10 Riley LE, Chen GJ, Latham HE. Comparison of noninvasive blood pressure monitoring with invasive arterial pressure monitoring in medical ICU patients with septic shock. Blood Press Monit 2017;22:202–7. doi:10.1097/MBP.0000000000000258.

11 Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, et al. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med 2019;45:190–200. doi:10.1007/s00134-019-05527-y.

12 American National Standard/Association for the Advancement of Medical Instrumentation. ANSI/AAMI SP10: 2002. Manual Electronic, or Automated Sphygmomanometers. Arlington, VA: Association for the Advancement of Medical Instrumentation; 2003. n.d.

13 Hiemstra B, Eck RJ, Koster G, Wetterslev J, Perner A, Pettilä V, et al. Clinical examination, critical care ultrasonography and outcomes in the critically ill: cohort profile of the Simple Intensive Care Studies-I. BMJ Open 2017;7:e017170. doi:10.1136/bmjopen-2017-017170.

14 Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60. doi:10.1177/096228029900800204.

15 Saugel B, Grothe O, Nicklas JY. Error grid analysis for arterial pressure method comparison studies. Anesth Analg 2018;126:1177–85. doi:10.1213/ANE.0000000000002585.

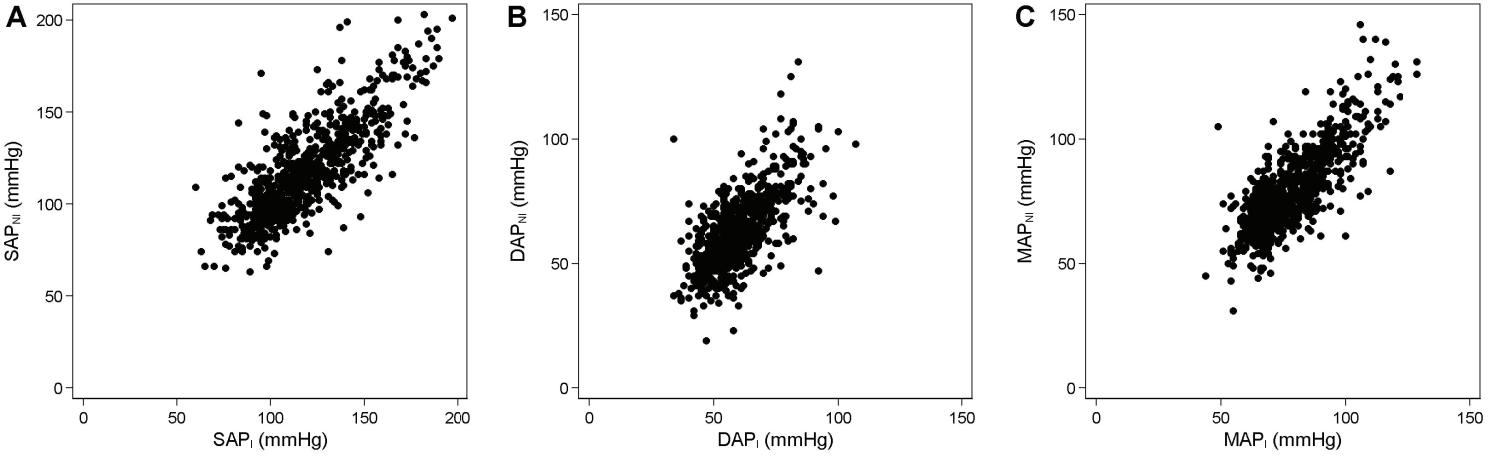
16 Grothe O, Kaplan A, Kouz K, Saugel B. Computer Program for Error Grid Analysis in Arterial Blood Pressure Method Comparison Studies. Anesth Analg 2019:1. doi:10.1213/ANE.0000000000004584.

17 Wax DB, Lin H-M, Leibowitz AB. Invasive and Concomitant Noninvasive Intraoperative Blood Pressure Monitoring. Anesthesiology 2011;115:973–8. doi:10.1097/aln.0b013e3182330286.

18 Hiemstra B, Eck RJ, Wiersema R, Kaufmann T, Koster G, Scheeren TWL, et al. Clinical Examination for the Prediction of Mortality in the Critically Ill. Crit Care Med 2019;47:1301–9. doi:10.1097/CCM.0000000000003897.
19 Lehman LWH, Saeed M, Talmor D, Mark R, Malhotra A. Methods of blood pressure measurement in the ICU. Crit Care Med 2013;41:34–40. doi:10.1097/CCM.0b013e318265ea46.
20 Scheer B, Perel A, Pfeiffer UJ. Clinical review: complications and risk factors of peripheral arterial catheters used for haemodynamic monitoring in anaesthesia and intensive care medicine. Crit Care 2002;6:199–204. doi:10.1186/cc1489.
21 Gershengorn HB, Wunsch H, Scales DC, Zarychanski R, Rubenfeld G, Garland A. Association between arterial catheter use and hospital mortality in intensive care units. JAMA Intern Med 2014;174:1746–54. doi:10.1001/jamainternmed.2014.3297.
22 Romagnoli S, Ricci Z, Quattrone D, Tofani L, Tujjar O, Villa G, et al. Accuracy of invasive arterial pressure monitoring in cardiovascular patients: an observational study. Crit Care 2014;18:644. doi:10.1186/s13054-014-0644-4.
23 Kim WY, Jun JH, Huh JW, Hong SB, Lim CM, Koh Y. Radial to femoral arterial blood pressure differences in septic shock patients receiving high-dose norepinephrine therapy. Shock 2013;40:527–31. doi:10.1097/SHK.0000000000000064.
24 Alpert BS, Quinn D, Gallick D. Oscillometric blood pressure: A review for clinicians. J Am Soc Hypertens 2014;8:930–8. doi:10.1016/j.jash.2014.08.014.
25 Smulyan H, Safar ME. Blood pressure measurement: Retrospective and prospective views. Am J Hypertens 2011;24:628–34. doi:10.1038/ajh.2011.22.

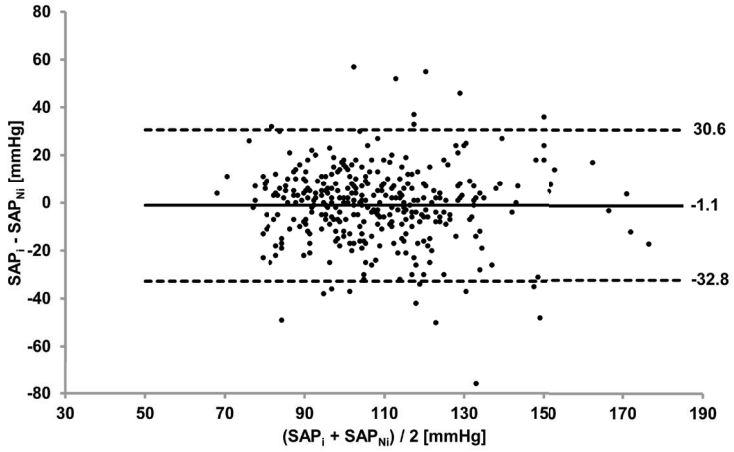
Supplements

eFigure 1. *Scatter plots comparing invasive to non-invasive blood pressure measurements*

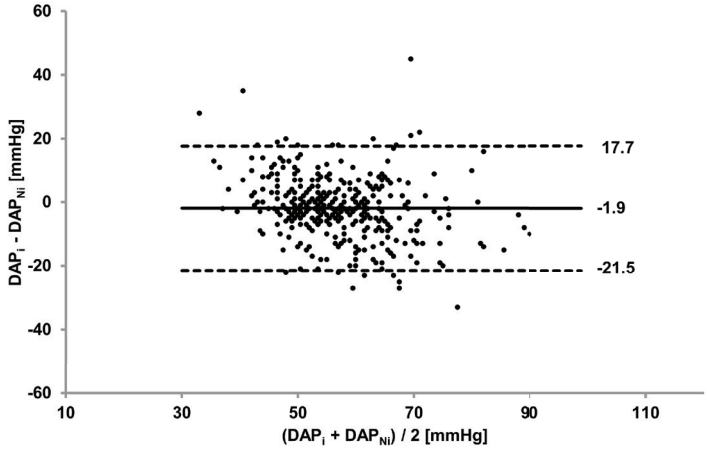


A. Scatter plot for systolic arterial pressure. Correlation between systolic arterial pressure data obtained with non-invasive oscillometric cuff (SAP_{NI}) and invasive arterial catheter (SAP_{I}) is illustrated. B. Scatter plot for diastolic arterial pressure. Correlation between diastolic arterial pressure data obtained with non-invasive oscillometric cuff (DAP_{NI}) and invasive arterial catheter (DAP_{I}) is illustrated. C. Scatter plot for mean arterial pressure. Correlation between mean arterial pressure data obtained with non-invasive oscillometric cuff (MAP_{NI}) and invasive arterial catheter (MAP_{I}) is illustrated.

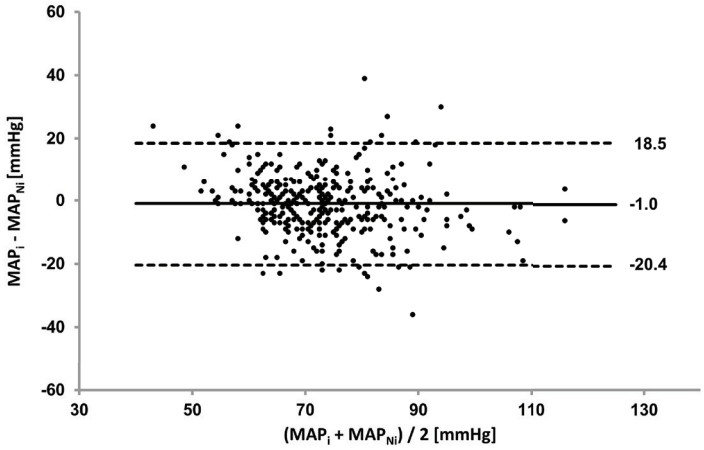
eFigure 2. *Bland-Altman plots for patients receiving norepinephrine (n=352)*



A. Bland-Altman plot for systolic arterial pressure for patients receiving norepinephrine. Comparison between measurements of non-invasive oscillometric cuff (SAP_{NI}) and invasive arterial catheter (SAP_I) is illustrated.



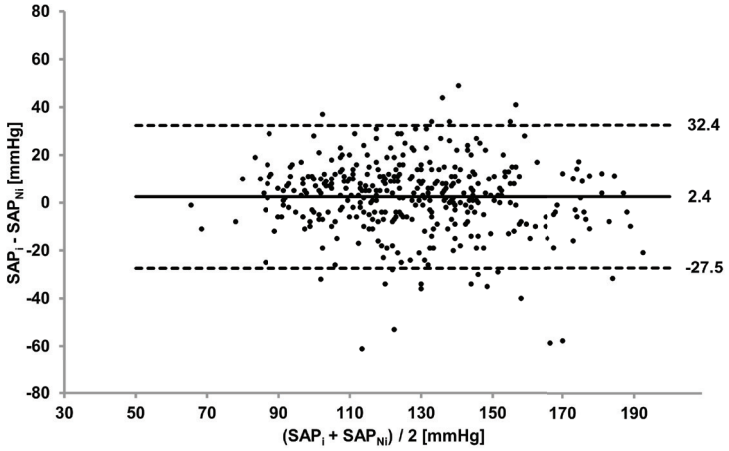
B. Bland-Altman plot for diastolic arterial pressure for patients receiving norepinephrine. Comparison between measurements of non-invasive oscillometric cuff (DAP_{NI}) and invasive arterial catheter (DAP_I) is illustrated.



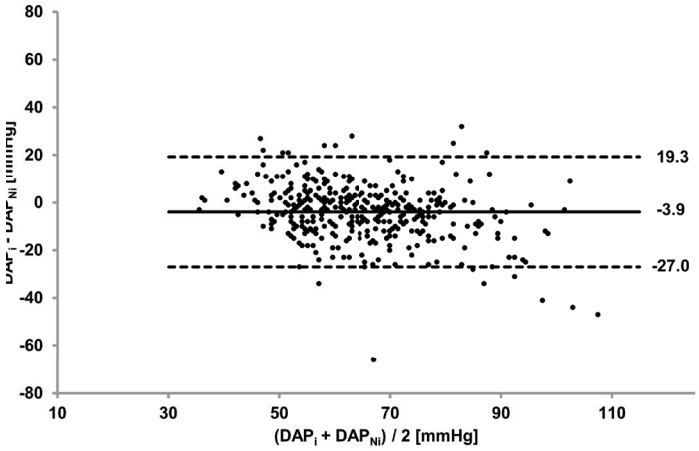
C. Bland-Altman plot for mean arterial pressure for patients receiving norepinephrine. Comparison between measurements of non-invasive oscillometric cuff (MAP_{NI}) and invasive arterial catheter (MAP_I) is illustrated.

In each plot, the continuous horizontal line represents the mean difference, and the upper and lower dashed lines represent the 95% limits of agreement.

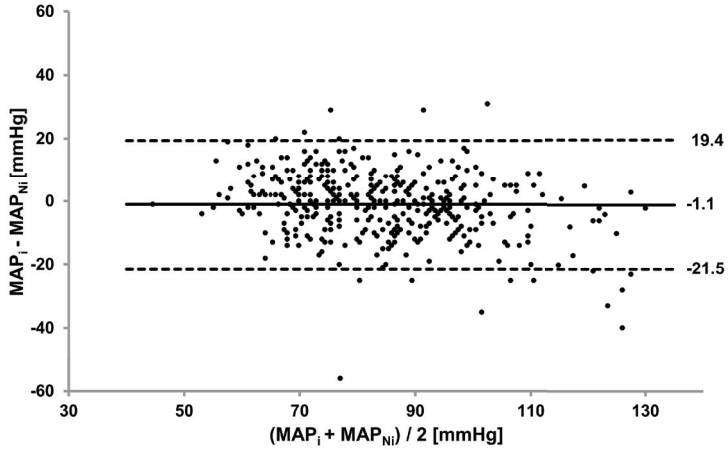
eFigure 3. *Bland-Altman plots for patients not receiving norepinephrine (n=384)*



A. Bland-Altman plot for systolic arterial pressure for patients not on norepinephrine. Comparison between measurements of non-invasive oscillometric cuff (SAP_{NI}) and invasive arterial catheter (SAP_I) is illustrated.



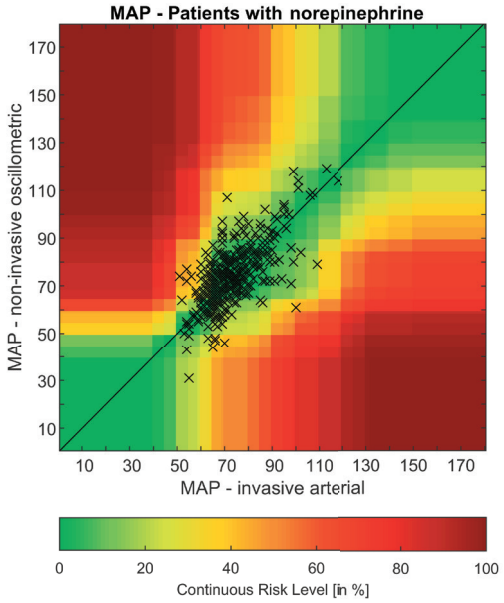
B. Bland-Altman plot for diastolic arterial pressure for patients not on norepinephrine. Comparison between measurements of non-invasive oscillometric cuff (DAP_{NI}) and invasive arterial catheter (DAP_I) is illustrated.



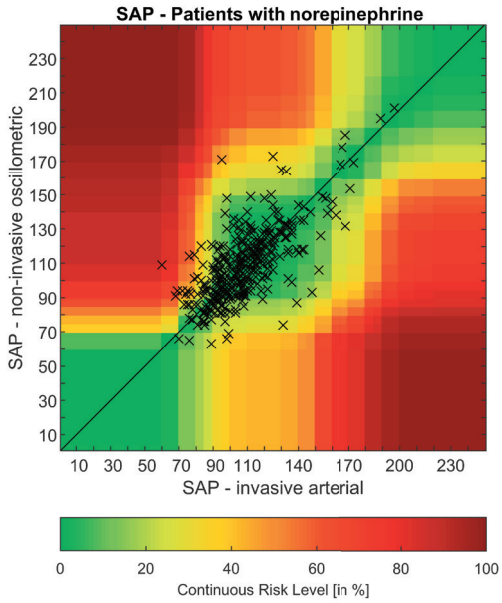
C. Bland-Altman plot for mean arterial pressure for patients not on norepinephrine. Comparison between measurements of non-invasive oscillometric cuff (MAP_{NI}) and invasive arterial catheter (MAP_I) is illustrated.

In each plot, the continuous horizontal line represents the mean difference, and the upper and lower dashed lines represent the 95% limits of agreement.

eFigure 4. *Error grid analyses for patients receiving norepinephrine (n=352)*



A. Error-grid analysis for mean arterial pressure. This figure illustrates the error grid for the test method (upper-arm cuff oscillometry) in comparison with the reference method (arterial catheter) for patients receiving norepinephrine regarding mean arterial pressure.

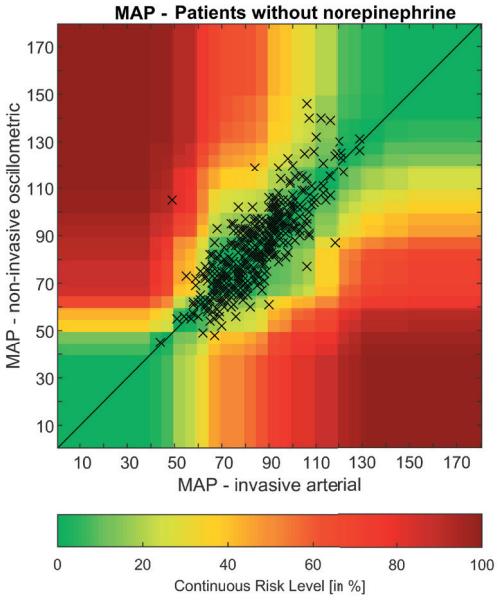


B. Error-grid analysis for systolic arterial pressure. This figure illustrates the error grid for the test method (upper-arm cuff oscillometry) in comparison with the reference method (arterial catheter) for patients receiving norepinephrine regarding systolic arterial pressure.

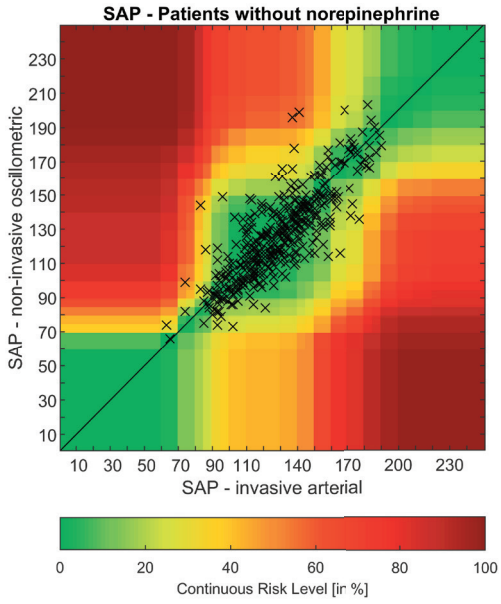
In each plot, the background colors correspond to the risk level for each pair of measurement. The risk ranges from 0% to 100%, as shown below the plots. A risk level between 0% and 100% is assigned to each combination of measurement and true value (test device versus gold standard).

Error grid analysis showed that the proportions of measurements in risk zones A to E were 75.6%, 17.6%, 6.5%, 0.3%, and 0% for systolic arterial pressure and 74.7%, 23.6%, 1.7%, 0%, and 0% for mean arterial pressure.

eFigure 5. *Error grid analyses for patients not receiving norepinephrine (n=384)*



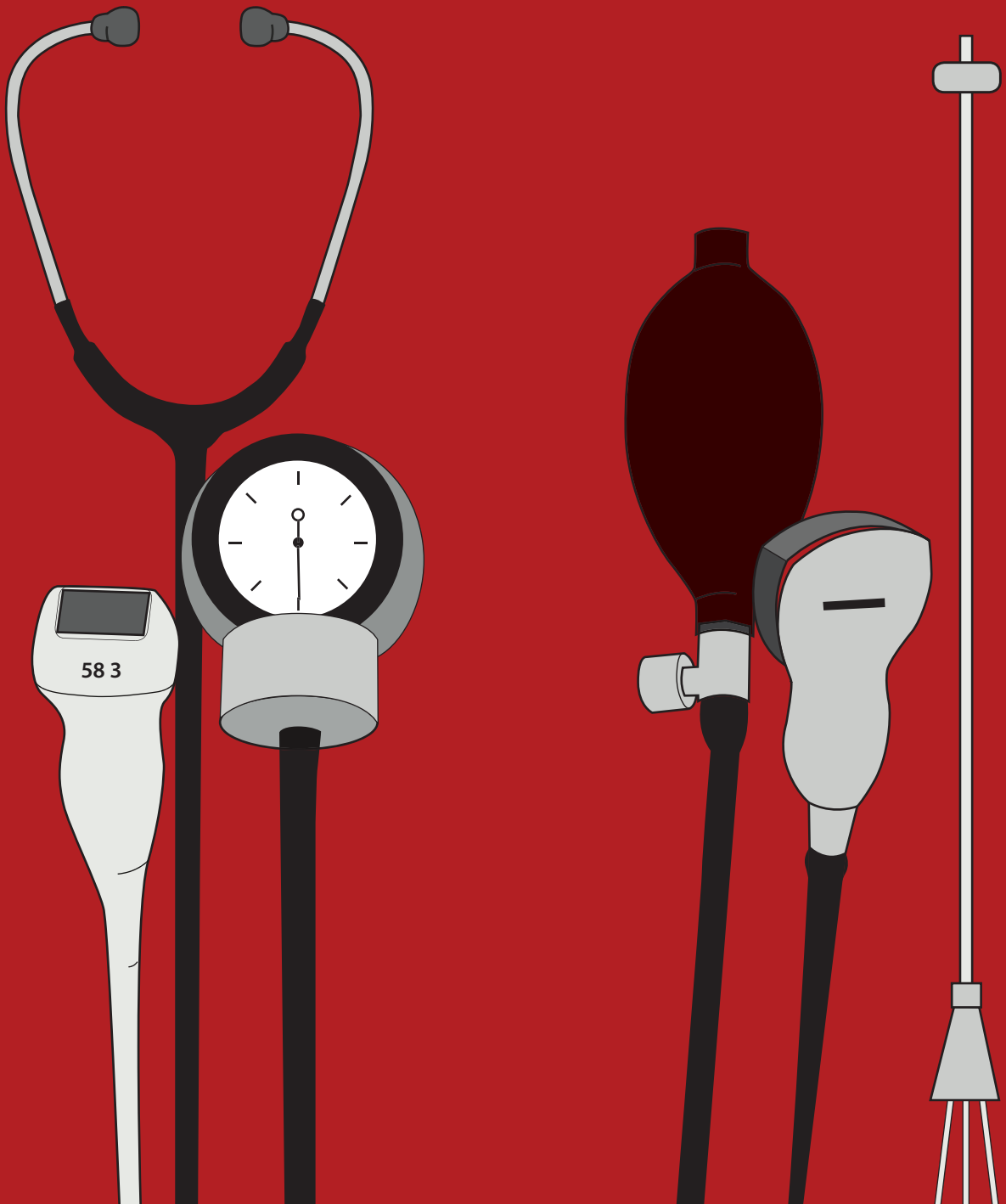
A. Error-grid analysis for mean arterial pressure. This figure illustrates the error grid for the test method (upper-arm cuff oscillometry) in comparison with the reference method (arterial catheter) for patients not receiving norepinephrine regarding mean arterial pressure.



B. Error-grid analysis for systolic arterial pressure. This figure illustrates the error grid for the test method (upper-arm cuff oscillometry) in comparison with the reference method (arterial catheter) for patients not receiving norepinephrine regarding systolic arterial pressure.

In each plot, the background colors correspond to the risk level for each pair of measurement. The risk ranges from 0% to 100%, as shown below the plots. A risk level between 0% and 100% is assigned to each combination of measurement and true value (test device versus gold standard).

Error grid analysis showed that the proportions of measurements in risk zones A to E were 88.5%, 9.1%, 2.1%, 0.3%, and 0% for systolic arterial pressure and 81.5%, 18.0%, 0.3%, 0%, and 0.3% for mean arterial pressure.


58 3

6

Feasibility of cardiac output measurements in critically ill patients by medical students

Geert Koster
Thomas Kaufmann
Bart Hiemstra
Renske Wiersema
Madelon E. Vos
Devon Dijkhuizen
Adrian Wong
Thomas W.L. Scheeren
Yoran M. Hummel
Frederik Keus
Iwan C.C. van der Horst
SICS Study Group

The Ultrasound Journal

Abstract

Background
Critical care ultrasonography (CCUS) is increasingly applied also in the intensive care unit (ICU) and performed by non-experts, including even medical students. There is limited data on the training efforts necessary for novices to attain images of sufficient quality. There is no data on medical students performing CCUS for the measurement of cardiac output (CO), a hemodynamic variable of importance for daily critical care.

Objective
The aim of this study was to explore the agreement of cardiac output measurements as well as the quality of images obtained by medical students in critically ill patients compared to the measurements obtained by experts in these images.

Methods
In a prospective observational cohort study, all acutely admitted adults with an expected ICU stay over 24 hours were included. CCUS was performed by students within 24 hours of admission. CCUS included the images required to measure the CO, i.e., the left ventricular outflow tract (LVOT) diameter and the velocity time integral (VTI) in the LVOT. Echocardiography experts were involved in the evaluation of the quality of images obtained and the quality of the CO measurements.

Results
There was an opportunity for a CCUS attempt in 1155 of the 1212 eligible patients (95%) and 1075 of the 1212 patients (89%) CCUS examination was performed by medical students. In 871 out of 1075 patients (81%) medical students measured CO. Experts measured CO in 783 patients (73%). In 760 patients (71%) CO was measured by both which allowed for comparison; bias of CO was 0.0 L/min with limits of agreement of -2.6 L/min to 2.7 L/min. The percentage error was 50%, reflecting poor agreement of the CO measurement by students compared with the experts CO measurement.

Conclusions
Medical students seem capable of obtaining sufficient quality CCUS images for CO measurement in the majority of critically ill patients. Measurements of CO by medical students, however, had poor agreement with expert measurements. Experts remain indispensable for reliable CO measurements.

Background

Critical care ultrasonography (CCUS) is a deliberately focused examination, aimed at rapidly answering straightforward clinical questions (1). In the field of emergency and critical care medicine, CCUS is increasingly used to guide interventions in critically ill patients in various settings by experts and novices (2–14). The training process required for users to attain competency in CCUS has varied widely between studies, reflecting the diversity in CCUS training between centers. Similarly, there is variability among statements from stakeholders regarding the type of training, the required number of hours spent and examinations performed by the trainee to achieve competency in CCUS (15–17). However, besides these disparities, individual physicians struggle with barriers to its use, such as perceived difficulty in obtaining adequate technical skills (13), limitations in training, need (perceived and real), and costs (6, 14).

One valuable CCUS hemodynamic measurement is the determination of the cardiac output (CO), especially if the patient is in circulatory shock (18). Circulatory shock occurs in one-third of patients admitted to the ICU (19), so being able to perform CCUS and measure CO is of importance. However, CO measurement by CCUS is considered an advanced level CCUS skill (20, 21). Whether trained novices (e.g., medical students or other less experienced physicians) are able to obtain reliable CO measurements has not yet been investigated. In a convenience sample of 100 adult patients in the emergency department (ED), two ultrasound-naive ED physicians were able to measure CO by ultrasonography accurately (22). Another study in the ED with a convenience sample of 80 patients, however, showed poor agreement in CO measurement by an emergency ultrasound fellow compared to an emergency cardiology fellow (23). At the start of our study there are were no data on medical students performing CO measurements by CCUS in critically ill patients, although medical students have been shown to be capable of performing CCUS after limited training (24). To our knowledge, only one small study investigated CCUS by medical students on a (cardiac) ICU, and CO was not measured (see supplements) (3).

The aim of this study was to explore the feasibility of a limited CCUS examination, consisting of CO measurements, performed by medical students in a protocolized manner, in critically ill patients. In addition, the quality of images required to calculate CO and the accuracy of CO measurements compared to those obtained by echocardiography experts were analyzed.

Methods

The Simple Intensive Care Studies (SICS)-I was a prospective, observational cohort study which followed a published protocol and statistical analysis plan (Clinicaltrials.gov; NCT02912624). The SICS-I was developed to unravel the diagnostic and prognostic value of a comprehensive selection of clinical, hemodynamic, and biochemical variables in critically ill patients, and details have been described elsewhere (25, 26). All acutely admitted adults with an expected ICU stay over 24 hours were included. Patients were excluded when admission was planned and if clinical care interfered

with acquiring research data (e.g., mechanical circulatory support). The local institutional review board approved the study (M15.168207).

Data collection and training

All patients underwent CCUS within 24 hours of ICU admission. Detailed information on the CCUS performed can be found in the supplements. Patients were enrolled by fourth-year to sixth-year medical students of a six-year medical school program. The training consisted of self-study on theoretical fundamentals and two practical sessions of at least two hours in total to learn how to operate the General Electric Vivid-S6 mobile ultrasonography machine using the cardiac phased-array probe (see appendix in supplements for detailed information). The theoretical self-study on how to perform CCUS and measure the CO consisted of study of the protocol (supplements), a website on the principles of echocardiography (27), and international guidelines (28, 29). This information became available two weeks before participation of the medical students. During the practical sessions, medical students learned to obtain the parasternal long axis (PLAX), apical four-chamber (AP4CH), and apical five-chamber (AP5CH) views, among others. The medical students alternated with obtaining the views and measurements of CO during the practical sessions. All medical students received at least two hours hands-on training from cardiologist-intensivists (GK and IVDH).

Views and images were obtained randomly during the respiratory cycle and/or phase of mechanical ventilation. In case of any arrhythmias, the average of multiple measurements over five heartbeats was taken.

The first 20 CCUS images and measurements of each medical student were supervised by medical students who had independently performed more than 50 CCUS examinations. After 20 scans, CCUS medical students were allowed to perform CCUS unsupervised, since previous studies showed acceptable capability for acquiring images beyond 20 exams (30).

Validation and definitions

For quality control, echocardiography technicians from an independent core laboratory (Groningen Image Core Lab, UMCG, Groningen, the Netherlands) assessed all CCUS images and measurements obtained by the medical students according to the study protocol. If the images were obtained according to guideline standards, the LVOTd and VTI were independently remeasured and CO recalculated (28, 29). Core laboratory technicians, which we refer to as experts throughout this report, were blinded to all other clinical measurements. The experts did not perform any CCUS examination.

Outcomes, index test and reference standard

The number of patients where CCUS could not be performed and reasons for unobtainable images by the medical students were reported. Patients were excluded from the analysis if, for research purposes, experts would also not be able to perform CCUS (i.e., drains, subcutaneous emphysema, surgical dressing/wounds). The number of patients in which CCUS images of PLAX

or AP5CH were obtained was analyzed (28, 29). Proportion of patients was reported wherein the CCUS images assessed by the experts was of insufficient quality for CO measurement.

We also evaluated the accuracy of CO measurements by medical students ($CO_{student}$) compared to CO measurements by experts (CO_{expert}). Moreover, the two components needed for CO calculation (i.e., LVOTd or VTI) were assessed to determine possible differences between medical students' and experts' measurements.

Sensitivity analyses were done with baseline characteristics to investigate reasons why experts could not measure a CO.

Sample size and missing data

Due to the observational nature of this study, no formal power calculation was performed. For the accuracy analysis on CO measurements, we only included patients if CO was measured by both medical students and experts.

Statistical analyis

Data were presented as mean with standard deviation (SD) when normally distributed or as median with interquartile ranges (IQR) in case of skewed data. Dichotomous and categorical data were presented in proportions. Intraclass correlation coefficients (ICC) were calculated to assess the concordance between the measurements made by the medical students and the experts. Bland-Altman analysis was performed to assess agreement of medical student versus expert measurements by calculating mean and SD of the differences, the 95% limits of agreement (LOA) (= mean of the difference ± 1.96 × SD of the difference), and the percentage error (31). In method comparison studies, a percentage error of 30% is considered acceptable if the error of the test and the reference method is 20%, which is the case when using the thermodilution method to calculate CO (32). Since there is no reference for CCUS, and only one method was used with comparison between the observers, a percentage error of less than 20% was defined as clinically acceptable. This would mean that the CO difference between medical students and experts would be less than 0.5 L/min in the lower end of the CO spectrum (e.g., when the experts measured a CO of 2.5 L/min, a CO of 2.0 – 3.0 L/min by the medical student would be clinically acceptable). An alpha error of 0.05 was used to indicate statistical significance. Statistical analyses were conducted using STATA version 15.0 (StataCorp, College Station, USA).

Results

CCUS acquisition and images

Between March 27th, 2015 and July 22nd, 2017, sixteen medical students were involved in the study and 1212 patients fulfilled inclusion criteria. Of these, in a total of 1155 patients CCUS was performed, as in 40 patients there was interference with clinical care during the first 24 hours of admission (e.g., the patient was in severe hemodynamic instability or an intervention was

being performed) and 17 patients had isolation restriction measures. Of these 1155 patients, in 80 patients, clinical conditions (i.e., thoracic drains, wounds, or subcutaneous emphysema) prohibited the image acquisition by CCUS, leaving 1075 patients with ultrasonography data (Figure 1).

The medical students deemed both LVOTd and VTI unmeasurable (i.e., images were of too low a quality and no or few structures could be identified) in 129 patients (12%), the LVOTd in 46 patients (4.2%), and the VTI in 29 patients (2.6%). The parasternal short axis view did not provide any additional measurements when the LVOTd was unmeasurable in the PLAX view. Thus, 204 patients (19%) out of 1075 had no CO measurement, leaving a total of 871 patients (81%) with a measured CO by medical students.

CCUS quality of images

The experts used the images obtained by the medical students and were unable to measure both the LVOTd and VTI in 152/1075 (14%), LVOTd in 76/1075 (7.1%), and VTI in 64/1075 (6.0%). While the experts deemed more measurements to be impossible in the obtained images compared to the medical students, the experts were also able to add 23 CO measurements in patients where medical students judged the images to be of too poor a quality and consequently did not perform the measurements. In total, the experts measured CO in 783 patients (73%). Comparisons of CO measurements by medical students and experts were possible in 760 (71%) out of 1075 patients in case of adequate image quality (Figure 1).

Differences in patient baseline characteristics were found between the group in which experts could measure a CO and the group in which experts could not measure a CO (Table 1). Patients without CO measured by experts were characterized by older age, greater illness severity (reflected in higher APACHE IV scores), higher heart rate, greater prevalence of chronic obstructive pulmonary disease (COPD), higher rates of mechanical ventilation, greater likelihood of being post-operative, and higher vasopressor dose

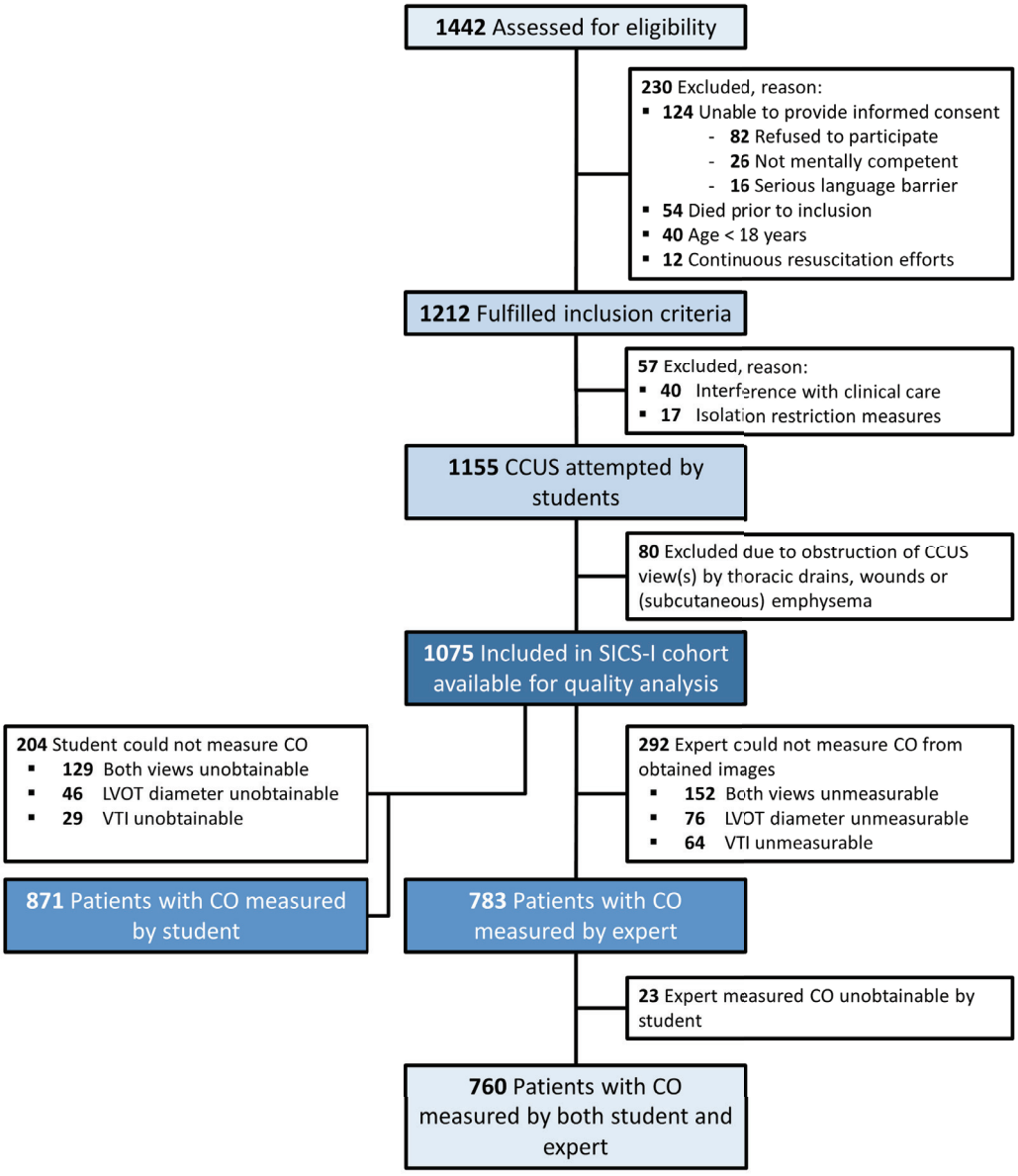


Figure 1. Flow diagram of the Simple Intensive Care Studies-I (SICS-I). Abbreviations: CCUS, critical care ultrasonography; CO, cardiac output; LVOT, left ventricular outflow tract; VTI, velocity time interval

Table 1. Patient characteristics

Variable		Patients without CO measurement (N = 292)	Patients with CO measurement (N = 783)	p value
Age (years)		64 ± 13	61 ± 15	0.004
Male gender		190 (65%)	484 (62%)	0.33
BMI (kg/m^2)		26.9 ± 5.3	26.9 ± 5.6	0.96
Respiratory rate (bpm)		18 ± 6	18 ± 6	0.88
Mechanical ventilation		194 (66%)	438 (56%)	0.002
PEEP (cm H_2O)		7 (5, 8)	7 (5, 8)	0.83
SBP (mmHg)		113 ± 25	120 ± 25	<0.001
DBP (mmHg)		59 ± 12	60 ± 12	0.44
MAP (mmHg)		76 ± 14	79 ± 14	0.014
Heart rate (bpm)		91 ± 22	87 ± 21	0.002
Atrial fibrillation		22 (8%)	56 (7%)	0.91
Norepinephrine		168 (58%)	361 (46%)	<0.001
CVP (mmHg)		9 (4 – 12)	9 (5 – 13)	0.84
Lactate (mmol/L)		1.5 (1.0 – 2.5)	1.3 (0.9 – 2.1)	<0.001
Consciousness	Alert	75 (26%)	254 (32%)	0.018
	reacting to Voice	49 (17%)	154 (20%)	
	reacting to Pain	22 (8%)	67 (9%)	
	Unresponsive	146 (49%)	308 (39%)	
COPD		54 (18%)	88 (11%)	0.002
Acute surgery		108 (37%)	230 (29%)	0.017
Post cardiothoracic surgery		40 (14%)	48 (6%)	<0.001
SAPS-II		49 ± 17	46 ± 17	0.004
APACHE IV score		80 ± 30	75 ± 29	0.017
90-day mortality		80 (27%)	217 (28%)	0.97

Abbreviations: APACHE; acute physiology and chronic health evaluation, BMI; body mass index, bpm; beats per minute, CO; cardiac output, CVP; central venous pressure, DBP; diastolic blood pressure, MAP; mean arterial pressure, PEEP; positive end-expiratory pressure, SAPS; simple acute physiology score, SBP; systolic blood pressure.

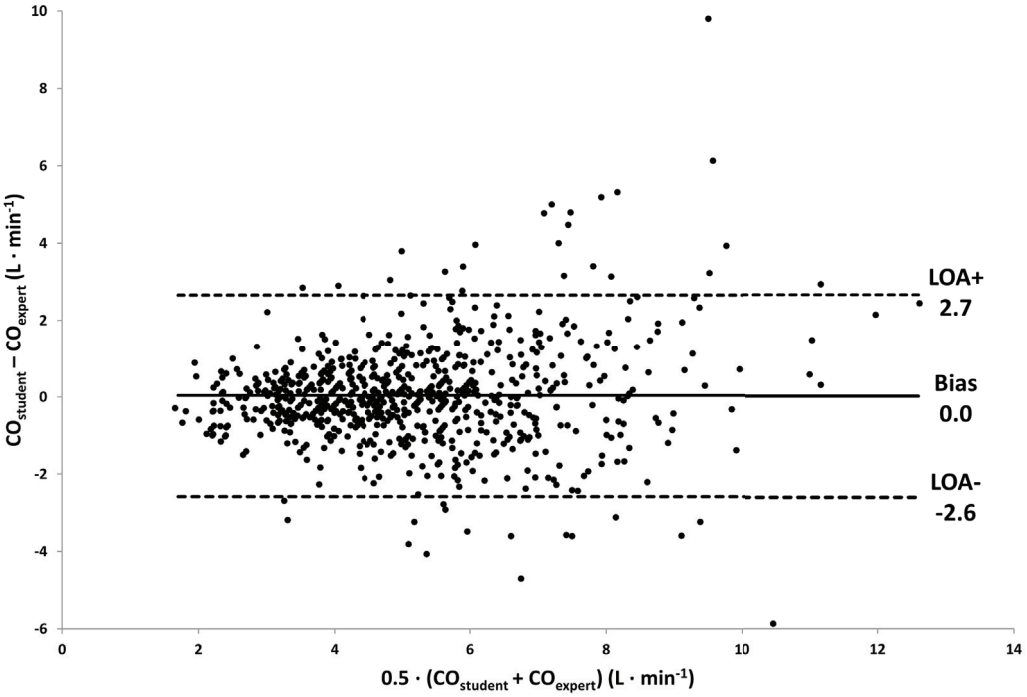


Figure 2. Bland-Altman plot showing the comparison between cardiac output measured by medical students ($CO_{student}$) and core lab experts (CO_{expert}). The mean bias between CO_{expert} and $CO_{student}$ and the upper and lower limits of agreement (LOA) are presented

Comparison of CO measurement by medical students and experts

The mean $CO_{student}$ was 5.2 ± 2.0 L/min and CO_{expert} was 5.2 ± 1.8 L/min (p=0.44). Bland-Altman analysis demonstrated a bias of 0.0 L/min (95% CI -0.06 – 0.13) with limits of agreement of -2.6 L/min (95% CI -2.7 – -2.4) to 2.7 L/min (95% CI 2.5 – 2.8) (Figure 2). Plotting a regression line in the Bland-Altman plot showed a proportional bias of 2%. The percentage error was 50% (95% CI 47 – 53). The ICC was 0.75 (95% CI 0.72 – 0.78).

Comparison of LVOTd and VTI measurements by medical students and experts

The medical students measured 900 LVOTd and the experts measured 847. There were 815 paired LVOTd measurements. Mean LVOTd by medical students ($LVOT_{dstudent}$) was 2.06 ± 0.24 cm, whereas the mean of the LVOTd measured by experts ($LVOTd_{expert}$) was 2.09 ± 0.18 cm (p<0.001). Bland-Altman analysis showed a bias of 0.0 cm (95% CI 0.0 – 0.0) with limits of agreement of -0.5 cm (95% CI -0.5 – -0.4) to 0.4 cm (95% CI 0.4 – 0.4) (see supplements). The percentage error was 21% (95% CI 20 – 23). There was a proportional bias of 20% (0.41 cm). The ICC was 0.43 (95% CI 0.37 – 0.48).

The medical students measured 917 VTI and the experts measured 859. There were 840 paired VTI measurements. Mean VTI by medical students ($VTI_{student}$) was 19.0 ± 5.6 cm compared to 18.5 ± 5.4 cm of the experts (VTI_{expert}) ($p<0.001$). Bland-Altman analysis showed a bias of 0.5 cm (95% CI 0.4 – 0.7) with limits of agreement of -5.0 cm (95% CI -5.3 – -4.6) to 6.1 cm (95% CI 5.7 – 6.4) (see supplements). The percentage error was 30% (95% CI 28 – 31). The ICC was 0.86 (95% CI 0.84 – 0.88).

Discussion

In this large prospective ICU cohort study with CCUS, we found that, after dedicated training, medical students were able to acquire a CO measurement in three out of every four patients (871 of 1155 patients). This finding is of interest considering that the medical students were ultrasound naïve, the CO measurement is considered an advanced CCUS skill, and the ICU population is known for technical difficulties in acquiring ultrasound images. In a minority of ICU patients (80 of the 1155 patients) CCUS was not possible due to clinical conditions hampering image acquisition, leaving 1075 patients with ultrasonography data. The CCUS images obtained by medical students were assessed by experts and rated to be of adequate quality in 73%. Patients (292 of 1075 patients) in which no adequate image quality could be obtained were more often mechanically ventilated, admitted after cardiothoracic surgery or were more severely ill.

Although the students reached a reasonable percentage on image acquisition/quality, our data do not support CO measurements by medical students (after limited training), as comparison to CO measurements by experts showed poor agreement. CCUS concerns more than acquiring the required images and any operator should be aware of the potential errors that can be made with ultrasonography, especially in complex critically ill patients (33). It is important to note that education on ultrasonography should focus on specific training and quality control on all aspects of ultrasonography in order to achieve accurate measurements (17). Our results are in line with recommendations by the European Association of Cardiovascular Imaging (EACVI) on point-of-care, problem-oriented focus cardiac ultrasound examination (FoCUS), which state that supervision and quality control by experts are essential for proper and complete examination. Quality control in our study was performed by an accredited echocardiographic laboratory as is recommended in this viewpoint (15).

To be able to compare our results to those of other studies, it is of utmost importance that every step, from eligible patients to the number of patients in which a reliable CO measurement by CCUS is obtained, are presented. Currently these numbers are often lacking, and this leads to varying success rates on the feasibility of CCUS. If reported, results may vary based on differences in ultrasonography training and experience, which impedes a comparison of image acquisition and quality. We found four studies, on measuring CO in critically ill patients by non-experts to compare with our study (see supplements) (22, 23, 34, 35). In two out of the three studies the operators

had previous experience with ultrasonography, but training varied (23, 34). The setting, sampling, and exclusion criteria may explain the reported high success rate in one study over another (22). Whether images obtained are of sufficient quality should preferably be judged by independent experts, as two out of three studies did (22, 34). In one study independent investigators assessed the quality, however, it is not clear if these were experts or not (35). The percentage of adequate/good-quality images in our study was comparable with Dinh et al. In the study of Betcher et al. and Villavicencio et al., image quality was generally (judged) overall lower. Duration of training or differences in baseline characteristics might explain part of these differences.

The final step to obtain a reliable CO measurement is to measure LVOTd and VTI on images of sufficient quality. Dinh et al. and Lee et al. reported data on measurement quality, and, furthermore, Dinh et al. reported a low bias between sonographers and independent experts. These studies and ours showed lack in precision for CO measurement by novices. Villavicencio et al. compared ultrasonography derived CO with the transpulmonary thermodilution technique and concluded that there was an acceptable level of agreement between the techniques. Furthermore, they found a high inter- and intra-observer reliability.

Ultrasonography in the acute setting remains challenging, and data regarding novice-based CCUS are limited (see supplements). In our study we chose for medical students as novices (i.e., non-experts), since non-experts constitute the majority of ultrasound trained personnel in an IC and as students would not interfere with daily ICU care. Five studies reported on medical students performing CCUS in critically ill patients (3 in ED setting, 1 in operating theatre and 1 in ICU) (3, 7–10). Four out of the five studies showed that images could be acquired in a promising 82-98% of cases. The studies reporting on image quality showed percentages of (at least) adequate imaging ranging from 89 to 98%, unfortunately by non-independent judging (3, 7). Furthermore, after training, medical students can adequately interpret images with a very simplified or binary assessment (36). A number of previous studies employed training curricula for medical students on ultrasonography protocols (37–39). Four other studies used a point-of-care ultrasonography training program to determine diagnostic performance in various clinical scenarios (36, 40–42). All studies showed feasibility to train medical students to perform ultrasonography after a relatively short amount of training, which is comparable to the training medical students received in our study.

In previous manuscripts on SICS study data we reported a higher percentage of images judged to be of sufficient quality (25, 26). The current results showed the percentage of measurements of CO considered of sufficient quality by a core-laboratory and not images with a LVOT and VTI. The higher level of quality considered necessary is according to internal protocol and is independently monitored.

Limitations

First, the proportion of patients with an acoustic window was based on the results of CCUS by medical students only. We did not check if more experienced sonographers were able to retrieve images in these cases, because the design of our study was to obtain images outside patient care. We believe image quality can only be assessed if the observers are blinded for all other study data and are not involved in the patient's clinical care. Ideally, independent experts perform ultrasonography themselves and make a direct comparison with the medical student. The availability of time and staff outside clinical care in our center was limited, leading us to include all consecutive patients and allow trained medical students to run the study.

Second, we did not check for interindividual variation of skills and quality of CCUS in each medical student who participated in the study, mainly to limit the time of investigation at the bedside.

Third, CCUS of the heart was limited to 2D imaging of the LVOTd, 2D imaging of the AP5CH, and pulse wave Doppler imaging of the LVOT. Therefore, valvular disease could have been missed.

Conclusions

Medical students as novices were capable of performing CCUS with adequate image acquisition in the majority of an ICU population of acutely admitted critically ill patients. However, they cannot accurately measure a CCUS derived cardiac output after limited training. Cardiac output measurements with CCUS in research and daily care should be interpreted with caution if not validated by experts; this is in concordance with the viewpoint of the EACVI on CCUS.

Supplementary material

Supplements are available online:
doi.org//10.1186/s13089-020-0152-5

References

1 Mayo PH, Beaulieu Y, Doelken P, Feller-Kopman D, Harrod C, Kaplan A, et al. American College of Chest Physicians/La Société de Réanimation de Langue Française statement on competence in critical care ultrasonography. Chest. 2009 Apr;135(4):1050–60.

2 Koster G, Van Der Horst ICC. Critical care ultrasonography in circulatory shock. Curr Opin Crit Care. 2017 Aug;23(4):326–33.

3 Filipiak-Strzecka D, John B, Kasprzak J, Michalski B, Lipiec P. Pocket-size echocardiograph - a valuable tool for non-experts or just a portable device for echocardiographers? Adv Med Sci. 2013 Jun;58(1):67–72.

4 AlEassa EM, Ziesmann MT, Kirkpatrick AW, Wurster CL, Gillman LM. Point of care ultrasonography use and training among trauma providers across Canada. Can J Surg. 2016 Feb 1;59(1):6–8.

5 Funcke S, Sander M, Goepfert MS, Groesdonk H, Heringlake M, Hirsch J, et al. Practice of hemodynamic monitoring and management in German, Austrian, and Swiss intensive care units: the multicenter cross-sectional ICU-CardioMan Study. Ann Intensive Care. 2016 Dec 31;6(1):49.

6 Yang Y, Royse C, Royse A, Williams K, Canty D. Survey of the training and use of echocardiography and lung ultrasound in Australasian intensive care units. Crit Care. 2016 Dec 24;20(1):339.

7 Panoulas VF, Daigeler A-L, Malaweera ASN, Lota AS, Baskaran D, Rahman S, et al. Pocket-size hand-held cardiac ultrasound as an adjunct to clinical examination in the hands of medical students and junior doctors. Eur Hear J - Cardiovasc Imaging. 2013 Apr;14(4):323–30.

8 Ho AM-H, Critchley LAH, Leung JYC, Kan PKY, Au SS, Ng SK, et al. Introducing Final-Year Medical Students to Pocket-Sized Ultrasound Imaging: Teaching Transthoracic Echocardiography on a 2-Week Anesthesia Rotation. Teach Learn Med. 2015 Jul 3;27(3):307– 13.

9 Krause C, Krause R, Krause R, Gomez N, Jafry Z, Dinh VA. Effectiveness of a 1-Hour Extended Focused Assessment With Sonography in Trauma Session in the Medical Student Surgery Clerkship. J Surg Educ. 2017 Nov;74(6):968–74.

10 Udrea DS, Sumnicht A, Lo D, Villarreal L, Gondra S, Chyan R, et al. Effects of Student-Performed Point-of-Care Ultrasound on Physician Diagnosis and Management of Patients in the Emergency Department. J Emerg Med. 2017 Jul;53(1):102–9.

11 Bobbia X, Chabannon M, Chevallier T, de La Coussaye JE, Lefrant JY, Pujol S, et al. Assessment of five different probes for lung ultrasound in critically ill patients: A pilot study. Am J Emerg Med. 2018 Jul;36(7):1265–9.

12 Rice BT, Vu H, Tran LD, Vo QX, Mowafi H. Survey of point of care ultrasound usage in emergency medicine by Vietnamese physicians. Emerg Med Australas. 2015 Dec;27(6):580–3.

13 Nielsen K, Lauridsen JRM, Laursen CB, Brabrand M. Physicians using ultrasound in Danish emergency departments are mostly summoned specialists. Scand J Trauma Resusc Emerg Med. 2015 Dec 7;23(1):51.

14 Sanders J, Noble V, Raja A, Sullivan A, Camargo C. Access to and Use of Point-of-Care Ultrasound in the Emergency Department. West J Emerg Med. 2015 Sep 15;16(5):747–52.

15 Neskovic AN, Skinner H, Price S, Via G, De Hert S, Stankovic I, et al. Focus cardiac ultrasound core curriculum and core syllabus of the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging. 2018 May 1;19(5):475–81.

16 International expert statement on training standards for critical care ultrasonography. Intensive Care Med. 2011 Jul 26;37(7):1077–83.

17 Price S, Via G, Sloth E, Guarracino F, Breitkreutz R, Catena E, et al. Echocardiography practice, training and accreditation in the intensive care: document for the World Interactive Network Focused on Critical Ultrasound (WINFOCUS). Cardiovasc Ultrasound. 2008 Dec 6;6(1):49.

18 Cecconi M, De Backer D, Antonelli M, Beale R, Bakker J, Hofer C, et al. Consensus on circulatory shock and hemodynamic monitoring. Task force of the European Society of Intensive Care Medicine. Intensive Care Med. 2014 Dec 13;40(12):1795–815.
19 Sakr Y, Reinhart K, Vincent J-L, Sprung CL, Moreno R, Ranieri VM, et al. Does dopamine administration in shock influence outcome? Results of the Sepsis Occurrence in Acutely Ill Patients (SOAP) Study. Crit Care Med. 2006 Mar;34(3):589–97.
20 Vieillard-Baron A, Mayo PH, Vignon P, Cholley B, Slama M, Pinsky MR, et al. International consensus statement on training standards for advanced critical care echocardiography. Intensive Care Med. 2014;40(5):654–66.
21 Millington SJ, Goffi A, Arntfield RT. Critical care echocardiography: a certification pathway for advanced users. Can J Anesth. 2018;65(4):345–9.
22 Dinh VA, Ko HS, Rao R, Bansal RC, Smith DD, Kim TE, et al. Measuring cardiac index with a focused cardiac ultrasound examination in the ED. Am J Emerg Med. 2012 Nov;30(9):1845–51.
23 Lee L, Tewalde S, Allison M, Witting M, Euerle B. Inter-Rater Reliability of Doppler Echocardiography Assessment of Cardiac Output by Emergency Physicians. Ann Emerg Med. 2014 Oct;64(4):S126.
24 Andersen GN, Viset A, Mjølstad OC, Salvesen Ø, Dalen H, Haugen BO. Feasibility and accuracy of point-of-care pocket-size ultrasonography performed by medical students. BMC Med Educ. 2014 Dec 28;14(1):156.
25 Hiemstra B, Eck RJ, Koster G, Wetterslev J, Perner A, Pettilä V, et al. Clinical examination, critical care ultrasonography and outcomes in the critically ill: cohort profile of the Simple Intensive Care Studies-I. BMJ Open. 2017 Sep;7(9):e017170.
26 Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, et al. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med. 2019 Feb;45(2):190–200.
27 Qasim A. Echocardiographer.org. [cited 2019 Oct 25]. Available from: http://echocardiographer.org/
28 Baumgartner H, Hung J, Bermejo J, Chambers JB, Edvardsen T, Goldstein S, et al. Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. Eur Heart J Cardiovasc Imaging. 2017 Mar;18(3):254–75.
29 Quiñones MA, Otto CM, Stoddard M, Waggoner A, Zoghbi WA, Doppler Quantification Task Force of the Nomenclature and Standards Committee of the American Society of Echocardiography. Recommendations for quantification of Doppler echocardiography: a report from the Doppler Quantification Task Force of the Nomenclature and Standards Committee of the American Society of Echocardiography. J Am Soc Echocardiogr. 2002 Feb;15(2):167–84.
30 Royse CF, Seah JL, Donelan L, Royse AG. Point of care ultrasound for basic haemodynamic assessment: novice compared with an expert operator. Anaesthesia. 2006 Jul 27;61(9):849–55.
31 Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999 Jun;8(2):135–60.
32 Critchley LAH, Critchley JAJH. A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. J Clin Monit Comput. 1999 Feb;15(2):85–91.
33 Orde S, Slama M, Hilton A, Yastrebov K, McLean A. Pearls and pitfalls in comprehensive critical care echocardiography. Crit Care. 2017 Dec 17;21(1):279.
34 Betcher J, Majkrzak A, Cranford J, Kessler R, Theyyunni N, Huang R. Feasibility study of advanced focused cardiac measurements within the emergency department. Crit Ultrasound J. 2018 May 25;10(1):10.

35 Villavicencio C, Leache J, Marin J, Oliva I, Rodriguez A, Bodí M, et al. Basic critical care echocardiography training of intensivists allows reproducible and reliable measurements of cardiac output. Ultrasound J. 2019 Dec 16;11(1):5.
36 Lebeau R, Sas G, Rayes M El, Serban A. Left ventricular ejection fraction assessment by non-cardiologists from transverse views using a simplified wall motion score index. 2015;(March).
37 Heiberg J, Hansen LS, Wemmelund K, Sørensen AH, Ilkjaer C, Cloete E, et al. Point-of-Care Clinical Ultrasound for Medical Students. Ultrasound Int open. 2015 Nov;1(2):E58-66.
38 Kobal SL, Lior Y, Ben-Sasson A, Liel-Cohen N, Galante O, Fuchs L. The feasibility and efficacy of implementing a focused cardiac ultrasound course into a medical school curriculum. BMC Med Educ. 2017 May 30;17(1):94.
39 Wong I, Jayatilleke T, Kendall R, Atkinson P. Feasibility of a focused ultrasound training programme for medical undergraduate students. Clin Teach. 2011 Mar;8(1):3–7.
40 Parks AR, Verheul G, Leblanc-Duchin D, Atkinson P. Effect of a point-of-care ultrasound protocol on the diagnostic performance of medical learners during simulated cardiorespiratory scenarios. Can J Emerg Med. 2015 May;17(3):263–9.
41 Arora S, Cheung AC, Tarique U, Agarwal A, Firdouse M, Ailon J. First-year medical students use of ultrasound or physical examination to diagnose hepatomegaly and ascites: a randomized controlled trial. J Ultrasound. 2017 Sep;20(3):199–204.
42 Ang J, Doyle B, Allen P, Cheek C. Teaching bedside ultrasound to medical students. Clin Teach. 2018;15(4):331–5.


58 3

7

Disagreement in cardiac output measurements between fourth-generation FloTrac and critical care ultrasonography in patients with circulatory shock: a prospective observational study

Thomas Kaufmann
Ramon P. Clement
Bart Hiemstra
Jaap Jan Vos
Thomas W.L. Scheeren
Frederik Keus
Iwan C.C. van der Horst
SICS Study Group

Journal of Intensive Care

Abstract

Background

Cardiac output measurements may inform diagnosis and provide guidance of therapeutic interventions in patients with hemodynamic instability. The FloTrac™ algorithm uses uncalibrated arterial pressure waveform analysis to estimate cardiac output. Recently, a new version of the algorithm has been developed. The aim was to assess the agreement between FloTrac™ and routinely performed cardiac output measurements obtained by critical care ultrasonography in patients with circulatory shock.

Methods

A prospective observational study was performed in a tertiary hospital from June 2016 to January 2017. Adult critically ill patients with circulatory shock were eligible for inclusion. Cardiac output was measured simultaneously using FloTrac™ with a fourth-generation algorithm (CO_{AP}) and critical care ultrasonography (CO_{CCUS}). The strength of linear correlation of both methods was determined by the Pearson coefficient. Bland-Altman plot and four-quadrant plot were used to track agreement and trending ability.

Result

Eighty-nine paired cardiac output measurements were performed in 17 patients during their first 24 h of admittance. CO_{AP} and CO_{CCUS} had strong positive linear correlation ($r^2 = 0.60$, $p < 0.001$). Bias of CO_{AP} and CO_{CCUS} was 0.2 L/min (95% CI – 0.2 to 0.6) with limits of agreement of – 3.6 L/min (95% CI – 4.3 to – 2.9) to 4.0 L/min (95% CI 3.3 to 4.7). The percentage error was 65.6% (95% CI 53.2 to 77.3). Concordance rate was 64.4%.

Conclusions

In critically ill patients with circulatory shock, there was disagreement and clinically unacceptable trending ability between values of cardiac output obtained by uncalibrated arterial pressure waveform analysis and critical care ultrasonography.

Background

Critically ill patients with circulatory shock have increased risks of multi-organ failure, long-term morbidity, and mortality (1). Advanced hemodynamic monitoring in these patients may inform diagnosis and simultaneously guide management by providing insight into cardiac function, cardiac preload, and afterload (2). Several methods for measuring cardiac output (CO) exist, varying from invasive (e.g., thermodilution by pulmonary artery catheter (PAC)) to minimally invasive (e.g., pulse contour analysis by FloTrac™ (Edwards Lifesciences, Irvine, USA)) or even non-invasive (e.g., transthoracic Doppler ultrasound by critical care ultrasonography (CCUS)). These methods all have their own merits, disadvantages and requirements (3).

One type of pulse contour analysis is the uncalibrated arterial pressure waveform analysis method to estimate CO (APCO). Reliability of APCO is questioned in patients with hemodynamic instability, and this occurs frequently in patients admitted to the ICU (4). Therefore, CO measurements obtained by APCO should be interpreted with caution in critically ill patients with circulatory shock (5,6).

The FloTrac™ system using the APCO method calculates CO based on the principle that aortic pulse pressure is proportional to stroke volume (SV) and inversely related to aortic compliance using a proprietary algorithm. FloTrac™ has been widely studied in more than 70 validation studies as of yet, mostly showing adequate performance in normo- and hypodynamic conditions, but not in patients with large changes in vascular tone which typically occur in patients with circulatory shock (7). However, these studies vary by the statistical methods and versions of the algorithm used. Recently, the fourth-generation algorithm was developed to improve performance.

Evaluation of the trending ability rather than the agreement of absolute values of CO monitoring devices is increasingly considered in validation studies for assessment of potential clinical usefulness (8). In addition to one single CO measurement for diagnosing circulatory shock, repeated measurements of CO informing the trending ability could be informative for monitoring and guidance of supportive treatments of patients with circulatory shock.

The aim of our study was to compare both agreements and trending ability for APCO measurements of CO (CO_{AP}) with CO routinely measured by CCUS (CO_{CCUS}) in critically ill patients with circulatory shock. CCUS was chosen as the reference standard since it is the preferred method for diagnosis, but not for monitoring, of circulatory shock in critically ill patients and is widely available (2, 9). Importantly, it should be noted that CCUS is not a gold standard reference technique for method comparison studies aiming to evaluate the validity of CO monitors (10).

Methods

This study was a substudy of the Simple Intensive Care Studies-I (SICS-I), which was a single-centre, prospective, observational cohort study in which all consecutive acutely admitted adult patients expected to stay beyond 24 h were included (NCT02912624) (16, 17). The STROBE guidelines for reporting observational studies were used (Additional file 1) (11). The checklist for CO monitor method comparison studies was used (10). The local institutional review board (Medisch Ethische Toetsingscommissie, University Medical Center Groningen) approved the study (M15.168207 and M16.193856). Written informed consent was obtained from all patients.

Selection criteria

In this substudy, all consecutive acutely admitted adult patients with suspected circulatory shock and expected to stay beyond 48 h were included from June 2016 to January 2017. The circulatory shock was defined as the requirement of any dose of vasopressor to maintain a mean arterial pressure (MAP) of 60 mmHg or if the MAP remained below 70 mmHg despite fluid resuscitation (defined by at least 1000 mL of crystalloids). In addition, at least one other sign of organ or tissue hypoperfusion had to be present: altered state of mind (Alert-Voice-Pain-Unresponsive scale) (12), mottled skin (Mottling score ≥ 1 (13)), decreased urine output (≤ 0.3 mL/kg/h) or increased serum lactate level (≥2 mmol/L). Exclusion criteria were inability to obtain sufficient quality CCUS images; no arterial line; atrial fibrillation; and aortic valve or mitral valve diseases known to impair the arterial waveform. We included this group of patients because CO measurements are indicated to identify the type of shock, select necessary therapeutic interventions and evaluate patient's response to therapy (2).

Objectives

The primary objective was to evaluate CO_{AP} measurements in terms of the agreement and trending ability against CO_{CCUS} as reference technique in patients with circulatory shock.

Definitions and bias

Patient characteristics including clinical, hemodynamic and laboratory variables as well as Acute Physiology and Chronic Health Evaluation (APACHE) IV and Simplified Acute Physiology Score (SAPS) II values were recorded (14,15). Measurements were performed following protocolized definitions and procedures (16, 17).

In short, CO_{CCUS} was measured by transthoracic echocardiography using the Vivid-S6 system (General Electric, Horton, Norway) with cardiac probe M3S or M4S, and with default cardiac imaging setting. The parasternal long axis was used to measure the left ventricular outflow tract diameter. In the apical five-chamber view, a pulse wave Doppler signal in the left ventricular outflow tract was used to measure the velocity time integral. CO_{CCUS} was calculated using an established formula (18). CCUS was performed after ICU admission within 6 h and repeated once every 24 h after admission provided there was no interference with clinical care. Researchers were trained in performing CCUS by experienced cardiologist-intensivists.

The FloTrac™ sensor was connected to an indwelling radial artery catheter and an EV1000TM monitor (version 4.00; Edwards Lifesciences, Irvine, USA), which continuously displayed CO_{AP} values. The value of CO_{AP} displayed on the EV1000™ monitor was registered simultaneously (i.e., 'beat-to-beat') with each CO_{CCUS} measurement.

All measurements, including CCUS findings, were kept blind for the caregivers. Quality of CCUS images and CO_{CCUS} measurements were validated by an independent specialized core laboratory (Groningen Image Core Lab) blinded for the CO_{AP} measurements.

Statistical analysis

No formal sample size calculation was performed due to lack of data on CO_{AP} variation in patients with circulatory shock. Therefore, this study has an exploratory nature.

Data were presented as means with standard deviations or medians with interquartile ranges depending on distributions. Normality of data was checked using the Shapiro-Wilk test. Dichotomous and categorical data were presented in proportions.

Correlations were assessed by scatter plot, and the strength of linear correlation was determined by calculating a Pearson (r) coefficient. Bland-Altman analyses of repeated measurements in each patient were performed and means (bias) and SD of the differences, 95% limits of agreement (LOA) (=mean difference ± 1.96 × SD of the difference) as well as the percentage error of CO_{AP} versus CO_{CCUS} were calculated (19, 20). To evaluate the trending ability of CO_{AP} versus CO_{CCUS} a four-quadrant plot was used and the concordance rate was calculated using an exclusion zone of 0.5 L min−1 (21). For statistical analysis, we used STATA version 15.0 (StataCorp, College Station, USA).

Results

Participants

During the study period, 184 patients were diagnosed with circulatory shock, but only 24 patients appeared eligible for this study. One hundred patients who had circulatory shock were not included as they were expected to stay for less than 48 h, and 60 patients with circulatory shock were not included because CCUS was not possible or image quality was insufficient to perform measurements. Six patients had to be excluded because study procedures interfered with clinical care, leaving 18 patients to be included. One patient was excluded afterwards for invalid CO_{AP} measurements due to improper use of a FloTrac™ sensor. Thus, 17 patients were included in the final analyses (Fig. 1).

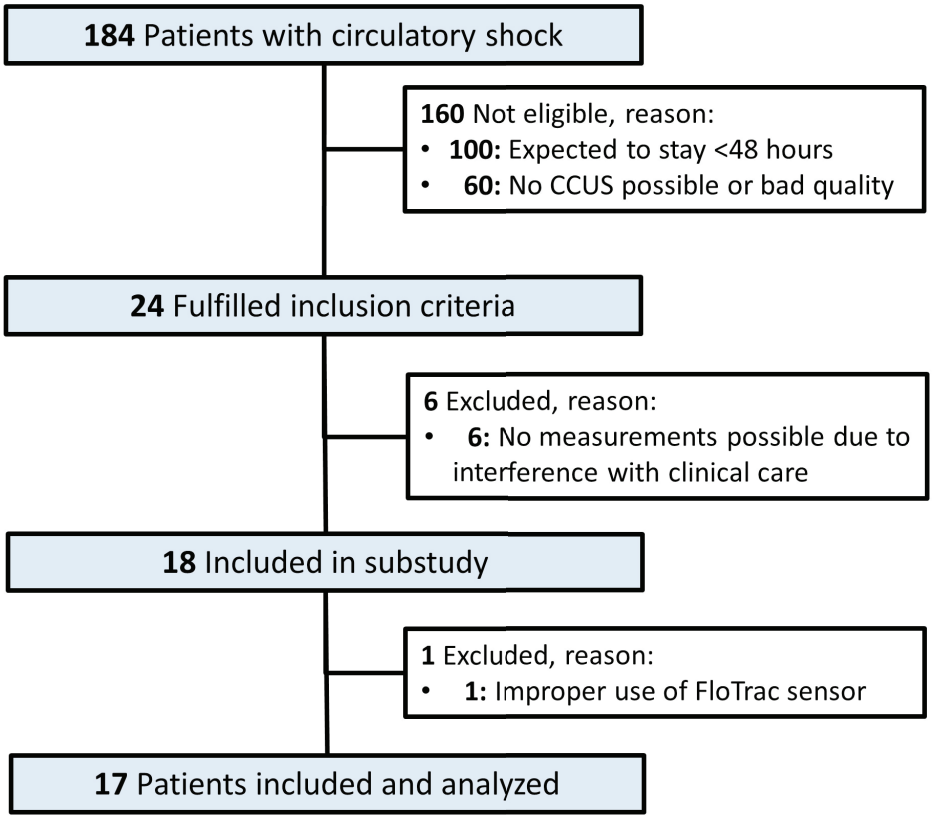


Figure 1. Patient flowchart. Abbreviations: CCUS, critical care ultrasonography

Bias, precision and correlation

The characteristics of the 17 included patients are shown in Table 1 (and Additional file 2: Table S1). The mean CO_{AP} and CO_{CCUS} for 89 paired measurements were 5.9 ± 1.9 L/min and 5.7 ± 2.0 L/min, respectively (p = 0.24). A significant correlation was observed for all CO measurements ($r^2 = 0.60$, $p < 0.001$) (Fig. 2). Bias was 0.2 L/min (95% CI − 0.2 to 0.6), with LOA of − 3.6 L/min (95% CI − 4.3 to − 2.9) to 4.0 L/min (95% CI 3.3 to 4.7) (Fig. 3). Plotting a regression line in the Bland-Altman plot gave no arguments for proportional bias (line not shown). The overall percentage error was 65.6% (95% CI 53.2 to 77.3). Individual cardiac output measurements for each patient are provided in Additional file 3: Table S2.

Table 1. *Patient characteristics*

Patient characteristics (n=17)		
Age (years)		65 (9)
Male gender, n (%)		14 (82%)
Body Mass Index (kg/m^2)		25.7 (4.7)
Clinical characteristics on study inclusion		
Heart rate (bpm)		95 (26)
Systolic arterial pressure (mmHg)		102 (15)
Diastolic arterial pressure (mmHg)		55 (6)
Mean arterial pressure (mmHg)		69 (7)
Norepinephrine therapy, n (%)		16 (94%)
Norepinephrine dose (µg/kg/min)		0.13 (0.05, 0.38)
Mechanical ventilation, n (%)		12 (71%)
Positive end-expiratory pressure (cm H_2O)		8 (6, 9)
AVPU score, n (%)	Alert	2 (12%)
	Verbal	4 (24%)
	Passive	1 (6%)
	Unresponsive	10 (58%)
Mottling score, n (%)	None	3 (19%)
	Modest	2 (13%)
	Mild	7 (43%)
	Moderate	3 (19%)
	Severe	1 (6%)
Urine output (mL/kg/h)		0.49 (0.26, 0.66)
Lactate (mmol/L)		1.7 (1.4, 3.4)
APACHE IV – score (points)		92 (32)
SAPS II – score (points)		56 (17)

Data are presented as the mean and standard deviation, median and interquartile ranges or as absolute frequencies with percentages as appropriate. Abbreviations: AVPU, alert, verbal pain, unresponsive; APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score

Trending ability

For assessment of trending ability 72 paired measurements were analyzed. Trending of measurements was evaluated using a four-quadrant plot (Fig. 4). Forty-five paired measurements showed a clinically relevant change, which was defined as larger than 0.5 L/min. The concordance rate was 64.4%.

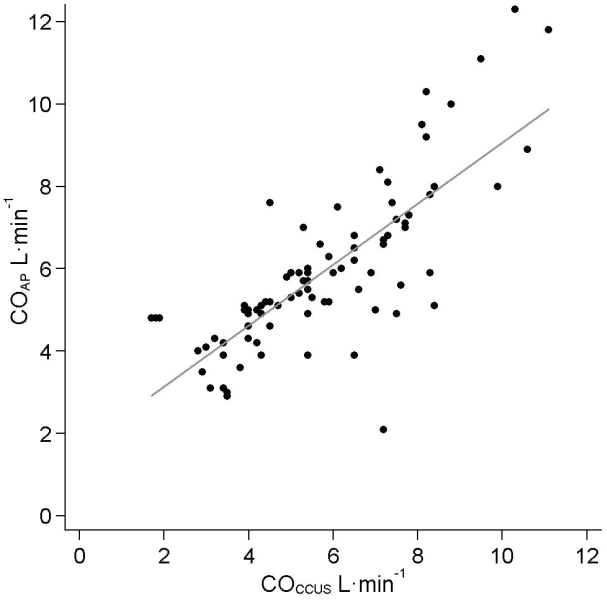


Figure 2. Scatter plot of cardiac output measured by FloTrac™ and CCUS. Abbreviations: CO_{AP}, cardiac output measured using fourth-generation FloTrac™ algorithm; CO_{CCUS}, cardiac output measured by critical care ultrasonography

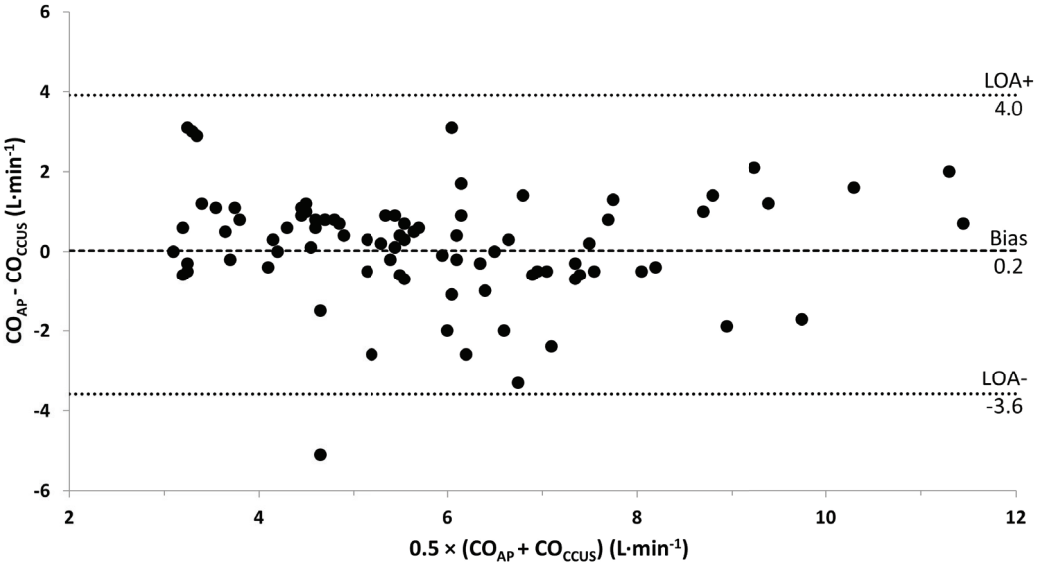


Figure 3. Bland-Altman plot for repeated measurements showing the comparison between CO_{AP} and CO_{CCUS}. The mean bias between CO_{AP} and CO_{CCUS} and the upper and lower limits of agreement (LOA) are presented. Abbreviations: CO_{AP}, cardiac output measured using fourth-generation FloTrac™ algorithm; CO_{CCUS}, cardiac output measured by critical care ultrasonography

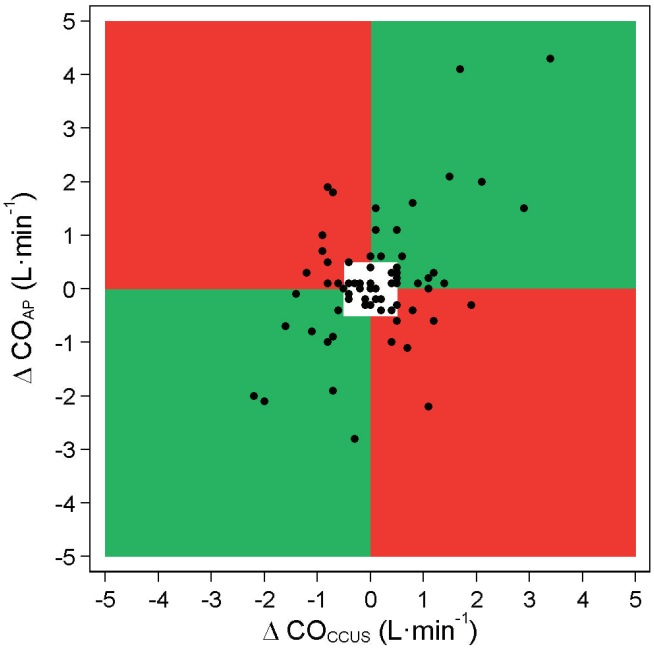


Figure 4. *Four-quadrant plot showing the trend of CO_{AP} versus CO_{CCUS}. Exclusion zone of 0.5 L/min (white area). Abbreviations: CO_{AP}, cardiac output measured using fourth-generation FloTrac™ algorithm; CO_{CCUS}, cardiac output measured by critical care ultrasonography*

Discussion

In this prospective observational study, agreement and trending ability of CO_{AP} was compared with CO_{CCUS} in critically ill patients with circulatory shock. CO_{AP} showed a low bias of 0.2 L/min but a large percentage error of 65.6% when compared with CO_{CCUS}, indicating disagreement (20). Trending ability was poor with a concordance rate of 64.4%. The new FloTrac™ algorithm should not be used for diagnosis or guidance of treatment in critically ill patients with circulatory shock.

Interpretation

There are no data on the reliability of CO measurements with the fourth-generation FloTrac™ software algorithm in critically ill patients with shock as of yet. The main concern with the previous version(s) of the APCO algorithm was the lack of reliability in tracking CO changes after hemodynamic interventions or in patients with sepsis (7, 22).

The low bias and the high percentage error of CO measurements are in accordance with results from another study, which tested the fourth-generation algorithm for tracking CO measurements after administration of phenylephrine to increase vasomotor tone in patients prior to cardiac surgery (bias –0.7 L/min; percentage error 55.4%) (23). Concordance rate for trending ability

was 87% which was higher than in our study. In that study, the chosen reference technique for measuring CO was thermodilution.

In a more recent study in patients undergoing cardiac surgery, the new FloTrac™ algorithm also showed lack of agreement and trending ability (bias –0.4 L/min; percentage error 37.1%; concordance rate 76%) (24). The reference technique was thermodilution and in that study, bias was influenced by systemic vascular resistance.

Another study tested the fourth-generation FloTrac™ algorithm in patients undergoing abdominal aortic aneurysm surgery and also found a low bias and high percentage error (bias 0.4 L/min; percentage error 46.7%) of CO measurements (25). The concordance rate for trending ability was 26.9% before and after aortic clamping and 47.3% before and after first unclamping of the iliac artery. The reference technique chosen in this study was transoesophageal echocardiography.

Advanced hemodynamic monitoring techniques are currently used to identify the type of shock, to guide choices of interventions and to evaluate the response to therapy. Less invasive hemodynamic monitoring techniques such as APCO are currently not recommended for use in patients with shock, especially when receiving vasopressors (2, 26). Our findings support this statement.

Implications and generalizability

Even though CO monitoring is considered a cornerstone in diagnosing and managing circulatory shock, the sequential evaluation of the hemodynamic state during shock is only a level 1 recommendation based on low quality of evidence (2).

The abovementioned studies validating the new fourth-generation FloTrac™ algorithm were performed in different target populations and contained different reference techniques, which limit comparability. There is a concern about the interchangeability of CO_{CCUS} and CO measurements by thermodilution, and tracking ability of the two methods has only been scarcely assessed and needs evaluation by larger studies (27)

Considerations and limitations

There are several considerations and limitations when interpreting the results of our study. First, since only parallel and no serial CO measurements were performed for each time point, the precision of individual measurements could not be assessed. While only few studies determined the precision of the CCUS and FloTrac™ technologies, it is a given that both methods have some degree of variation which influences precision of agreement (28). This might influence—and possibly overestimate—the observed bias and precision to an unknown extent, since the precision of the CCUS as reference method was not incorporated.

Second, a stepwise approach and checklist for the complete presentation of CO method comparison research have been published (10). This checklist includes a design study phase where it is encouraged that criteria for acceptable bias and LOA or percentage error are defined,

and a sample size calculation should be performed prior to the conduct of method comparison studies. In our study, we defined clinically acceptable limits based on available literature, but we did not specify a sample size in advance. The current study could serve as a pilot for a further validation study.

Third, during the study period, we included only 17 patients. Patients with circulatory shock were eligible only if they were expected to stay for longer than 48 h and if it was possible to perform CCUS. We chose this definition to ensure that a complete picture of shock treatment could be presented which allowed for the best comparison between the two methods. Last, CCUS was used as a reference technique in our study despite pulmonary or transpulmonary thermodilution being the gold standard for CO method comparison studies (10). Therefore, we cannot prove direct superiority of either method. In order to do this, a comparison with a thermodilution method will have to be performed. We chose CCUS as reference because it is currently the first-line evaluation modality in patients with circulatory shock and also because it is widely available and used in the ICU for diagnostic purposes (2, 29). However, images required to make CO_{CCUS} measurements are unobtainable in up to 20% of patients (30).

FloTrac™ measurements of CO are still not recommended in critically ill patients (5, 6), and further clinical studies comparing minimally invasive techniques for CO estimation with a reference technique are needed for further validation of these techniques and also for extending applicability to other types of patients, who were initially not the target population.

Conclusions

In critically ill patients with circulatory shock, there was disagreement and clinically unacceptable trending ability between values of cardiac output obtained by uncalibrated arterial pressure waveform analysis and critical care ultrasonography.

Supplementary material

Supplements are available online:
https://doi.org/10.1186/s40560-019-0373-5

References

1 Sakr Y, Reinhart K, Vincent J-L, Sprung CL, Moreno R, Ranieri VM, et al. Does dopamine administration in shock influence outcome? Results of the Sepsis Occurrence in Acutely Ill Patients (SOAP) Study. Crit Care Med. 2006 Mar;34(3):589–97.

2 Cecconi M, De Backer D, Antonelli M, Beale R, Bakker J, Hofer C, et al. Consensus on circulatory shock and hemodynamic monitoring. Task force of the European Society of Intensive Care Medicine. Intensive Care Med. 2014 Dec 13;40(12):1795–815.

3 Clement RP, Vos JJ, Scheeren TWL. Minimally invasive cardiac output technologies in the ICU: Putting it all together. Curr Opin Crit Care. 2017;23(4):302–9.

4 Saugel B, Wagner JY, Scheeren TWL. Cardiac output monitoring: less invasiveness, less accuracy? J Clin Monit Comput. 2016;30(6):753–5.

5 De Backer D, Bakker J, Cecconi M, Hajjar L, Liu DW, Lobo S, et al. Alternatives to the Swan– Ganz catheter. Intensive Care Med. 2018 May 3;44(6):730–41.

6 Saugel B, Vincent J-L. Cardiac output monitoring: how to choose the optimal method for the individual patient. Curr Opin Crit Care. 2018;24(3):165–72.

7 Slagt C, Malagon I, Groeneveld ABJ. Systematic review of uncalibrated arterial pressure waveform analysis to determine cardiac output and stroke volume variation. Br J Anaesth. 2014 Apr;112(4):626–37.

8 Critchley LA, Lee A, Ho AMH. A critical review of the ability of continuous cardiac output monitors to measure trends in cardiac output. Anesth Analg. 2010;111(5):1180–92.

9 Koster G, Van Der Horst ICC. Critical care ultrasonography in circulatory shock. Curr Opin Crit Care. 2017 Aug;23(4):326–33.

10 Montenij LJ, Buhre WF, Jansen JR, Kruitwagen CL, De Waal EE. Methodology of method comparison studies evaluating the validity of cardiac output monitors: A stepwise approach and checklist. Br J Anaesth. 2016;116(6):750–8.

11 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007 Oct 20;370(9596):1453–7.

12 American College of Surgeons Committee on Trauma. Advanced Life Support Course for Physicians. Chicago, IL: American College of Surgeons; 1993.

13 Ait-Oufella H, Lemoinne S, Boelle PY, Galbois A, Baudel JL, Lemant J, et al. Mottling score predicts survival in septic shock. Intensive Care Med. 2011 May;37(5):801–7.

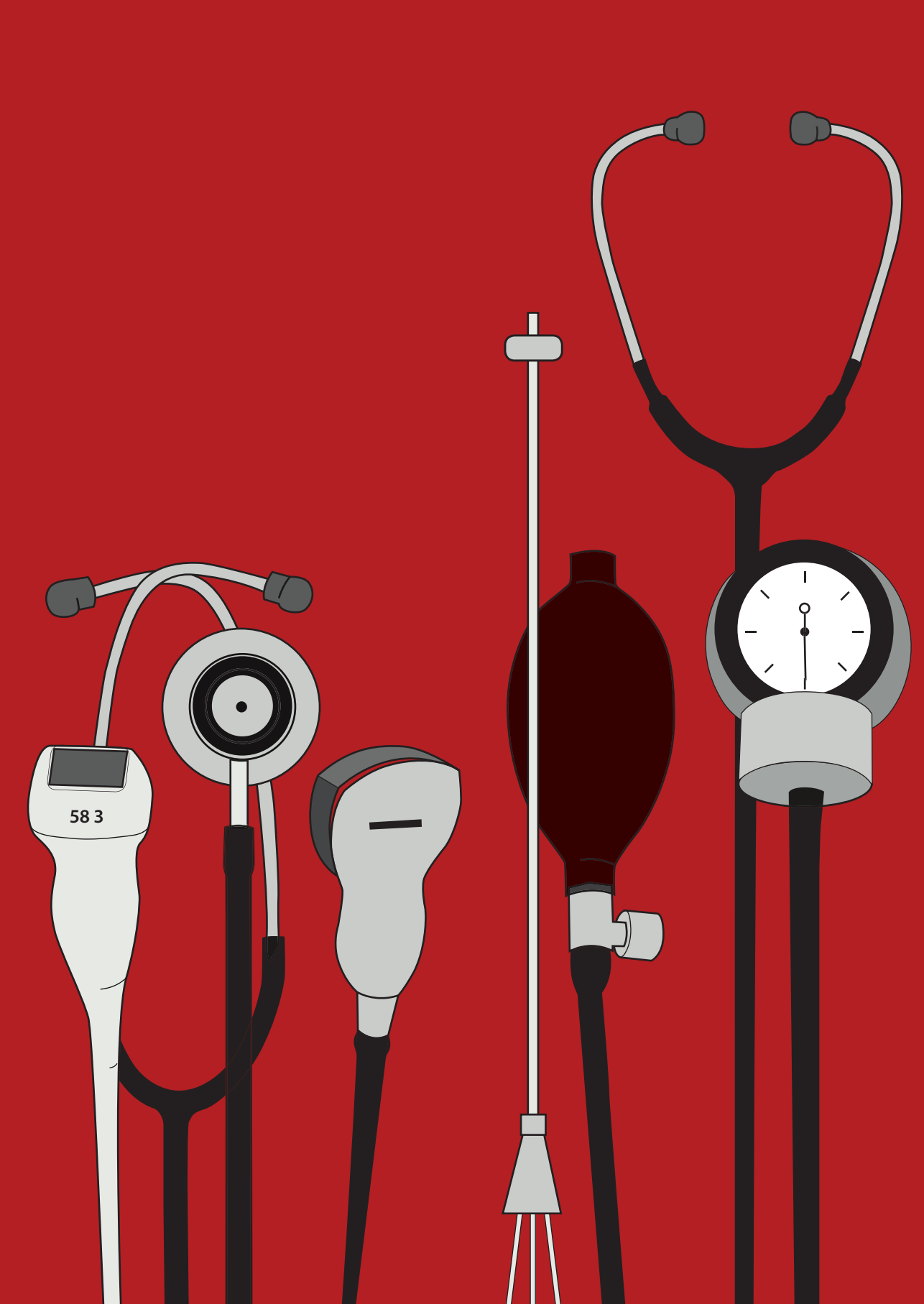
14 Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006 May;34(5):1297–310.

15 Gall JR, Lemeshow S, Saulnier F, Le Gall JR, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. JAMA J Am Med Assoc. 1993 Aug 14;270(24):2957–63.

16 Hiemstra B, Eck RJ, Koster G, Wetterslev J, Perner A, PettilÃ¤ V, et al. Clinical examination, critical care ultrasonography and outcomes in the critically ill: cohort profile of the Simple Intensive Care Studies-I. BMJ Open. 2017 Sep;7(9):e017170.

17 Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, et al. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med. 2019 Feb;45(2):190–200.

18 Lewis JF, Kuo LC, Nelson JG, Limacher MC, Quinones MA. Pulsed Doppler echocardiographic determination of stroke volume and cardiac output: clinical validation of two new methods using the apical window. Circulation. 1984 Sep;70(3):425–31.
19 Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat. 2007;17(4):571–82.
20 Critchley LAH, Critchley JAJH. A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. J Clin Monit Comput. 1999 Feb;15(2):85–91.
21 Saugel B, Grothe O, Wagner JY. Tracking changes in cardiac output: Statistical considerations on the 4-quadrant plot and the polar plot methodology. Anesth Analg. 2015;121(2):514–24.
22 Monnet X, Vaquer S, Anguel N, Jozwiak M, Cipriani F, Richard C, et al. Comparison of pulse contour analysis by Pulsioflex and Vigileo to measure and track changes of cardiac output in critically ill patients. Br J Anaesth. 2015;114(2):235–43.
23 Suehiro K, Tanaka K, Mikawa M, Uchihara Y, Matsuyama T, Matsuura T, et al. Improved Performance of the Fourth-Generation FloTrac/Vigileo System for Tracking Cardiac Output Changes. J Cardiothorac Vasc Anesth. 2015;29(3):656–62.
24 Kusaka Y, Ohchi F, Minami T. Evaluation of the Fourth-Generation FloTrac/Vigileo System in Comparison With the Intermittent Bolus Thermodilution Method in Patients Undergoing Cardiac Surgery. J Cardiothorac Vasc Anesth. 2018 Jun;000:4–11.
25 Maeda T, Hattori K, Sumiyoshi M, Kanazawa H, Ohnishi Y. Accuracy and trending ability of the fourth-generation FloTrac/Vigileo SystemTM in patients undergoing abdominal aortic aneurysm surgery. J Anesth. 2018 Jun 3;32(3):387–93.
26 Teboul JL, Saugel B, Cecconi M, De Backer D, Hofer CK, Monnet X, et al. Less invasive hemodynamic monitoring in critically ill patients. Intensive Care Med. 2016;42(9):1350–9.
27 Wetterslev M, Møller-Sørensen H, Johansen RR, Perner A. Systematic review of cardiac output measurements by echocardiography vs. thermodilution: the techniques are not interchangeable. Intensive Care Med. 2016;42(8):1223–33.
28 Hapfelmeier A, Cecconi M, Saugel B. Cardiac output method comparison studies: the relation of the precision of agreement and the precision of method. J Clin Monit Comput. 2016 Apr;30(2):149–55.
29 Saugel B, Reese PC, Wagner JY, Buerke M, Huber W, Kluge S, et al. Advanced hemodynamic monitoring in intensive care medicine. Medizinische Klin - Intensivmed und Notfallmedizin. 2017;113(3):192–201.
30 Jensen MB, Sloth E, Larsen KM, Schmidt MB. Transthoracic echocardiography for cardiopulmonary monitoring in intensive care. Eur J Anaesthesiol. 2004 Sep;21(9):700–7.


58 3

8

Mortality prediction models in the adult critically ill: A scoping review

Britt E. Keuning
Thomas Kaufmann
Renske Wiersema
Anders Granholm
Ville Pettilä
Morten Hylander Møller
Christian Fynbo Christiansen
José Castela Forte
Harold Snieder
Frederik Keus
Rick G. Pleijhuis
Iwan C.C. van der Horst

Acta Anaesthesiologica Scandinavica

Abstract

Background
Mortality prediction models are applied in the Intensive Care Unit (ICU) to stratify patients into different risk categories and to facilitate benchmarking. To ensure that the correct prediction models are applied for these purposes, the best performing models must be identified. As a first step, we aimed to establish a systematic review of mortality prediction models in critically ill patients.

Methods
Mortality prediction models were searched in four databases using the following criteria: developed for use in adult ICU patients in high-income countries, with mortality as primary or secondary outcome. Characteristics and performance measures of the models were summarized. Performance was presented in terms of discrimination, calibration and overall performance measures presented in the original publication.

Results
In total, 43 mortality prediction models were included in the final analysis. Fifteen models were only internally validated (35%), 13 externally (30%) and 10 (23%) were both internally and externally validated by the original researchers. Discrimination was assessed in 42 models (98%). Commonly used calibration measures were the Hosmer-Lemeshow test (60%) and the calibration plot (28%). Calibration was not assessed in 11 models (26%). Overall performance was assessed in the Brier score (19%) and the Nagelkerke's R^2 (4.7%).

Conclusions
Mortality prediction models have varying methodology, and validation and performance of individual models differ. External validation by the original researchers is often lacking and head-to-head comparisons are urgently needed to identify the best performing mortality prediction models for guiding clinical care and research in different settings and populations.

Introduction

Outcome prediction models, severity scales and risk scores are prognostic tools to estimate the probability for a pre-specified outcome (1). These prognostic tools use variables (e.g., about the severity of illness) to predict outcome, often mortality, in a specific patient population such as the critically ill. In the ICU, mortality prediction models may be applied to stratify patients in different risk categories and to facilitate benchmarking by using standardized mortality rates. An accurate mortality prediction model provides a stratification of the risk of an outcome at a population level. These models generally provide a numerical estimate of that risk based on estimates from previous populations (2). Per definition, all mortality prediction models are best suited for use at a population level and not for individual prognostication, as uncertainty for individual patients remains high (3,4).

Several models are widely known and broadly applied such as the Acute Physiology, and Chronic Health Evaluation (APACHE) I-IV, the Mortality Prediction Model (MPM), and the Simplified Acute Physiology Score (SAPS) I-III (5), whereas others like the Intensive Care National Audit & Research Centre (ICNARC) are used solely in one country (6). Previous literature has only reviewed commonly used models, models with different outcome than mortality or disease- or organ specific prognostic models (3-5,7,8). To the best of our knowledge, no study has systematically assessed which mortality prediction models have been developed and validated for broad cohorts of adult critically ill patients.

Rationale and objective

The objective of this study was to provide an overview of available mortality prediction models in adult critically ill patients as a step-up towards future head-to-head comparison of model performance through systematic external validation.

Methods

Protocol and registration

This scoping review was performed following our protocol (Supplement 1) and was reported in accordance with the PRISMA-ScR checklist (9). Notably, we aimed to publish the protocol on PROSPERO, but during the process it showed that PROSPERO currently does not accept registrations for scoping reviews, literature reviews or mapping reviews.

Search strategy

We conducted a systematic search of MEDLINE, EMBASE, Web of Science and The Cochrane Central Register of Controlled Trials (CENTRAL) to identify relevant ICU mortality prediction models (Supplement 1). Mortality was chosen as the outcome of interest, as prediction models were originally developed to identify patients with high mortality risk. For all databases, except the CENTRAL database, the search period encompassed a period starting from the 1st of January

2008 to the 21st of April 2019. We used snowballing, that is, searching references and related articles, to identify additional prediction models that were published before 2008.

One author ran the search, after which the screening of records and data extraction was performed in duplicate. All records were screened based on title and/or abstract. Papers clearly irrelevant to the purpose were excluded. The remaining articles were screened for eligibility. Consulting a third opinion solved disagreements. More detailed information is presented in the protocol (Supplement 1).

Eligibility criteria

To be considered eligible, mortality prediction models had to meet the following criteria: 1) originally developed specifically for use in adult critically ill patients as defined by the included studies, 2) representing broad groups of ICU patients (with large diversity of admission diagnoses, e.g., non-diabetic patients, medical admissions, surgical admissions, etc.), 3) availability of the original article in English, and 4) mortality at any time as (primary or secondary) outcome of interest.

Prediction models were excluded 1) when developed for low or middle-income countries, as characteristics of ICU patients in these countries often substantially differ from those in high-income countries and, epidemiological data from low-income countries has been frequently unavailable (10,11), 2) when developed as a digital model or derived from a machine-learning algorithm, since code and data availability is not a requirement in all journals. Since our utmost goal is to make a head-to-head comparison of available mortality prediction models using an independent external validation cohort, the code or data necessary to retrieve the underlying prediction model formula are required to reproduce the prediction models. 3) When the development of multiple customized prediction models was described in one article, but no final model was proposed, the prediction models were excluded. Finally, 4) we excluded prediction models specifically developed for subgroups of intensive care patients such as those with sepsis, trauma, cardiac and neurological patients. Studies not specifying inclusion of these subgroups within a wider, general ICU population were considered to be eligible. Prediction models developed in a medical or surgical ICU were included.

Data extraction

If multiple mortality outcomes (e.g., at different time points) were used, we used the primary outcome in the original publication (or the first mortality outcome if the primary outcome was not mortality) to describe the performance of the prediction model.

Details on the development process of the mortality prediction models included were shown, as well as the number of variables included in the prediction models, mortality rate in each development setting and method of handling of missing data. To give an overview of the performance of all mortality prediction models, e.g., values from discrimination, calibration and overall performances measures (12) for mortality were presented for development and internal or external validation cohorts in the original publication (if available).

The discrimination measure presented was the C-statistic (area under the receiver operating characteristic curve (AUROC)), calibration measures presented were goodness-of-fit tests like the Hosmer-Lemeshow (HL) test, calibration plot and calibration slope, and the overall performance measures presented were the Nagelkerke's R^2 and the Brier score (12).

Preferable values from external validation were presented if both internal and external validation values were present in the original publication. If not available, values of internal validation cohorts were presented. External validation was defined as using a separate individual dataset for validation of the mortality prediction model (i.e., no split sampling of a dataset also used for the development of the model).

Citations of original publications were screened for internal and/or external validation articles and shown as being present (+) or absent (-). A list of variables sought for in the identified articles can be found in Supplement 1.

Results

The selection of sources of evidence can be found in the flowchart (Figure 1). Articles evidently developed for specific groups of patients (i.e., sepsis, trauma, cardiac, neurological patients) were excluded based on the title or abstract. Evaluating 99 full-text articles for eligibility resulted in exclusion of another 39 articles, leaving 60 articles that were screened for original publications. Eventually, 43 relevant mortality prediction models reported in 38 publications were extracted and included in the final analysis.

Characteristics of the included mortality prediction models

Characteristics of the mortality prediction models and underlying derivation cohorts are presented in Table 1. Nineteen mortality prediction models (44%) were developed using prospectively collected data specifically gathered for the development of the prediction model (6,13–27), whereas 24 (56%) were developed using either retrospective data (28-44) or prospective data previously collected for other purposes (45-49). The start of data collection for the development cohorts spanned 36 years (1979-2015), and the duration of the cohort studies varying from two months up to 10 years for each cohort. Two mortality prediction models (4.7%) did not report the timespan during which their development cohort was assembled (22,33). Thirty-one mortality prediction models (74%) were developed in a single country (14,18–26,27,29,31,33–44,47,49,50), six (14%) in neighboring countries (two or more) (6,13,28,30,32,46), and five (12%) were developed in multiple countries worldwide (15–17,48). The number of patients included in the development databases ranged from 232 to 731,611 patients with a median of 4,895 (IQR 528 – 35,878). The minimum age at which patients were included was 15 years (2.3%) (35). Eleven mortality prediction models (26%) did not specify age (6,13,23,25,29,31,36,38,42,46). The number of variables included in the mortality prediction models varied from 5 up to 5,695, with a median of 16 (IQR 9 – 24).

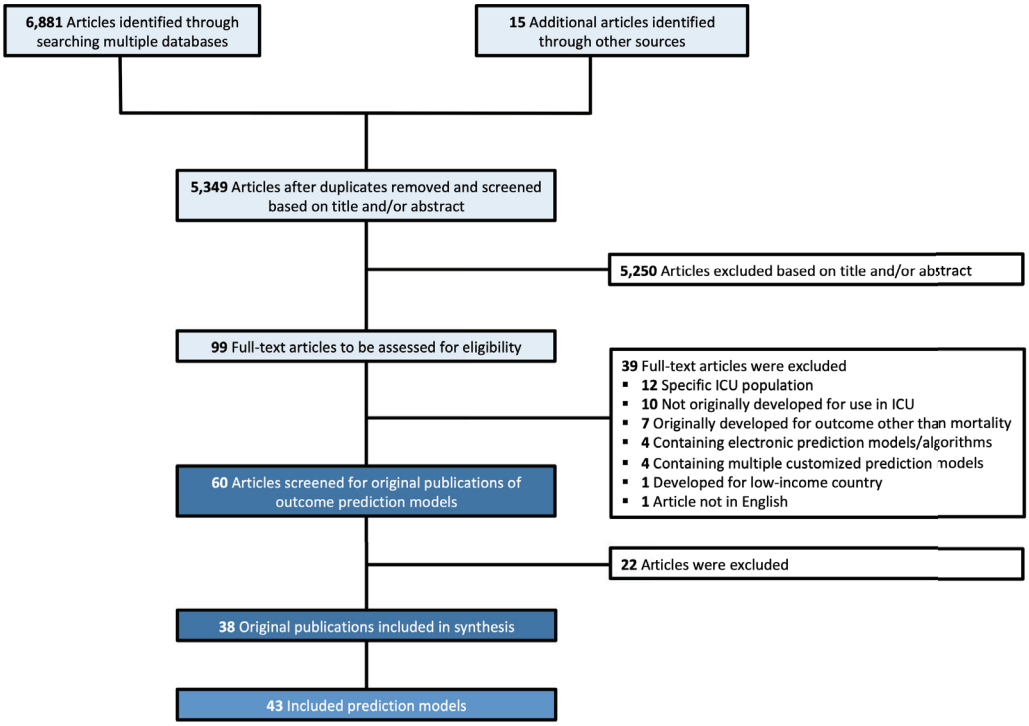


Figure 1. Flow diagram of the search

Table 1. *Characteristics of the development of the 43 mortality prediction models*

Mortality prediction model	Year published	Development database	Cohort assembly period	ICU population	Number of variables*	Outcome		Hospital mortality rate in each development setting	Data collection	Handling of missing data
						Primary	Secondary			
ICNARC[6] Harrison et al.	2007	216,626 Prospective	December 1995 - August 2003	General, adult patients in England, Wales and Ireland	16	Hospital mortality	-	Not reported	Worst values and total urine output in initial 24h in ICU	Exclusion
ICNARC-II[13] Ferrando-Vivas et al.	2017	155,239 Prospective	01/01/2012 - 31/12/2012	General, adult patients in England, Wales and Ireland	23	Hospital mortality	-	32,064/155,239 (20.7%)	Worst values and total urine output in initial 24h in ICU	No missing data
APACHE IV[14] Zimmerman et al.	2006	66,270 Prospective	01/01/2002 - 31/12/2003	General, adult (≥16 yr) patients in the USA	142	Hospital mortality	-	9,013/66,270[#] (13.6%)	Worst values in initial 24h in ICU	Exclusion
SAPS III[15] Moreno et al.	2005	13,428[#] Prospective	14/10/2002 - 15/12/2002	General, adult (≥16 yr) patients worldwide	20	Hospital mortality	-	Not reported	ICU admission ±1h	Imputation of normal values
MPM_{0}-II[16] Lemeshow et al.	1993	12,610 Prospective	17/04/1989 - 31/07/1990 (dataset I) and 30/09/1991 - 27/12/1991 (dataset II)	General, adult (≥18 yr) patients in Europe and the USA	15	Hospital mortality	-	2,632/12,610 (20.9%)	ICU admission	Exclusion
MPM_{24}-II[16,21] Lemeshow et al.	1993	10,357 Prospective	17/04/1989 - 31/07/1990 (dataset I) and 30/09/1991 - 27/12/1991 (dataset II)	General, adult (≥18 yr) patients in Europe and the USA	13	Hospital mortality	-	2,261/10,357 (21.8%)	At 24h in ICU	Exclusion

Mortality prediction model	Year published	Development database	Cohort assembly period	ICU population	Number of variables*	Outcome		Hospital mortality rate in each development setting	Data collection	Handling of missing data
						Primary	Secondary			
SAPS II[17] Le Gall et al.	1993	8,369 Prospective	30/09/1991 - 28/02/1992	General, adult (≥18 yr) patients in Europe and North-America	17	Hospital mortality	-	1,824/8,369[#] (21.8%)	Worst values in initial 24h in ICU	Imputation of normal values
APACHE III[18] Knaus et al.	1991	7,848[#] Prospective	May 1988 - November 1989	General, adult (≥16 yr) patients in the USA	26	Hospital mortality	-	Not reported	Worst values in initial 24h in ICU	Imputation of normal values
APACHE II[19] Knaus et al.	1985	5,030 Prospective	1979 - 1982	General, adult (≥16 yr) patients in the USA	18	Hospital mortality	-	993/5,030 (19.7%)	Worst values in initial 24h in ICU	Exclusion
SUPPORT[20] Knaus et al.	1995	4,301 Prospective	June 1989 - June 1991	General, adult (≥ 18 yr) patients in the USA	15	180-day mortality	-	2,072/4,301 (48.2%)	After 3 days	Imputation of normal values, missing data at day 3 were imputed with day 1 values
MPM_{48}-II[21] Lemeshow et al.	1994	2,049 Prospective	17/04/1989 - 31/07/1990	General, adult (≥18 yr) patients in the USA	13	Hospital mortality	-	307/2,049[#] (15.0%)	At 48h in ICU	Exclusion
MPM_{72}-II[21] Lemeshow et al.	1994	1,497 Prospective	17/04/1989 - 31/07/1990	General, adult (≥18 yr) patients in the USA	13	Hospital mortality	-	418/1,497[#] (27.9%)	At 72h in ICU	Exclusion
TRIOS[22] Timsit et al.	2001	893 Prospective	Not reported (validation dataset in March 1999)	General, adult (≥ 16 yr) patients, hospitalized >48h in France	32	Hospital mortality	-	268/893 (30.0%)	First 3 days in ICU	Imputation of normal values
Mortality Risk Score[23] Dólera-Moreno et al.	2016	844 Prospective	January 2013 - April 2014	General, adult patients in Spain	6	ICU mortality	-	91/844 (10.8%)	ICU admission	Not reported

Mortality prediction model	Year published	Development database	Cohort assembly period	ICU population	Number of variables*	Outcome		Hospital mortality rate in each development setting	Data collection	Handling of missing data
						Primary	Secondary			
Mortality Multifactor Model[24] Li et al.	2017	500 Prospective	01/03/2014 - 30/04/2014	General, adult (≥18 yr) patients in China	36	Hospital mortality	Mortality 30 days after ICU admission, LOS	102/500 (20.4%)	First 24h in ICU	Exclusion
Mortality Prognostic Model[25] Hadique et al.	2017	500 Prospective	November 2013 - April 2014	Medical, adult patients in the USA	44	6-month mortality	-	180/500 (36.0%)	ICU admission, SQ within 12-24h of admission	Not reported
Mortality Prediction Model[26] Fika et al.	2018	400 Prospective	January 2012 - July 2013	General, adult (≥18 yr) patients in Greece	12	ICU mortality	-	131/400 (23.8%)	Worst values in initial 24h in ICU	Not reported
APACHE II-APM[27] Nematifard et al.	2018	304 Prospective	June 2014 - November 2016	General, adult (≥16 yr) patients in Iran	19	Hospital mortality	-	96/304 (31.6%)	Worst values in initial 24h in ICU	Exclusion
APACHE III-APM[27] Nematifard et al.	2018	304 Prospective	June 2014 - November 2016	General, adult (≥16 yr) patients in Iran	27	Hospital mortality	-	96/304 (31.6%)	Worst values in initial 24h in ICU	Exclusion
ANZROD0[28] Paul et al.	2017	731,611 Retrospective	01/01/2006 - 31/12/2015	General, adult (≥16 yr) patients in Australia and New Zealand	11	Hospital mortality	-	69,503/731,611[#] (9.5%)	ICU admission	Exclusion
MMI[29] Min et al.	2017	354,154[#] Retrospective	January 2003 - December 2013	Medical, veteran ICU patients in the USA	5,695	All-cause mortality at 6- and 12-months post-hospital discharge	Hospital mortality	Not reported	Worst values of 24h before and 24h after admission	Imputation of mean values
ANZROD[30] Paul et al.	2013	304,149 Retrospective	01/01/2004 - 31/12/2009	General, adult (≥16 yr) patients in Australia and New Zealand	38	Hospital mortality	-	34,369[#] (11.3%)	Worst values in initial 24h in ICU	Exclusion

Mortality prediction model	Year published	Development database	Cohort assembly period	ICU population	Number of variables*	Outcome		Hospital mortality rate in each development setting	Data collection	Handling of missing data
						Primary	Secondary			
Customized APACHE IV[31] Brinkman et al.	2013	77,616 Retrospective	01/01/2008 - 01/07/2011	Non-CABG, adult critically ill patients in the Netherlands	142	Hospital mortality	Mortality at 1, 3 and 6 months after ICU admission	12,186/77,616# (15.7%)	First 24h in ICU	Not reported
MPM_0-III[32] Higgins et al.	2005	74,578 Retrospective	October 2001 - March 2004	General, adult (≥18 yr) patients in the USA, Canada and Brazil	16	Hospital mortality	-	10,292/74,578 (13.8%)	ICU admission	Exclusion
NQF-ICOMmort[33] Philip R. Lee Institute	2016	40,395 Retrospective	Not reported	General, adult (≥18 yr) patients in the USA	17	Hospital mortality	-	Not reported	1h prior to ICU admission to 1h after admission	Not reported
OASIS[34] Johnson et al.	2013	39,070 Retrospective	01/01/2007 - 15/09/2011	General, adult (≥16 yr) patients in the USA	10	ICU mortality	Hospital mortality	4,571/39,070# (11.7%)	Worst values and total urine output in initial 24h in ICU	Exclusion
COPE-4[35] Duke et al.	2013	35,878 Retrospective	01/07/2004 - 30/06/2006	General, adult (≥15 yr) patients in Australia	6	Hospital mortality	-	4415/35,878 (12.3%)	ICU admission (mechanical ventilation *during* ICU admission)	No missing data
RDW-SAPS[50] Hunziker et al.	2012	17,922 Retrospective	January 2001 - December 2008	General, adult (≥18 yr) patients in the USA	15	Hospital mortality	ICU mortality, 1-year mortality	2,007/17,922# (11.2%)	ICU admission	Not reported
COPE[36] Duke et al.	2008	17,880 Retrospective	01/07/2004 - 30/06/2005	General, adult patients in Australia	5	Hospital mortality	-	2,186/17,880 (12.1%)	ICU admission (mechanical ventilation *during* ICU admission)	No missing data
PREDICT[37] Ho et al.	2008	11,930 Retrospective	1989 - 2002	General, adult (≥16 yr) patients in Australia	6	15-year mortality	-	829/11,930# (6.9%)	First 5 days in ICU	No missing data
High-Risk Selection System[46] Iapichino et al.	2006	8,248 Retrospective	October 1994 – February 1995	General, adult patients (>24h in ICU) in Europe	16	Hospital mortality	-	1,617/8,248# (19.6%)	ICU admission	Not reported
GV-SAPS II[47] Liu et al.	2016	4,895 Retrospective	2001 - 2008	Non-diabetic, adult (≥18 yr) patients in the USA	20	30-day mortality	9-month mortality	649/4,895 (13.3%)	First 24h in ICU	When >5% exclusion, <5% not reported

Mortality prediction model	Year published	Development database	Cohort assembly period	ICU population	Number of variables*	Outcome		Hospital mortality rate in each development setting	Data collection	Handling of missing data
						Primary	Secondary			
MODS/NEMS[38] Kao et al.	2016	4,321 Retrospective	01/01/2009 - 30/11/2012	General, adult patients in Canada	32	Hospital mortality	-	986/4,321 (22.8%)	First 24h in ICU	Not reported
SMS-ICU[48] Granholm et al.	2018	4,086 Retrospective	23/12/2009 - 30/06/2016	General, adult (≥18 yr), acutely admitted patients worldwide	7	90-day mortality	-	1,403/4,086 (34.3%)	Worst values in initial 24h in ICU	Multiple imputations, exclusion when >25%
P- model[39] Umegaki et al.	2010	3,505 Retrospective	01/01/2007 - 31/12/2007	General, adult (≥20 yr) patients in Japan	10	Mortality at 28 days after the first ICU day	-	336/3,505[#] (9.6%)	First 24h in ICU	Not reported
BCV model[40] Huang et al.	2013	1,624 Retrospective	01/01/2006 - 01/12/2008	General, adult (≥18 yr) patients in Taiwan	6	Daily probability of mortality from day 3 to day 28 post ICU admission	-	Not reported	Daily complete blood count	Exclusion
BCV/APACHE II model[40] Huang et al.	2013	1,624 Retrospective	01/01/2006 - 01/12/2008	General, adult (≥18 yr) patients in Taiwan	24	Daily probability of mortality from day 3 to day 28 post ICU admission	-	Not reported	Daily complete blood count, APACHE II score in the first 24h in ICU	Exclusion
CREEK[41] Stachon et al.	2008	528 Retrospective	April 2003 - January 2004	Medical, adult (≥18 yr) patients in Germany	8	Hospital mortality	-	87/528 (16.5%)	ICU admission	Not reported
SAPS-R[42] Viviand et al.	1991	351 Retrospective	01/01/1986 - 31/10/1988	General, adult patients in France	5	Hospital mortality	-	Not reported	Worst values in initial 24h in ICU	Exclusion

Mortality prediction model	Year published	Development database	Cohort assembly period	ICU population	Number of variables*	Outcome		Hospital mortality rate in each development setting	Data collection	Handling of missing data
						Primary	Secondary			
SAPS-E[42] Viviand et al.	1991	351 Retrospective	01/01/1986 - 31/10/1988	General, adult patients in France	7	Hospital mortality	-	Not reported	Worst values in initial 24h in ICU	Exclusion
25OHD Deyo-Charlson Comorbidity Index[49] Mahato et al.	2016	310 Retrospective	01/06/2012 - 30/05/2015	General, adult (≥18 yr) patients in the USA	18	90-day mortality after ICU admission	-	59/310 (19.0%)	First 24h in ICU	Not reported
DELAWARE[43] Stachon et al.	2008	271 Retrospective	April 2003 – January 2004	Surgical, adult (≥18 yr) patients in Germany	9	Hospital mortality	-	67/271 (24.7%)	ICU admission	Exclusion
Simplified Mortality Score[44] Goag et al.	2018	232 Retrospective	June 2015 - February 2016	Medical, adult (≥18 yr) patients in Korea	8	28-day mortality	-	72/232[#] (31.1%)	Within 24h of ICU admission	Exclusion

** When (parts of) other mortality prediction models were used as variables in a mortality prediction model (e.g., the Charlson Comorbidity Index and APACHE III as variable in the Mortality Prognostic Model), variables included in these specific mortality prediction models were also taken into account.*

 # Estimated based on information in original publication.

Abbreviations: ANZROD, Australian and New Zealand Risk Of Death; APACHE, Acute Physiology and Chronic Health Evaluation; APM, adductor pollicis muscle; BCV, Blood Cell Variability; COPE, Critical care Outcome Prediction Equation; CREEK, Critical Risk Evaluation by Early Keys; DELAWARE, Dense Laboratory Whole Blood Applied Risk Estimation; GV, glucose variability; ICNARC, Intensive Care National Audit Research Centre; ICU, intensive care unit; LOS, length of stay; MMI, Multi-morbidity Index; MODS, Multiple Organs Dysfunctional Score; MPM; mortality prediction model; NEMS, Nine Equivalents Nursing Manpower use Score; NQF-ICOMmort, National Quality Forum ICU outcomes model (mortality); OASIS, Oxford Acute Severity of Illness Score; PREDICT, Predicted Risk, Existing Diseases and Intensive Care Therapy; RDW, red cell distribution width; SAPS, Simplified Acute Physiology Score; SMS-ICU, Simplified Mortality Score for the Intensive Care Unit; SQ, surprise question; SUPPORT, Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments; TRIOS, Three day Recalibrating ICU Outcomes.

Outcome measures

The timing of mortality outcome varied between the studies. Hospital mortality was the most frequently used primary outcome in 29 (67%) mortality prediction models (6,13-19,21,22,24,27,28,30-33,35,36,38,41-43,45,46). Other primary outcome variables were ICU mortality (7%) (23,26,34) 28-day mortality (4.7%) (39,44), 90-day mortality (4.7%) (48,49), 3 to 28-day mortality (4.7%) (40), 30-day mortality (2.3%) (47), 180-day mortality (2.3%) (20), 6-month mortality (2.3%) (25), 15-year mortality (2.3%) (37), and 6 and 12-month mortality (2.3%) (29).

Secondary outcomes were 1-month mortality after ICU admission (4.7%),(24,31) hospital mortality (4.7%),(29,34) ICU mortality (2.3%),(50) 3-month mortality after ICU admission (2.3%),(31) 6-month mortality after ICU admission (2.3%),(31) 9-month mortality (2.3%),(47) 1-year mortality (2.3%),(50) and length of stay (2.3%).(24) Of the 43, thirty-seven mortality prediction models (86%) did not prognosticate any secondary outcome (6,13-23,25-28,30,32,33,35-44,46,48,49).

Hospital mortality rates of the development cohorts varied from 6.9% to 48% and were not reported for nine mortality prediction models (21%) (6,15,18,29,33,40,42).

For 21 mortality prediction models (49% of 43), data were collected within the first 24 hours after patient admission to the ICU (6,13,14,17-19,24,26,27,30,31,34,38,39,42,44,47-49). For 11 prediction models (26%) data on ICU admission were collected (16,23,25,28,32,35,36,41,43,45,46) whereas for the remaining prediction models data timing varied from 24 days before admission up to five days after patient admission to the ICU.

Handling of missing data was not reported in 11 mortality prediction models (26%) (23,25,26, 31,33,38,39,41,45,46,49), 20 prediction models (47% of 43) excluded records with missing data (6,14,16,19,21,24,27,28,30,32,34,40,42-44), six prediction models (14%) imputed values with normal or mean values (15,17,18,20,22,29), and four prediction models (9.3%) reported no missing data (13,35–37). The remaining two prediction models (4.7%) excluded patients when more than a certain percentage of the data was missing (>5% or >25%) (47,48).

Discrimination, calibration and overall performance measures

Discrimination, calibration and overall performance measures are presented in Table 2. Of the 43 mortality prediction models, 15 (35%) were only internally validated (23,26,28-31,33,38-41,44,46,48), 13 (30%) only externally (16,19-21,25,35,36,42,43,47), 10 (23%) were both internally and externally validated (6,13-15,17,18,22,32,34,37), and 5 prediction models (12%) were not validated at all (24,27,45,49). Fifteen prediction models (35%) included a description of an external validation in their original publication (13,16,20-22,25,34-36,42,43,47).

Discrimination was expressed as the area under the receiver operating characteristics curve (AUROC) in 42 of the 43 mortality prediction models original publications (98%). Only the APACHE II model didn't report an AUROC value in the original publication (19). In the development cohorts, the lowest discrimination was AUROC 0.72 (95% CI 0.71-0.74) (48), and the highest AUROC 0.91

(95% CI not specified) (30). In the validation cohorts, the lowest AUROC was 0.58 (95% CI not specified) (44), and the highest AUROC 0.95 (0.91-0.99) (23).

Calibration measures were expressed by various statistical measures. The Hosmer-Lemeshow goodness-of-fit test was used in 26 mortality prediction models (60%) (14-17,21,22,24-26,28,30,32-36,38-41,43,46,48). Calibration plot was expressed for 12 prediction models (28%) (13,15,20,24,28,30,31,33,35,37,43,48), and two prediction models (4.7%) presented the calibration slope value (30,48). Finally, one prediction model (2.3%) used the likelihood ratio test chi-squared value (23), and one prediction model (2.3%) used the Quasi likelihood under the Independence Criterion (QIC) (45). In eleven prediction models (26%), calibration was not assessed (6,18,19,27,29,42,44,47,49).

Overall performance was expressed as the Brier score in eight mortality prediction models (19%) 6,13,28-31,34,41), and as Nagelkerke's R^2 in two prediction models (4.7%) (37,48).

Table 2. *Performance of the 43 mortality prediction models*

Mortality prediction model	Validated?*		AUROC (95% CI) Development cohort#	Calibration Development cohort#	Overall performance Development cohort#	Type of validation cohort in original publication	AUROC (95% CI) Validation cohort	Calibration Validation cohort	Overall performance Validation cohort
	Internally	Externally							
ICNARC[6] Harrison et al.	+ *data splitting*	+	-	-	-	Internal validation dataset	0.87 (n.s.)	-	Brier score: 0.132
ICNARC-II[13] Ferrando-Vivas et al.	+ *bootstrapping*	+ *original publication*	0.89 (0.89-0.89)	-	Brier score: 0.103	External validation dataset	0.89 (0.88-0.89)	Calibration plot present	Brier score: 0.108
APACHE IV[14] Zimmerman et al.	+ *data splitting*	+	-	-	-	External validation dataset	0.88 (n.s.)	HL X^2: 16.8 (P = 0.08)	-
SAPS III[15] Moreno et al.	+ *cross-validation*	+	-	-	-	Internal validation dataset	0.85 (n.s.)	HL H-statistic: 10.6 (P = 0.39) HL C-statistic: 14.3 (P = 0.16) Calibration plot present	-
MPM_0-II[16] Lemeshow et al.	-	+ *original publication*	0.84 (n.s.)	HL C-statistic: 6.2 (P = 0.62)	-	External validation dataset	0.82 (n.s.)	HL C-statistic: n.s. (P = 0.33)	-
MPM_{24}-II[16,21] Lemeshow et al.	-	+ *original publication*	0.84 (n.s.)	HL C-statistic: 4.9 (P = 0.76)	-	External validation dataset	0.84 (n.s.)	HL C statistic: 12.9 (P = 0.23)	-
SAPS II[17] Le Gall et al.	+ *data splitting*	+	0.88 (0.87-0.90)	HL H-statistic: 3.70 (P = 0.88)	-	Internal validation dataset	0.86 (0.84-0.88)	HL H statistic: n.s. (P = 0.10)	-
APACHE III[18] Knaus et al.	+ *data splitting*	+	-	-	-	Internal validation dataset	0.90 (n.s.)†	-	-
APACHE II[19] Knaus et al.	-	+	-	-	-	-	-	-	-
SUPPORT[20] Knaus et al.	-	+ *original publication*	0.79 (n.s.)	-	-	External validation dataset	0.78 (n.s.)	Calibration plot present	-
MPM_{48}-II[21] Lemeshow et al.	-	+ *original publication*	0.81 (n.s.)	HL C-statistic: 11.7 (P = 0.31)	-	External validation dataset	0.80 (n.s.)	HL C statistic: 8.4 (P = 0.59)	-
MPM_{72}-II[21] Lemeshow et al.	-	+ *original publication*	0.79 (n.s.)	HL C-statistic: 11.6 (P = 0.31)	-	External validation dataset	0.75 (n.s.)	HL C statistic: 10.4 (P = 0.41)	-

Mortality prediction model	Validated?*		AUROC (95% CI) Development cohort#	Calibration Development cohort#	Overall performance Development cohort#	Type of validation cohort in original publication	AUROC (95% CI) Validation cohort	Calibration Validation cohort	Overall performance Validation cohort
	Internally	Externally							
TRIOS[22] Timsit et al.	+ *bootstrapping*	+ *original publication*	0.79 (0.77-0.82)	HL C-statistic: 5.6 (P = 0.70)	-	External validation dataset	0.83 (0.78-0.87)	HL C statistic: 7.1 (P = 0.55)	-
Mortality Risk Score[23] Dólera-Moreno et al.	+ *data splitting*	-	-	-	-	Internal validation dataset	0.95 (0.91-0.99)	Likelihood ratio test X^2: 296.8†	-
Mortality Multifactor Model[24] Li et al.	-	-	0.84 (0.80-0.87)	HL X^2: 12.3 (P = 0.14) Calibration plot present	-	-	-	-	-
Mortality Prognostic Model[25] Hadique et al.	-	+ *original publication*	0.83 (0.80-0.87)	HL statistic: 6.5 (P = 0.59)	-	External validation dataset	0.84 (0.81-0.88)	HL statistic: 9.2 (P = 0.33)	-
Mortality Prediction Model[26] Fika et al.	+ *data splitting*	-	-	-	-	Internal validation dataset	0.85 (0.73-0.97)	HL X^2: 4.9 (P = 0.77)	-
APACHE II-APM[27] Nematifard et al.	-	-	0.85 (0.81-0.90)	-	-	-	-	-	-
APACHE III-APM[27] Nematifard et al.	-	-	0.87 (0.82-0.91)	-	-	-	-	-	-
ANZROD0[28] Paul et al.	+ *data splitting*	-	0.85 (0.85-0.86)	HL C-statistic: 459.3	Brier score: 0.069 Adjusted Brier score: 0.196	Internal validation dataset	0.85 (0.85-0.85)	HL C-statistic: 264.9 Calibration plot present	Brier score: 0.069 Adjusted Brier score: 0.190
MMI[29] Min et al.	+ *data splitting*	-	-	-	-	Internal validation dataset	*6-month mortality*: 0.86 (0.85-0.86) *12-month mortality*: 0.84 (0.83-0.84)	-	Brier score: 0.21†‡

Mortality prediction model	Validated?*		AUROC (95% CI) Development cohort#	Calibration Development cohort#	Overall performance Development cohort#	Type of validation cohort in original publication	AUROC (95% CI) Validation cohort	Calibration Validation cohort	Overall performance Validation cohort
	Internally	Externally							
ANZROD[30] Paul et al.	+ *data splitting*	-	0.91 (n.s.)	HL C-statistic: 189.5 HL H-statistic: 174.1 Cox calibration regression slope: 1	Brier score: 0.065	Internal validation dataset	0.90 (n.s.)	HL C-statistic: 104.9 HL H-statistic: 111.4 Cox calibration regression slope: 0.98 Calibration plot present	Brier score: 0.066
Customized APACHE IV[31] Brinkman et al.	+ *bootstrapping*	-	0.88 (0.88-0.88)	Calibration plot present	Brier score: 0.09	Internal validation dataset	-	-	-
MPM_0-II[32] Higgins et al.	+ *data splitting*	+	0.83 (0.82-0.83)	HL statistic: 11.5 (P = 0.17)	-	Internal validation dataset	0.82 (0.82-0.83)	HL statistic: 11.6 (P = 0.31)	-
NQF-ICOMmort[33] Philip R. Lee Institute	+ *data splitting*	-	-	-	-	Internal validation dataset	0.82 (0.81-0.83)	HL C statistic: 12.0 (P = 0.28) HL H statistic: 16.9 (P = 0.08) Calibration plot present	-
OASIS[34] Johnson et al.	+ *data splitting*	+ *original publication*	-	-	-	External validation dataset	0.90 (P < 0.0003)§	HL X^2: 19.6§	Brier score: 0.048§
COPE-4[35] Duke et al.	-	+ *original publication*	-	-	-	External validation dataset	- (0.82-0.83)	HL H-statistic: 14.8 (P = 0.06) Correlation of calibration plot R^2: 0.99 Calibration plot present	-
RDW-SAPS[50] Hunziker et al.	-	-	0.77 (n.s.)	Quasi Likelihood under the Independence model Criterion (QIC) X^2: 1.83	-	-	-	-	-
COPE[36] Duke et al.	-	+ *original publication*	- (0.83-0.84)	HL X^2: 23.1 (P < 0.01)	-	External validation dataset	- (0.83-0.84)	HL X^2: 26.9 (P < 0.01)	-
PREDICT[37] Ho et al.	+ *bootstrapping*	+	-	-	-	Internal validation dataset	0.76 (0.75-0.77)	Calibration plot present	Nagelkerke's R^2: 0.255
High-Risk Selection System[46] Iapichino et al.	+ *data splitting*	-	0.81 (n.s.)	HL X^2: n.s. (P = 0.21)	-	Internal validation dataset	0.81 (n.s.)	HL X^2: n.s. (P = 0.22)	-

Mortality prediction model	Validated?*		AUROC (95% CI) Development cohort#	Calibration Development cohort#	Overall performance Development cohort#	Type of validation cohort in original publication	AUROC (95% CI) Validation cohort	Calibration Validation cohort	Overall performance Validation cohort
	Internally	Externally							
GV-SAPS II[47] Liu et al.	-	+ *original publication*	0.83 (0.81-0.84)	-	-	External validation dataset	0.82 (0.81-0.83)	-	-
MODS/NEMS[38] Kao et al.	+ *bootstrapping*	-	0.79 (n.s.)	-	-	Internal validation dataset	0.76 (n.s.)	HL X^2: 5.48 (P = 0.32)†	-
SMS-ICU[48] Granholm et al.	+ *bootstrapping*	+	0.72 (0.71-0.74)	HL X^2: 9.0 (P = 0.34)† Calibration slope: 0.99 Calibration plot present	Nagelkerke's R^2: 0.191	Internal validation dataset	0.73 (n.s.)	Calibration slope: 0.99 Calibration plot present	Nagelkerke's R^2: 0.193
P- model[39] Umegaki et al.	+ *cross-validation*	-	0.87 (0.85-0.90)	HL X^2: 14.5 (P = 0.07)	-	Internal validation dataset	0.90 (0.88-0.92)	HL X^2: 13.5 (P = 0.10)	-
BCV model[40] Huang et al.	+ *data splitting*	-	0.79 (0.76-0.81)	HL X^2: 8.7 (P = 0.37)	-	Internal validation dataset	0.76 (0.71-0.81)	HL X^2: 11.1 (P = 0.19)	-
BCV/APACHE II model[40] Huang et al.	+ *data splitting*	-	0.80 (0.78-0.83)	HL X^2: 6.2 (P = 0.63)	-	Internal validation dataset	0.78 (0.73-0.83)	HL X^2: 5.4 (P = 0.72)	-
CREEK[41] Stachon et al.	+ *cross-validation*	-	0.86 (n.s.)	HL C-statistic: 10.7 (P = 0.22) HL H-statistic: 10.1 (P = 0.26)	Brier score: 0.096	Internal validation dataset	0.832 (n.s.)	-	-
SAPS-R[42] Viviand et al.	-	+ *original publication*	-	-	-	External validation dataset	0.76 (n.s.)	-	-
SAPS-E[42] Viviand et al.	-	+ *original publication*	-	-	-	External validation dataset	0.79 (n.s.)	-	-
25OHD Deyo-Charlson Comorbidity Index[49] Mahato et al.	-	-	0.75 (0.67-0.83)	-	-	-	-	-	-
DELAWARE[43] Stachon et al.	-	+ *original publication*	0.86 (0.80-0.91)	HL statistic: n.s. (P = 0.28) Calibration plot present	-	External validation dataset	0.81 (0.75-0.87)	HL statistic: 0.44 (P = n.s.) Calibration plot present	-

Mortality prediction model	Validated?*		AUROC (95% CI) Development cohort#	Calibration Development cohort#	Overall performance Development cohort#	Type of validation cohort in original publication	AUROC (95% CI) Validation cohort	Calibration Validation cohort	Overall performance Validation cohort
	Internally	Externally							
Simplified Mortality Score[44] Goag et al.	+ *data splitting*	-	-	-	-	Internal validation dataset	0.58 (n.s.)	-	-

** Citations of original publications were screened on internal and/or external validation articles and shown as being present (+) or not present (-). When internal validation was present, the method of internal validation used in the original publication was presented. When external validation in the original publication was present, original publication was added in the column*

 # Development cohort indicates the cohort in whom the prediction model was developed, sometimes also referred to as training cohort † Not clear whether the value was derived from the development or validation dataset in the original publication, or value was derived from the development and validation dataset together

‡ Not clear whether this value is calculated for the 6-month mortality outcome or 12-month mortality.

§ Not clear whether the value was derived from the internal or external validation dataset in the original publication

Abbreviations: ANZROD, Australian and New Zealand Risk Of Death; APACHE, Acute Physiology and Chronic Health Evaluation; APM, adductor pollicis muscle; AUROC; area under the receiving operating curves; BCV, Blood Cell Variability; CI, confidence interval; COPE, Critical care Outcome Prediction Equation; CREEK, Critical Risk Evaluation by Early Keys; DELAWARE, Dense Laboratory Whole Blood Applied Risk Estimation; GV, glucose variability; HL, Hosmer-Lemeshow; ICNARC, Intensive Care National Audit Research Centre; ICU, intensive care unit; MMI, Multi-morbidity Index; MODS, Multiple Organs Dysfunctional Score; MPM; mortality prediction model; NEMS, Nine Equivalents Nursing Manpower use Score; NQF-ICOMmort, National Quality Forum ICU outcomes model (mortality); n.s., not specified; OASIS, Oxford Acute Severity of Illness Score; PREDICT, Predicted Risk, Existing Diseases and Intensive Care Therapy; RDW, red cell distribution width; SAPS, Simplified Acute Physiology Score; SMS-ICU, Simplified Mortality Score for the Intensive Care Unit; SUPPORT, Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments; TRIOS, Three day Recalibrating ICU Outcomes.

Discussion

Main findings

In this scoping review, we presented a contemporary overview of 43 mortality prediction models used in adult ICU patients in high-income countries. We found varying methodology, and the validation and performance of individual prediction models differ. Only 23 mortality prediction models of the 43 (53%) were externally validated. This overview provides a basis for head-to-head comparison of existing mortality prediction models through systematic external validation, with the ultimate goal to identify the most suitable prediction model for a certain cohort of patients.

Summary of evidence

In previous literature, the maximum number of ICU mortality prediction models reviewed was 12 (7), which is considerably less than the 43 prediction models identified by this review. Where we included all developed prediction models specifically designed to assess mortality, other reviews regarding ICU mortality prediction models focused mainly on commonly used models like the APACHE, SAPS and MPM (3–5), or identified models with different outcome than mortality (e.g., organ dysfunction) or disease- or organ specific prognostic models (4,5,7,8). Additionally, only Siontis et al. and Strand et al. applied a systematic search to identify the models and discussed the validation of the models (5,8). Where we included all developed mortality prediction models, Strand et al. did only include prediction models when the search for the specific scoring system yielded more than 50 citations (5). Siontis et al. conducted an evaluation of validated tools for hospitalized patients to predict all-cause mortality. However, their analysis included specific patient groups (e.g., heart or liver patients) rather than general ICU patients as included in the current review (8).

Model performance is affected by the choice of outcome (31,51). Most mortality prediction models used hospital mortality as outcome measure (6,13-19,21,22,24,27,28,30-33,35,36,38,41-43,45,46). In general, longer fixed-time outcome measures used in some models (20,24,25,29,31,37,39,40,44,45,47-49) are currently recommended (51). To elaborate, hospital mortality is dependent on discharge practices and availability of post-ICU care, and is therefore a subjective measure. Furthermore, critical illness affects patients after hospital discharge.

The time span during which the mortality prediction models gathered their data varied from short (e.g., upon ICU admission or during the first initial hour of admission to the ICU) to long (e.g., during the first 24 hours of admission). Concerning complexity (time consumption) and missing data problems, it may be better in some situations to use a simpler model with less missing data than a more complex model built from a dataset with more missing data which achieves a slightly better performance (52). Longer collection periods may lead to more complete data, as incompleteness is often substantial for biochemical variables for patients with short-duration admissions (i.e., less than 24 hours). However, sampling rate affects predictions (53). This limitation is considered less important in models with shorter data collection. Similarly, the treatments administered during the first 24 hours in the ICU obviously also affect predictions.

Comparison of performance

We reported the performance of mortality prediction models in terms of discrimination, calibration and overall performance values. Direct comparison of prediction models predictive performances is not possible, as the development cohorts differed substantially from one another. As a consequence, prediction models cannot be considered interchangeable. Comparisons that are not done head-to-head in external samples independent of all models developed are at high risk of being misleading and may lead to inappropriate conclusions and resource use (12).

Of 43, 26 (60%) mortality prediction models used the HL goodness-of-fit test for calibration (14-17,21,22,24-26,28,30,32-36,38-41,43,46,48). The HL test is commonly used, despite being frequently non-significant for small data cohorts and nearly always significant for large data cohorts (54–57). When only the HL test is reported without any calibration plot or table comparing predicted and observed outcome frequencies, inadequate information regarding calibration is provided (1).

Many ICU mortality prediction models are available and comparatively assessing their performance is a crucial task (4). Twenty-five articles compared the performance of the new model with existing models but used the same cohort of patients that was used in the development of the "novel" model (6,13,14,16-18,20,22,24,26-30,32,34,40-47,49). This methodology is inherently biased in favor of the "novel" model (54,57). Comparisons between prediction models should therefore only be executed in independent external validation samples not used to develop any of the models.

Machine-learning algorithms

Mortality prediction models developed as an electronic model or derived from a machine-learning algorithm such as AutoTriage (58) were excluded in our manuscript since code and data availability is not a requirement in all journals and this is necessary to reproduce the specific prediction model. However, code availability appears to be a rising trend (59). Machine learning-based prediction models seem to achieve increasingly higher accuracies and are becoming more dynamic (60), although they still have to include a sufficiently large development and validation cohort to adequately assess performance and the risk of overfitting. However, a recent systematic review concluded that machine learning did not have superior performance over logistic regression for clinical prediction models (61).

The association between mortality and variables may have changed since the original mortality prediction models were developed, e.g., as a result of advancements in diagnostics and therapeutics (62). Mortality alone however is rarely the only outcome measure for interventional studies in ICU patients, and many trials, especially in sepsis, include an organ dysfunction score as part of on-going patient assessment so that effects on morbidity can also be evaluated (3).

Misuse of mortality prediction models can lead to inappropriate use of resources and potentially even mismanagement of patient care due to incorrect stratification (57). Awareness of the differences in model design, the variance of predictions across different ICU settings, and the effect of heterogeneity in populations are of utmost importance.

Limitations

Some limitations of this study need to be addressed. First, having restricted our search to the period from 2008, relevant mortality prediction models might have been overlooked. Even though some of the most widely used mortality prediction models precede the screening period we identified 16 prediction models that were published before 2008, but optimally searches have no time limit (63). Second, we only included mortality prediction models originally developed for use in the ICU. Mortality prediction models not originally developed for mortality prediction in the ICU could still be valuable clinically. Third, in some original publications, it was unclear whether the presented discrimination, calibration and/or overall performance values were derived from the development cohort or from the validation dataset. We aimed to clarify these, but certain values might reflect another dataset from the original publication. Fourth, we only provided a systematic overview of all developed mortality prediction models in the adult critically ill patients. We did not perform a systematic review of every retrieved model complete with all consecutive internal and external validations, as results from different external validations in different cohorts are not directly comparable due to differences in populations, case-mix and settings. We restricted the scope of this review to only identify whether internal or external validation had been performed as a measure of thoroughness of development of the identified models. For this reason, only screening of citations of the original articles was done to identify internal and/or external validation articles. Therefore, we should address that our assessment on mortality prediction models not being internally and/or externally validated might be incomplete if validation in different publications was missed. A systematic search specifically designed for retrieving validation papers is advised when systematically reviewing the internal and external validations of mortality prediction models (64).

Unanswered research questions

Although we retrieved many developed mortality prediction models that can be used as a step towards future head-to-head comparison, with the results of this scoping review it is not possible to make a recommendation on what mortality prediction models to use and it was not our intention to do so. External validation involving direct head-to-head comparisons in independent cohorts is needed to unravel the comparable performance of individual models. Although we provide a systematic overview of mortality prediction models and describe whether these were internally and/or externally validated, it was not desirable to give an overview of all external validations of the prediction models since this would require a specific search strategy for each model. Moreover, we would have liked to assess risk of bias using the recently developed PROBAST score (1). However, this was not feasible because of the amount of prediction models.

Future perspectives

To identify the most suitable mortality prediction model for a certain patient cohort, ideally a head-to-head comparison of available models should be performed through systematic external validation using prospectively obtained datasets and appropriate statistical methods. The eventual aim will be to use this review to identify, update, and implement the best performing mortality prediction models in daily practice. We are in the process of validating the found

prediction models in independent contemporary cohorts to provide external validation of these models. Second, the process should be performed in different cohorts as heterogeneity of ICU patients exists on multiple levels, i.e., patient level, hospital level, region and country level (65). The best mortality prediction model in one setting is not necessarily the best performing prediction model in another setting. Third, it is worth mentioning that ICU patients have reduced long-term survival and impaired quality of life after ICU discharge compared to the general population (66). Future research should also look at determinants of poor outcomes in ICU survivors to help guide long-term follow-up (67).

Conclusions

In this review, 43 mortality prediction models have been studied. The validation and performance of individual prediction models differ and the best prediction models for guiding clinical care and research is still to be established.

Supplementary material

Supplements are available online:
https://onlinelibrary.wiley.com/doi/10.1111/aas.13527

References

1 Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1.

2 Lemeshow S, Le J-R. Modeling the Severity of Illness of ICU Patients: A Systems Update. JAMA. 1994 Oct;272(13):1049–55.

3 Vincent JL, Moreno R. Clinical review: Scoring systems in the critically ill. Crit Care. 2010;14(2):207.

4 Bouch C, Thompson J. Severity scoring systems in the critically ill. Contin Educ Anaesth Crit Care Pain. 2008;8(5):181–5.

5 Strand K, Flaatten H, K. S, H. F. Severity scoring in the ICU: A review. Acta Anaesthesiol Scand. 2008;52(4):467–78.

6 Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. Crit Care Med. 2007;35(4):1091–8.

7 Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. Indian J Crit Care Med. 2014 Apr;18(4):220–8.

8 Siontis GCM, Tzoulaki I, Ioannidis JPA. Predicting Death: An Empirical Evaluation of Predictive Tools for Mortality. Arch Intern Med. 2011 Oct;171(19):1721–6.

9 Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and ExplanationThe PRISMA-ScR Statement. Ann Intern Med. 2018 Oct;169(7):467–73.

10 Riviello ED, Kiviri W, Fowler RA, Mueller A, Novack V, Banner-Goodspeed VM, et al. Predicting Mortality in Low-Income Country ICUs: The Rwanda Mortality Probability Model (R-MPM). Lazzeri C, editor. PLoS One. 2016 May;11(5):e0155858.

11 Adhikari NKJ, Fowler RA, Bhagwanjee S, Rubenfeld GD. Critical care and the global burden of critical illness in adults. Lancet. 2010;376(9749):1339–46.

12 Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010 Jan;21(1):128–38.

13 Ferrando-Vivas P, Jones A, Rowan KM, Harrison DA. Development and validation of the new ICNARC model for prediction of acute hospital mortality in adult critical care. J Crit Care. 2017;38:335–9.

14 Timsit J, Fosse J, Troché G, De Lassence A, Alberti C, Garrouste-Orgeas M, et al. Accuracy of a composite score using daily SAPS II and LOD scores for predicting hospital mortality in ICU patients hospitalized for more than 72 h. Intensive Care Med. 2001;27(6):1012–21.

15 Dólera-Moreno C, Palazón-Bru A, Colomina-Climent F, Gil-Guillén VF. Construction and internal validation of a new mortality risk score for patients admitted to the intensive care unit. Int J Clin Pract. 2016;70(11):916–22.

16 Li Z, Cheng B, Wang J, Xie G, Yu X, Huang M, et al. A multifactor model for predicting mortality in critically ill patients: A multicenter prospective cohort study. J Crit Care. 2017;42:18–24.

17 Hadique S, Culp S, Sangani RG, Chapman KD, Khan S, Parker JE, et al. Derivation and validation of a prognostic model to predict 6-month mortality in an intensive care unit population. Ann Am Thorac Soc. 2017;14(10):1556–61.

18 Fika S, Nanas S, Baltopoulos G, Charitidou E, Myrianthefs P. A novel mortality prediction model for the current population in an adult intensive care unit. Hear Lung J Acute Crit Care. 2018;47(1):10–5.

19 Nematifard E, Ardehali SH, Shahbazi S, Eini-Zinab H, Vahdat Shariatpanahi Z. Combination of APACHE Scoring Systems with Adductor Pollicis Muscle Thickness for the Prediction of Mortality in Patients Who Spend More Than One Day in the Intensive Care Unit. Crit Care Res Pract. 2018;2018.

20 Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006 May;34(5):1297–310.

21 Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 - From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31(10):1345–55.

22 Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA. 1993 Nov 24;270(20):2478–86.

23 Gall JR, Lemeshow S, Saulnier F, Le Gall JR, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. JAMA J Am Med Assoc. 1993 Aug 14;270(24):2957–63.

24 Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system: Risk prediction of hospital mortality for critically III hospitalized adults. Chest. 1991;100(6):1619–36.

25 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985 Oct;13(10):818–29.

26 Knaus WA, Harrell FEJ, Lynn J, Goldman L, Phillips RS, Connors AFJ, et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med. 1995 Feb;122(3):191–203.

27 Lemeshow S, Klar J, Teres D, Avrunin JS, Gehlbach SH, Rapoport J, et al. Mortality probability models for patients in the intensive care unit for 48 or 72 hours: A prospective, multicenter study. Crit Care Med. 1994 Sep;22(9):1351–8.

28 Paul E, Bailey M, Kasza J, Pilcher DV. Assessing contemporary intensive care unit outcome: Development and validation of the Australian and New Zealand Risk of Death admission model. Anaesth Intensive Care. 2017;45(3):326–43.

29 Min H, Avramovic S, Wojtusiak J, Khosla R, Fletcher RD, Alemi F, et al. A Comprehensive Multimorbidity Index for Predicting Mortality in Intensive Care Unit Patients. J Palliat Med. 2017;20(1):35–41.

30 Kao R, Priestap F, Donner A. To develop a regional ICU mortality prediction model during the first 24 h of ICU admission utilizing MODS and NEMS with six other independent variables from the Critical Care Information System (CCIS) Ontario, Canada. J Intensive Care. 2016;4(1):1–12.

31 Umegaki T, Sekimoto M, Hayashida K, Imanaka Y. An outcome prediction model for adult intensive care. Crit Care Resusc. 2010 Jun;12(2):96–103.

32 Huang YC, Chang KY, Lin SP, Chen K, Chan KH, Chang P. Development of a daily mortality probability prediction model from Intensive Care Unit patients using a discrete-time event history analysis. Comput Methods Programs Biomed. 2013;111(2):280–9.

33 Stachon A, Segbers E, Hering S, Kempf R, Holland-Letz T, Krieg M. A laboratory-based risk score for medical intensive care patients. Clin Chem Lab Med. 2008;46(6):855–62.

34 Viviand X, Gouvernet J, Granthil C, Francois G. Simplification of the SAPS by selecting independent variables. Intensive Care Med. 1991;17(3):164–8.

35 Stachon A, Becker A, Holland-Letz T, Friese J, Kempf R, Krieg M. Estimation of the mortality risk of surgical intensive care patients based on routine laboratory parameters. Eur Surg Res. 2008;40(3):263–72.

36 Goag EK, Lee JW, Roh YH, Leem AY, Kim SY, Song JH, et al. A Simplified Mortality Score Using Delta Neutrophil Index and the Thrombotic Microangiopathy Score for Prognostication in Critically Ill Patients. Shock. 2018;49(1):39–43.

37 Paul E, Bailey M, Pilcher D. Risk prediction of hospital mortality for adult patients admitted to Australian and New Zealand intensive care units: Development and validation of the Australian and New Zealand Risk of Death model. J Crit Care. 2013;28(6):935–41.

38 Brinkman S, Abu-Hanna A, De Jonge E, De Keizer NF. Prediction of long-term mortality in ICU patients: Model validation and assessing the effect of using in-hospital versus long-term mortality on benchmarking. Intensive Care Med. 2013;39(11):1925–31.

39 Higgins TL, Teres D, Copes W, Nathanson B, Stark M, Kramer A. UPDATED MORTALITY PROBABILITY MODEL (MPM -III). Chest. 2005 Oct;128(4):348S.

40 Studies PRLI for HP. Summary of NQF-endorsed intensive care outcomes models for risk adjusted mortality and length of stay (ICOMmort and ICOMlos). 2009. Available from: https://healthpolicy.ucsf.edu/icu-outcomes

41 Johnson AEW, Kramer AA, Clifford GD. A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. Crit Care Med. 2013;41(7):1711–8.

42 Duke GJ, Barker A, Rasekaba T, Hutchinson A, Santamaria JD. Development and validation of the critical care outcome prediction equation, version 4. Crit Care Resusc. 2013;15(3):191–7.

43 Duke GJ, Santamaria J, Shann F, Stow P, Pilcher D, Ernest D, et al. Critical care outcome prediction equation (COPE) for adult intensive care. Crit Care Resusc. 2008;10(1):35–41.

44 Ho KM, Knuiman M, Finn J, Webb SA. Estimating long-term survival of critically ill patients: The PREDICT model. PLoS One. 2008;3(9).

45 Hunziker S, Celi LA, Lee J, Howell MD. Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients. Crit Care. 2012;16(3):R89.

46 Iapichino G, Mistraletti G, Corbella D, Bassi G, Borotto E, Miranda DR, et al. Scoring system for the selection of high-risk patients in the intensive care unit. Crit Care Med. 2006;34(4):1039– 43.

47 Liu WY, Lin SG, Zhu GQ, Van Poucke S, Braddock M, Zhang Z, et al. Establishment and validation of GV-SAPS II scoring system for non-diabetic critically Ill patients. PLoS One. 2016;11(11):1–13.

48 Granholm A, Perner A, Krag M, Hjortrup PB, Haase N, Holst LB, et al. Development and internal validation of the Simplified Mortality Score for the Intensive Care Unit (SMS-ICU). Acta Anaesthesiol Scand. 2018;62(3):336–46.

49 Mahato B, Otero TMN, Holland CA, Giguere PT, Bajwa EK, Camargo CA, et al. Addition of 25- hydroxyvitamin D levels to the Deyo-Charlson Comorbidity Index improves 90-day mortality prediction in critically ill patients. J Intensive Care. 2016;4(1):1–9.

50 Rydenfelt K, Engerström L, Walther S, Sjöberg F, Strömberg U, Samuelsson C. In-hospital vs. 30-day mortality in the critically ill - a 2-year Swedish intensive care cohort analysis. Acta Anaesthesiol Scand. 2015 Aug;59(7):846–58.

51 Jammer I, Wickboldt N, Sander M, Smith A, Schultz MJ, Pelosi P, et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions: A statement from the ESA- ESICM joint taskforce on perioperative outcome measur. Eur J Anaesthesiol. 2015;32(2):88– 105.

52 Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338.

53 Suistomaa M, Kari A, Ruokonen E, Takala J. Sampling rate causes bias in APACHE II and SAPS II scores. Intensive Care Med. 2000 Dec 14;26(12):1773–8.

54 Labarère J, Bertrand R, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med. 2014;40(4):513–27.

55 Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016 Jun;74:167–76.

56 Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Vol. 19. 2009.

57 Granholm A, Perner A, Jensen AKG, Møller MH. Important methodological flaws in the recently published clinical prediction model the REMEMBER score. Crit Care. 2019;23(1):4–5.

58 Calvert J, Mao Q, Hoffman JL, Jay M, Desautels T, Mohamadlou H, et al. Using electronic health record collected clinical variables to predict medical intensive care unit mortality. Ann Med Surg. 2016 Nov;11:52–7.

59 Meiring C, Dixit A, Harris S, MacCallum NS, Brealey DA, Watkinson PJ, et al. Optimal intensive care outcome prediction over time using machine learning. PLoS One. 2018;13(11):e0206862.

60 Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning. Sci Rep. 2019 Feb;9(1):1879.

61 Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019 Jun;110:12–22.

62 Poncet A, Perneger T V, Merlani P, Capuzzo M, Combescure C. Determinants of the calibration of SAPS II and SAPS 3 mortality scores in intensive care: a European multicenter study. Crit Care. 2017 Dec 4;21(1):85.

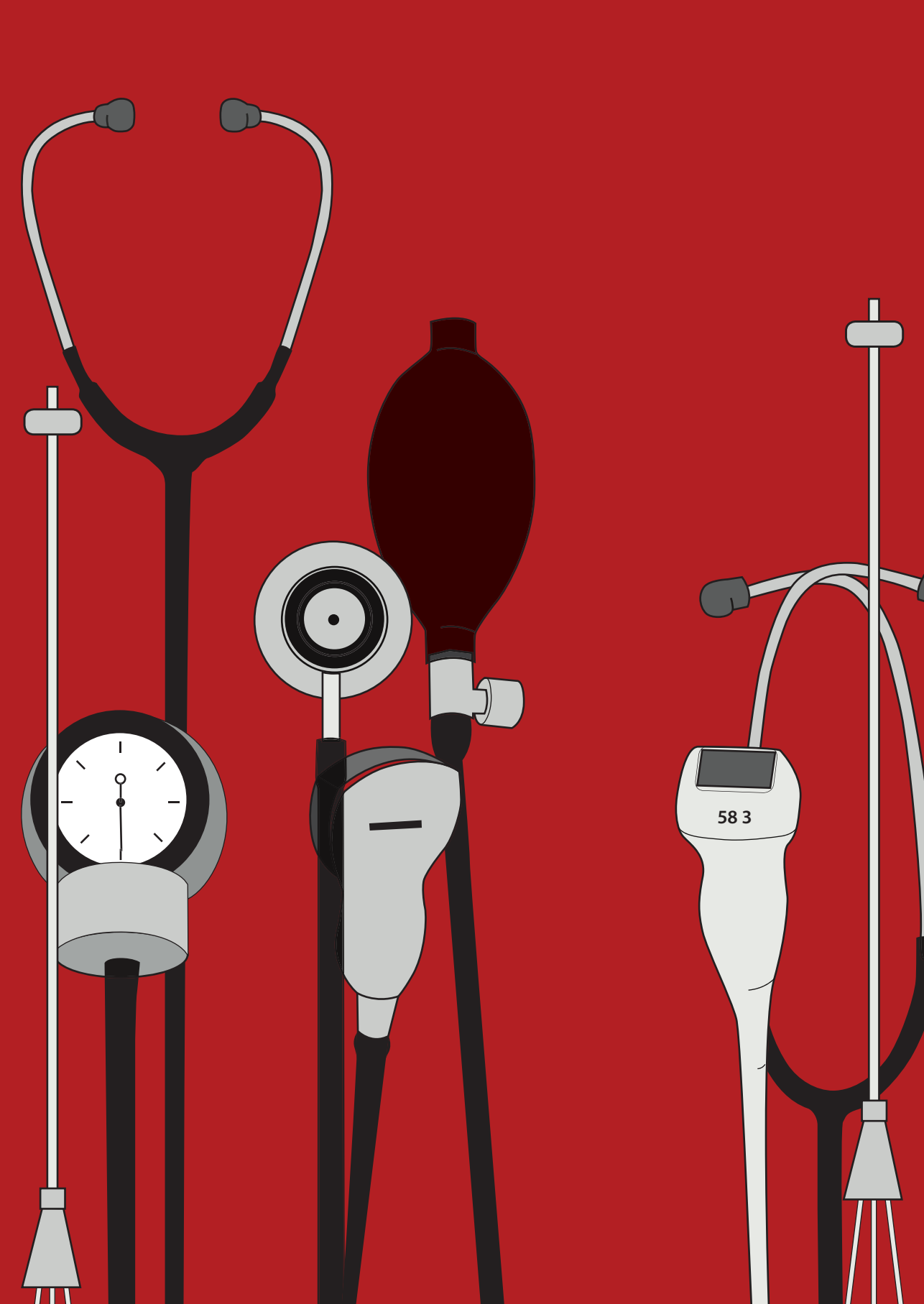
63 Kitchenham B. Procedures for performing systematic reviews. Keele, UK, Keele Univ. 2004;33(2004):1–26

64 Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews. PLoS One. 2012 Feb;7(2):e32844.

65 Prin M, Wunsch H. International comparisons of intensive care: informing outcomes and improving standards. Curr Opin Crit Care. 2012 Dec;18(6):700–6.

66 Winters BD, Eberlein M, Leung J, Needham DM, Pronovost PJ, Sevransky JE. Long-term mortality and quality of life in sepsis: a systematic review. Crit Care Med. 2010 May;38(5):1276– 83.

67 Gayat E, Cariou A, Deye N, Vieillard-Baron A, Jaber S, Damoisel C, et al. Determinants of long-term outcome in ICU survivors: results from the FROG-ICU study. Crit Care. 2018;22(1):8.


58 3

9

Foresight over hindsight: Mandatory publication of clinical research protocols prior to conduct

Thomas Kaufmann
Anders Granholm
Frederik Keus
John Myburgh
Anders Perner
Morten Hylander Møller
Iwan C.C. van der Horst

Acta Anaesthesiologica Scandinavica

Introduction

Recently, the Extracorporeal membrane Oxygenation (ECMO) to res cue Lung Injury in severe Acute respiratory distress syndrome (ARDS) (EOLIA) trial was published (1). This randomized controlled trial (RCT) evaluated early initiation of ECMO compared with standard care in patients with severe ARDS. At a planned interim analysis after the enrolment of 240 (73%) of the targeted 331 patients, the trial was stopped prematurely, according to prespecified stopping rules, as continuation of the trial would likely not lead to a statistically significant difference in the primary outcome (60-day mortality). The sample size estimation was based on an unlikely effect size of 20% absolute risk reduction (2). In the final analysis performed after the interim analysis, the 60-day mortality was 35% in the ECMO group and 46% in the standard care group (relative risk 0.76; 95% CI: 0.55-1.04, P = 0.09). The true effect may thus vary from a 45% relative risk reduction to a 4% relative risk increase, but the trial provides no firm evidence in favor or against the intervention. Comments emerged in medical journals and on social media about the decision to stop the trial. One of the accompanying editorials stated: "We are disappointed that the Data and Safety Monitoring Board (DSMB) acted so quickly to stop the trial, but others may have reached the same decision as the DSMB" (3). Also, on social media many suggested that stopping the trial early for futility was unethical, while others supported the investigators for following the predefined criteria (4).

The EOLIA trial renewed the discussion of hindsight on any study, especially since the trial had a protocol and followed it. The debate about this RCT after publication in a peer-reviewed high impact journal and the ensuing social media discussion highlight the need to consider key aspects of trial methodology. First, trial planning must include both realistic and biologically plausible hypothesized effect sizes, assumed event rates, and sample size estimations. In critical care research, effect sizes are invariably inflated and implausible — for example, an absolute reduction in mortality >10% is frequently reported (5,6) that is not considered plausible by trialists and clinicians (2). Second, interim analyses and stopping criteria must be clearly specified before trial onset, and should consider different scenarios including possible harms, likely benefits and in some cases futility. Currently, research is usually registered in a registry (e.g., ClinicalTrials.gov or PROSPERO) prior to initiation; however, these registrations are often not detailed, not peer-reviewed and they may go unnoticed by other experts in the field. This is sometimes followed by publication of a full protocol in a peer-reviewed journal, often while the research is being performed and changes in study design may be difficult to implement. We propose that full protocols of trials and other clinical studies are published in peer-reviewed journals prior to the commencement of patient recruitment (for prospective studies), data access (for retrospective studies), or search conduction (systematic reviews), so methodology may be optimized.

Current initiatives for improved conduct of research

The International Committee of Medical Journal Editors (ICMJE) has developed recommendations regarding planning and reporting of medical studies (7). The Enhancing the QUAlity and Transparency Of health Research (EQUATOR) network hosts an increasing number of checklists

that may guide researchers (8), including guidelines for reporting systematic reviews and meta-analyses (9), randomised trials (10), and observational studies (11). In addition, extensions of guidelines have addressed the planning and conduct of randomized trials (12,13) and systematic reviews with meta-analyses (14). These initiatives add to achieving results that are reproducible, transparent and have rigor, so that clinicians, patients, and society may trust them (15). Even though these initiatives are in place, many trials still have avoidable methodological flaws and inconsistencies (16).

Perceived advantages of prepublication of study protocols

- Publishing study protocols will likely improve the quality of planning and conduct of studies. Protocols can be checked according to the checklist fitting the study design, that is, Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) for RCTs, and Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) for systematic reviews and meta-analyses (12,14). Modifications can be made based on established indices of high internal validity. Qualified expert readers assessing study protocols by means of peer review may identify important flaws or omissions, allowing corrections in a timely manner, that is, before the study is commenced.
- Hypothesized effect size and sample size estimation are open to discussion, which may allow alterations of implausible effect sizes and increase study robustness and the chance of obtaining useful evidence.
- A mandatory published study protocol adds transparency as it allows methods to become publicly available.
- If other researchers or research groups aim to study a similar research question, these studies could serve as validation and confirmation. This may provide opportunities for trial protocol harmonization of definitions (population, intervention, comparator and outcomes), that may reduce clinical heterogeneity and strengthen the interpretation of results of subsequent trial- or patient-level systematic reviews and meta-analyses.
- Accepting publication of a study before completion based on a published, peer-reviewed protocol may reduce time between completion and publishing due to rejections, revisions, and repetitive submissions. Reviewers and editors will only have to focus on protocol adherence, clarity of presentation of the results and discussion, as the methods will have already been reviewed.
- Publishing protocols may facilitate the peer review process.
- Knowledge on upcoming multicenter trials may facilitate the recruitment of centers.

Perceived disadvantages of prepublication of study protocols

- Potential delay in time between getting the idea for a research project and publishing the manuscript, and in the time between finishing the protocol and beginning the research project/study.
- Funding agencies require a near-final version of a protocol for their funding applications. Funding becomes more difficult to obtain if protocol amendments must be implemented. Protocol amendments become more difficult to implement if funding has already been obtained.

- Innovative ideas, included in the protocols, are open and could lead to incorporation in studies of other researchers or research groups.

Suggestions for improved implementation in dissemination of study protocols

- Publication in dedicated protocol journals or as a section in existing journals, either online or in print.
- Create an electronic platform to facilitate suggestions and comments potentially amending the study protocol.
- Set reasonable time limits for suggestions and comments to be considered for the final protocol.
- Replace the initial protocol by the amended protocol once the definite publication emerges and add the initial protocol as a supplementary file, publish the amendments as a second supplementary file and cite the authors of the suggestions and comments that were the reason for the amendment to acknowledge their intellectual contribution.
- Publish the amended protocol on a dedicated database, such as PROSPERO or ClinicalTrials.gov.
- Opportunities for journals to commit themselves to publishing the outcomes of amended study protocols if conducted accordingly.

Conclusions

We recommend peer-review of study protocols by way of prepublication prior to study commencement as it may improve study design, increase transparency of study conduct and reporting, and increase trust in the results. Several obstacles need to be considered, but the benefits are likely to outweigh the perceived disadvantages.

References

1 Combes A, Hajage D, Capellier G, Demoule A, Lavoué S, Guervilly C, et al. Extracorporeal Membrane Oxygenation for Severe Acute Respiratory Distress Syndrome. N Engl J Med. 2018 May 24;378(21):1965–75.

2 Ridgeon EE, Bellomo R, Aberegg SK, Sweeney R Mac, Varughese RS, Landoni G, et al. Effect sizes in ongoing randomized controlled critical care trials. Crit Care. 2017 Dec 5;21(1):132.

3 Harrington D, Drazen JM. Learning from a Trial Stopped by a Data and Safety Monitoring Board. N Engl J Med. 2018 May 24;378(21):2031–2.

4 ADAlthousePhD. This article calls for a TWEETORIAL. https://twitter.com/ADAlthousePhD/status/999626627381329920 2018 May 24 [cited 2018 September 5] [Tweet]. Available from: @ADAlthousePhD.

5 Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, et al. Outcomes and Statistical Power in Adult Critical Care Randomized Trials. Am J Respir Crit Care Med. 2014 Jun 15;189(12):1469–78.

6 Aberegg SK, Richards DR, O'Brien JM. Delta inflation: a bias in the design of randomized controlled trials in critical care medicine. Crit Care. 2010;14(2):R77.

7 International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. Accessed on 5 September 2018. Available from: http://www.ICMJE.org.

8 Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR network's survey of guideline authors. PLoS Med. 2008 Jun 24;5(6):0869–74.

9 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009 Jul 21;339(jul21 1):b2535–b2535.

10 Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar 23;340(mar23 1):c332–c332.

11 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007 Oct 20;370(9596):1453–7.

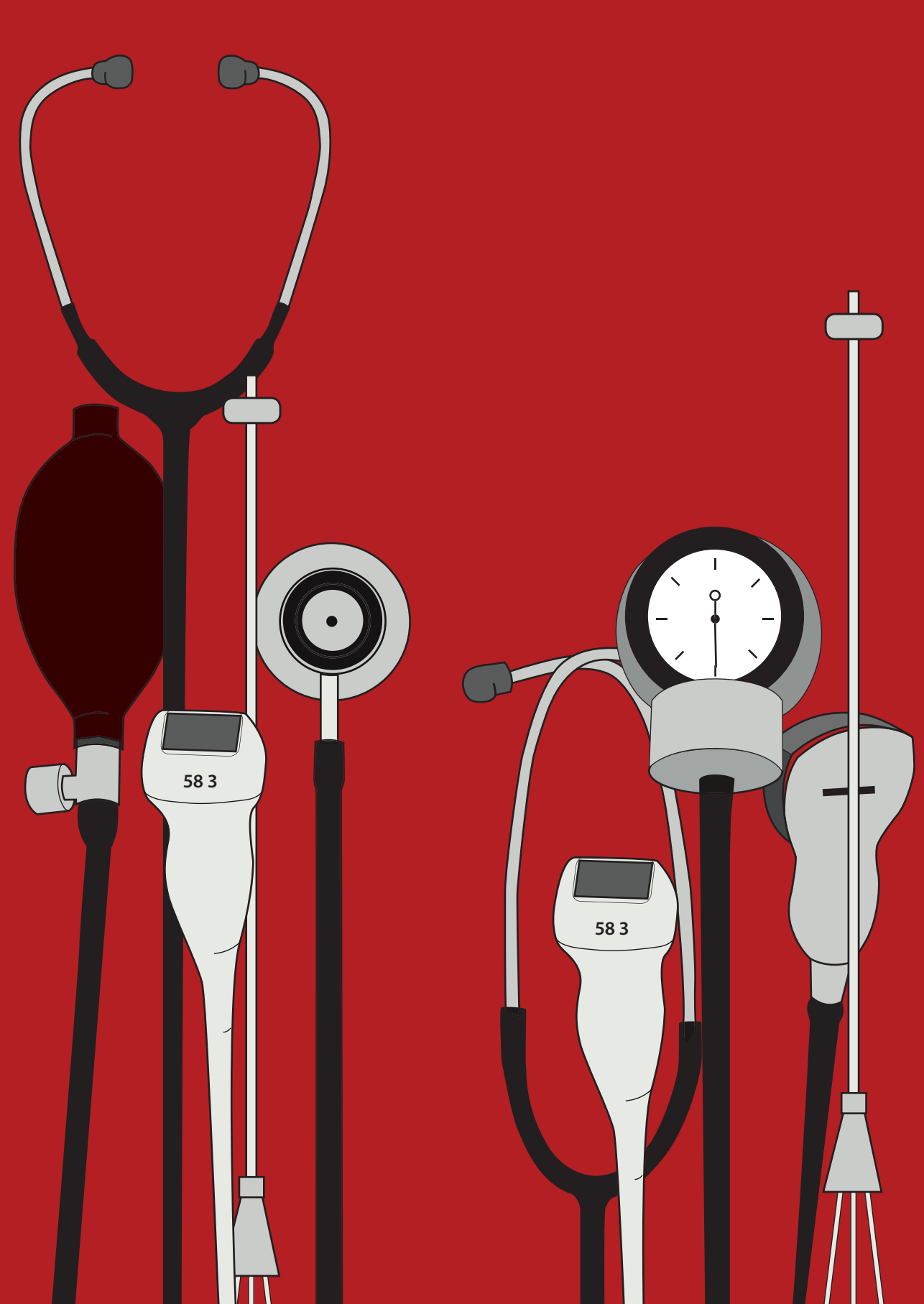
12 Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. Ann Intern Med. 2013 Feb 5;158(3):200.

13 Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan A-W, King MT, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols. JAMA. 2018 Feb 6;319(5):483.

14 Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015 Dec 1;4(1):1.

15 McNutt M. Journals unite for reproducibility. Science (80-). 2014 Nov 7;346(6210):679–679.

16 Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. BMJ. 2015 Jun 11;350(1):87.


58 3
58 3

10

Summary, discussion, and future perspectives

Summary

This thesis aimed to extend the evidence on the applicability of hemodynamic monitoring to prevent and treat hemodynamic instability during the perioperative period and admission to the ICU. Second, we aimed to gain knowledge on how to improve the conduct of studies in perioperative and critical care medicine.

Hemodynamic monitoring used in perioperative goal-directed therapy

Perioperative goal-directed therapy is more than a single intervention

The perioperative goal-directed therapy (PGDT) intervention using hemodynamic monitoring has been evaluated in many different randomized controlled trials (RCTs). In our systematic review, however, a meta-analysis of data obtained from these trials was considered inappropriate due to clinical heterogeneity (***chapter 2***). The use of different hemodynamic variables as a target for intervention and different hemodynamic monitors were only some of the aspects of extensive heterogeneity observed in the complex PGDT intervention. Given the existing standards requested for high-quality evidence, it was unjustified to draw a uniform conclusion on the effect of PGDT on patient-centered outcomes. We demonstrated the need for universal and uniform definitions of the PGDT intervention to make a comparison between different trials possible. In addition, whenever results of individual RCTs are pooled in a meta-analysis, this inherent heterogeneity needs to be considered by readers before conclusions are drawn.

Even though we showed that high-quality evidence was not available for the PGDT intervention, i.e., a meta-analysis pooling data from trials with low risk of bias, a combination of several aspects needs to be considered to make sure that the right patients will benefit from the PGDT intervention. First, the PGDT intervention needs to be applied in high-risk patients undergoing high-risk surgery. Second, the initiation of the PGDT intervention needs to be started early during the perioperative period. Third, the PGDT algorithm needs to combine flow optimization, i.e., cardiac output (CO) or stroke volume (SV), with dynamic preload variables, e.g., stroke volume variation (SVV), to assess fluid responsiveness (***chapter 3***).

Hemodynamic monitoring in the intensive care unit

Data obtained as part of the prospective observational Simple Intensive Care Studies-I (SICS-I) were used to expand the available evidence regarding hemodynamic monitoring in the ICU. The primary hemodynamic assessment when a circulatory shock is suspected can be described as a step-by-step approach where initially clinical examination is performed, followed by a CO measurement using critical care ultrasonography (CCUS), and concluded by more advanced hemodynamic monitoring if a complex shock state requires so. In this thesis, we expanded the evidence for each of these steps.

Clinical examination is the first step but needs to be improved

We assessed the current practice of clinical examination in critically ill patients and mapped a Bayesian network, which showed that physicians regard the presence of a high-dose norepinephrine infusion and a prolonged capillary refill time or mottling of the knee as being

conditional dependencies of a low cardiac pump function (***chapter 4***). The methodology employed to construct the Bayesian network was used for the first time in this setting so that rigorous testing and validation were necessary. Our findings give an insight into the possible thought process of physicians when performing a clinical examination and show that improvements regarding the diagnostic accuracy of the clinical examination for hemodynamic assessment are needed.

Blood pressure measurement methods are not interchangeable

Blood pressure measurements obtained with a non-invasive oscillometric upper-arm cuff showed significant limits of agreement compared to invasive arterial pressure measurement, and error-grid analysis showed that one in four patients on norepinephrine would potentially have had at least a low-risk treatment decision (***chapter 5***). Blood pressure measurements obtained with an invasive arterial catheter are at the moment the clinical standard in patients with circulatory shock, and our findings do not support replacing this method with a non-invasive alternative. We believe that in clinical practice, it is essential to remain critical on the interchangeability of methods of blood pressure measurement. The analyses that were used to compare the two blood pressure measurement methods provide limited data on measurement differences between different methods in individual patients. For clinical practice, this means that in every individual patient, the optimal applicable method of measurement needs to be determined based on the benefits and disadvantages.

Experts remain essential for the interpretation of images obtained with critical care ultrasonography

Medical students are capable of obtaining critical care ultrasonography (CCUS) images of sufficient quality in critically ill patients admitted to the ICU, but experts are still required for proper interpretation of these images (***chapter 6***). The quality of the images obtained by the medical students was of sufficient quality, but the measurements of CO that were made by the medical students based on these images showed significant limits of agreement, i.e., unacceptable accuracy of the measurement differences, compared to measurements made by experts from the same images. Being able to acquire the required CCUS images and perform the calculation of CO is essential as the utility of CCUS in the ICU increases. Our findings were of interest as being able to calculate CO using CCUS is considered an advanced skill, and the medical students had no prior experience in using CCUS. For future research and clinical application of CCUS, it is feasible to have non-experts acquire the images which can then be interpreted by experts, thus saving time and resources.

Uncalibrated pulse wave analysis to measure cardiac output is not indicated in patients with circulatory shock

CO measurements obtained with an uncalibrated pulse wave analysis device showed significant limits of agreement and clinically unacceptable trending ability compared to CCUS performed in patients with circulatory shock (***chapter 7***). At the moment, the uncalibrated pulse wave analysis method of measuring CO is not recommended in patients with circulatory shock, and our findings support this. Due to the chosen methods, it was not possible to prove direct superiority

or non-inferiority of one method over the other, as CCUS is not a gold standard reference technique required for formal method comparison. Comparison in prospective observational studies, however, may reflect clinical practice better and allows for validation of these methods in populations of patients who were initially not the target population.

Initiatives to improve methodology and conduct of studies

External validation of research findings in independent cohorts are needed

We presented an overview of 43 identified mortality prediction models for estimating mortality in unselected critically ill patients with varying quality of methodology, and the validation of the individual models differed and was most often incomplete (***chapter 8***). External validation of the mortality prediction models is often lacking, and head-to-head comparisons are needed to identify the best performing model out of the available mortality prediction models for guiding clinical care and research in different settings and populations.

Mandatory publication and peer-review of research protocols

Peer-review of study protocols by way of prepublication before initiation of a study may increase the transparency of study conduct and reporting (***chapter 9***). We believe that a structured discussion of aspects of study design in advance may limit the discussion in hindsight regarding the conduct and interpretation of the study results. Protocols can be reviewed according to the checklist fitting the study design, and modifications can be made based on established indices of high internal validity. Notable flaws and omissions can be identified, which allows corrections on time. Potential delays in study commencement need to be considered, and innovative ideas are in the open and could be adopted by other researchers. Transparency and optimization of study design and conduct benefit not only researchers but also patients, which is the ultimate goal of all conducted research.

Discussion and future perspectives

The chapters of this thesis combine to show that it is hard to obtain high-quality evidence needed to demonstrate the benefit of the application of hemodynamic monitors in perioperative goal-directed therapy (PGDT) in the operating room (OR) or patients treated in the intensive care unit (ICU). Even though several national and international guidelines support the use of the PGDT intervention in major surgery (1,2), the beneficial effect of PGDT is not based on high-quality evidence yet. The same principle goes for the application of hemodynamic monitors in patients in the ICU as there is no high-quality evidence on the interchangeability of different monitors that measure cardiac output (CO) or blood pressure. The ultimate goal in this research field is to find the best applicable methods to measure hemodynamic variables reliably and that these variables are valuable to guide treatments that benefit individual patients. Here, we address the possible reasons for the findings of this thesis and elaborate on the future of hemodynamic monitoring.

Considerations for hemodynamic monitoring in the operating room

Hemodynamic monitors are applied in the context of the PGDT intervention, and this intervention consists of composite components, which makes a direct comparison between studies problematic. Whenever a systematic review with meta-analysis of individual PGDT trials is completed, the limitation section of these articles contains the following standard phrases: presence of clinical and statistical heterogeneity, limited comparability between different hemodynamic monitors, and high risk of bias among included studies. While some parts of these limitations are inherent to the PGDT intervention, other parts may be preventable if more attention is paid to them. One should be aware that the fact that limitations may be inherent does not mean that they should be neglected when accumulating and assessing the quality of evidence, and consideration of these limitations when designing the study is crucial.

Heterogeneity

Studies brought together in a systematic review will differ because of the varying characteristics of the individual trials. The Cochrane Handbook for Systematic Reviews of Interventions describes variability in the participants, interventions, comparisons, and outcomes studied as clinical heterogeneity and variability in the intervention effects being evaluated as statistical heterogeneity (3). Statistical heterogeneity can be quantified using various statistical tests, but the interpretation of clinical heterogeneity is left to the reader. As the PGDT intervention studies often use different types of monitoring devices, hemodynamic variables, target values, types of fluids, and types of vasoactive medication, and are performed in different patient groups undergoing different types of surgery, heterogeneity is inevitable. For example, there is still a research debate regarding the beneficial effect of using certain types of fluids (i.e., crystalloids versus colloids) on postoperative outcomes of patients undergoing surgery (4). The possible implications regarding the type of fluid to use underline the rationale to standardize these components of the PGDT intervention when performing a meta-analysis. Because of clinical heterogeneity alone, we decided that pooling data from the individual PGDT trials in one meta-analysis was not appropriate.

Research in other areas of critical care has shown that clinical heterogeneity may influence the results of systematic reviews with meta-analyses (5,6). Overall, heterogeneity limits the external validity of research findings. The consequence for PGDT is that it is sometimes considered to be a poorly defined term with impaired implementation in clinical practice as a result (7). Novel trials on the benefit of the PGDT intervention need to report all individual components of PGDT to allow for a more uniform comparison. In addition, the first trial on the PGDT intervention was published over thirty years ago (8), and this trial is still included in present-day systematic reviews with meta-analyses (9). Mortality in perioperative medicine has decreased substantially over the past fifty years (10), and thus with the general advancements in medicine over time, it seems inappropriate to merge data from these early PGDT trials with contemporary trials to estimate a single pooled intervention effect.

Furthermore, research on PGDT will benefit from trials with a sufficiently large sample size as most of the trials published at the moment only report on small patient groups, which increases the risk of heterogeneity (9). Although, if large groups of patients are needed to show a statistically significant benefit, it might not be a clinically relevant effect. On the other hand, statistical techniques such as heterogeneity treatment effect analysis using Bayesian hierarchical models included in protocols of extensive trials may be one way to limit the clinical heterogeneity (11).

Comparability of hemodynamic monitors

Various hemodynamic monitors are used to determine the target hemodynamic variables used in the PGDT intervention. Hemodynamic variables in PGDT algorithms pooled in systematic reviews are handled on the assumption that measurements produce the same value regardless of the type of hemodynamic monitor used. However, each of these monitors performs the measurements of these variables via different methods, and they do not naturally agree on values. Method comparison studies are used to determine the validity and interchangeability of hemodynamic monitors measuring hemodynamic variables (12). These studies consist of assessments of precision, accuracy, and trending ability of a test method against an established reference method (i.e., gold standard). The established gold standard produces the measurement value that is considered to be the correct one, and the agreement of the test value is compared to the correct value. High precision means that the values of the two methods will have little spread, and high accuracy means that the values on average will be close to the correct value (figure 1).

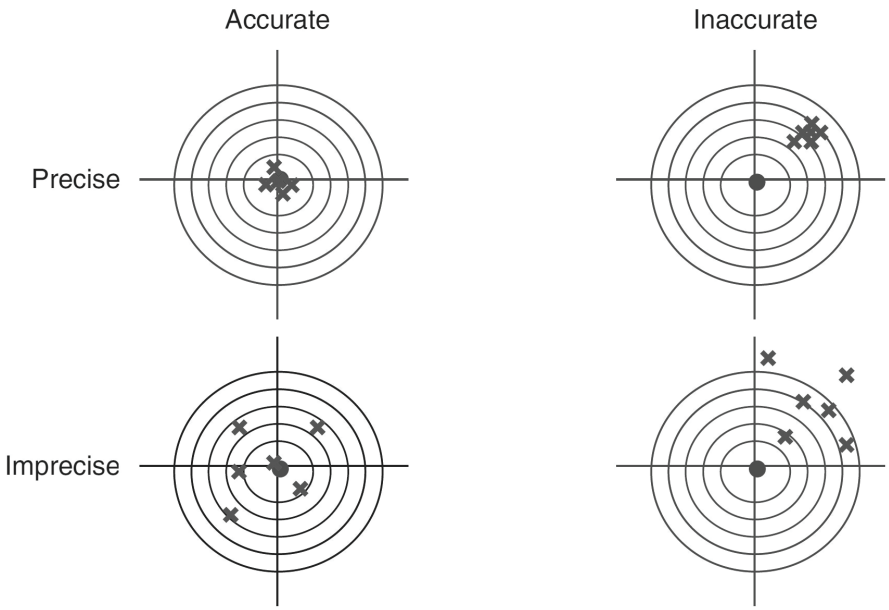


Figure 1. *Accurate measurements are close to the correct value, irrespective of the spread of the measurements; precise measurements are close to each other, regardless of the deviation from the correct value. Valid hemodynamic monitors should be both accurate and precise*

There are several concerns regarding the assumption that measurements of hemodynamic variables using different hemodynamic monitors are interchangeable. First, not all monitors show the acceptable agreement of measurements of hemodynamic variables (13). Second, most method comparison studies were performed in patients undergoing cardiac surgery (13). One of the reasons is that this patient group is routinely monitored using the gold standard pulmonary artery catheter, which facilitates performing method comparison studies. However, results of agreement of hemodynamic monitors obtained in cardiac surgery patients cannot be directly transferred to other patient populations as physiology varies in different populations. For example, patients undergoing liver surgery may experience changes in arterial resistance, which makes the uncalibrated pulse wave analysis method to measure CO less reliable (14). Third, studies assessing interchangeability often fail to report all necessary assessments of agreement: precision, accuracy, and trending ability (15,16). Incomplete reporting of items in publications is a general problem, as intervention studies in anesthesiology often lack in full reporting of items necessary for a comprehensive evaluation of an intervention (17). A checklist was developed to assist with providing a complete method comparison, but many studies were performed before this quality check was available (12). We found a total of 18 different hemodynamic monitors used in the PGDT trials in one of our studies. In another study, we applied the checklist to compare CO measurements made using the uncalibrated pulse wave analysis method with critical care ultrasonography (CCUS) in critically ill patients. As more hemodynamic monitors are still being developed, method comparison studies need to optimize their application and reporting of

the methodology to allow for proper validation and implementation of the novel monitoring technique (18,19).

Risk of bias

A bias is a systematic error influencing the results of a study. This type of error implies that multiple replications of the same study would reach the wrong answer on average (3). Bias can be interpreted as a target in figure 1 with low accuracy; high bias means that there is a low accuracy for the test results as there is a systematic deviation from the correct value. A risk of bias assessment is an essential component of performing a systematic review with meta-analysis. There are several forms of bias that are considered in a formal risk of bias assessment, such as selection bias, detection bias, attrition bias, and reporting bias. These different biases are known to contribute to an overestimation or underestimation of the true intervention effect (20).

Sometimes, a particular risk of bias is inherent to a field of research. For example, in the PGDT intervention, it is impossible to blind participants and study personnel to invasive hemodynamic monitoring devices. Some of these devices are connected to indwelling catheters in large arteries or veins, or a probe is placed in the esophagus of the patient (21). That an intervention cannot be blinded should not mean that a lack of blinding is not considered a risk of bias, although there is no evidence that a lack of blinding in itself influences intervention effects on mortality (22). We found only one trial on the PGDT intervention with an overall low risk of bias. The most common cause for the risk of bias of the PGDT intervention was lack of blinding or collaboration with the industry by at least one of the authors. Even if blinding in itself is not considered as an additional risk, most studies that we identified had a high risk of bias.

Collaboration with the industry is a specific bias of interest for the PGDT intervention. As clinical research is widely sponsored by companies, there may be a financial incentive to publish positive results. It has been shown that drug and device studies sponsored by companies have a more favorable intervention effect and more often have a positive conclusion (23). In PGDT research, there exists a significant impact of authors' collaboration with the industry on publishing a positive conclusion to a trial regarding the PGDT intervention (24). Research on the benefit of PGDT will be helped by trials with an as low as possible risk of bias as it is an intervention prone to biases that influence the overall assessment of the intervention effect.

Hemodynamic monitoring in the ICU – improving on the current application

Improving on the contemporary clinical examination for hemodynamic assessment

Clinical examination of critically ill patients comes first in the hemodynamic assessment of suspected circulatory shock (25). Based on available evidence, the clinical examination should be used as a trigger for additional measures in patients with circulatory shock, such as a measurement of CO using CCUS or additional hemodynamic monitoring (26). Even though several clinical examination findings are independently associated with CO, multivariable analyses of these findings showed that clinical examination alone was insufficient to estimate CO (27). It is vital to improve the hemodynamic assessment made with clinical examination. Although technological

aids are becoming more widely available, further research into the value of clinical examination may potentially limit inappropriate overuse of technological aids (28).

Clinical examination by physicians can be improved by identifying some of the cognitive biases physicians may be subjected to and negating them (29). Physicians may be prone to premature closure in which the physician fails to consider reasonable alternatives after an initial diagnosis is made (30). In addition, a physician may be prone to confirmation bias in which the physician is prevented from considering other information having already reached a conclusion, and involuntarily dismisses information which does not support their decision. For example, a physician may examine a critically ill patient who is receiving a high dose of norepinephrine infusion and may believe that the presence of this value must imply that CO is low. From earlier studies, we know that CO does not necessarily decrease with norepinephrine infusion, and CO response to norepinephrine infusion varies with underlying disease and comorbidities (31,32). These cognitive biases and how they limit diagnostic accuracy can be identified by analyzing clinical examination data using machine learning methods.

We demonstrated that a crude insight into the thought process regarding the estimation of cardiac function by physicians using clinical examination is possible with Bayesian networks. This method has been developed and tested in an article describing the evolution of human orthodontic features (33). While we demonstrated that this method was feasible to give an insight into the overall thought process of multiple physicians and researchers, a similar approach could also be applicable for individual physicians. The physician can then be made aware of their thinking process and can be trained to use or leave out specific values obtained with the clinical examination, or the physician can be trained to reconsider their conclusion when new information is presented. Individual physicians can be trained using this method built into a game, and if multiple physicians are trained, the overall diagnostic accuracy of clinical examination can be improved. Using games for these types of educational purposes has been shown to enhance knowledge gain in various settings (34,35).

Clinical trials are needed to find the optimal method of blood pressure measurement in every individual patient

Reliable blood pressure measurements are of vital importance for the care of the critically ill patient in the OR and the ICU. A mean arterial pressure (MAP) of 65 mmHg is the lower limit of blood pressure to target (25), and this is chosen because research in the OR (36) and the ICU (37,38) has shown that a MAP below this value is associated with complications and increased mortality. An invasive arterial catheter is the clinical gold standard for blood pressure measurements in critically ill patients and blood pressure monitoring during high-risk surgery. Placing an arterial catheter has other benefits besides continuous blood pressure monitoring, such as facilitating the regular drawing of blood samples and blood gas analyses. Risks such as thrombosis and catheter infections should be considered when placing an arterial catheter (39). Various alternatives have been developed to try and reduce the invasiveness of hemodynamic monitoring. These alternatives, such as the non-invasive oscillometric upper-arm cuff, are also

regularly used in critically ill patients (40). Because it is essential to target a specific value of blood pressure in critically ill patients, it is also vital that the values of the different methods used agree. Most studies that assessed the agreement between the non-invasive oscillometric upper-arm cuff and the invasive arterial catheter are retrospective analyses (41,42). These studies were limited by the inability to correct potential issues that influence measurements, such as ensuring that the correct size upper-arm cuff was used for each patient or leveling and zeroing the arterial system to prevent under- or overdamping. We performed a prospective observational study where we considered these potential issues. Besides the statistical agreement, it is also beneficial to look at the clinical significance of measurement differences. Error grid-analysis was developed for this purpose (43), and we applied this method in our study.

The methods used in method comparison studies give overall conclusions on the interchangeability of different methods of measurement. Bland-Altman analysis and error-grid analysis both provide an answer to questions regarding interchangeability based on measurements of an entire group of patients. For individual patients, one method may be preferred over the other, but at the moment, there is no evidence to support the preferred use of one method over another. Ultimately, clinical trials are needed to ensure a potential improvement of patient safety and cost-effectiveness, and the appropriate method to measure blood pressure can be designated for specific subgroups of critically ill patients (44).

Critical care ultrasonography will become standard of care in the intensive care unit

After clinical examination, including blood pressure monitoring, performing a CO measurement is the next step of hemodynamic assessment in critically ill patients suspected of circulatory shock (25,26). The routine use of CCUS for this purpose is becoming more widespread, and ultrasonography machines are increasingly available at the bedside in the ICU (45). It is now expected that every new ICU physician is familiar with basic assessments performed with CCUS (45). These basic assessments are designed to answer a simple binary question, such as whether a patient is fluid responsive or whether there is pericardial effusion (46). Measurement of CO is considered to be an advanced assessment, but the ability to perform this measurement could be of significant interest in the daily hemodynamic assessment of critically ill patients (47).

Our finding that, after a short training, even novices in CCUS can obtain the necessary images but are unable to produce a correct CO in those images supports the fact that measuring CO is considered an advanced assessment. This finding, however, also supports the use of CCUS novices to acquire the necessary images to measure CO. The actual CO measurement is reserved for experts. We used medical students as novices, and even though they did not follow a formal training program, the training they received was similar to the national requirements for CCUS of many countries (48). The full implementation of CCUS in the ICU requires clearly defined competencies (48). Once these have been agreed upon, a supporting training infrastructure can be set up. Eventually, a clinical trial has to be performed, demonstrating improved patient outcomes when CCUS is used for the daily hemodynamic assessment of critically ill patients. Recently, such a trial was conducted to determine the benefit of hemodynamic monitoring using transesophageal echocardiography in critically ill patients with circulatory shock (49).

Prepublication of research protocols will improve the conduct and reporting of studies

Published studies can contain misleading results and waste valuable resources as a consequence of weaknesses in the design and conduct of the study (50). Waste of research can prevent physicians from making well-informed decisions, with severe consequences if harmful or ineffective treatments are promoted (51). This waste may be in part due to the reasons mentioned in this discussion, which influence study results. As said before, some of these reasons are inherent to study design. However, there are also correctable flaws in study methodology and conduct (52).

We looked at mortality prediction models and observed an example of studies with potential flaws in methodology. We provided an overview of all available mortality prediction models in adult critically ill patients and demonstrated that a large number of these models lack complete development. A critical aspect of the development of a prediction model is to test the model in external cohorts, i.e., external validation (53). External validation was one of the factors lacking in the majority of the mortality prediction models we identified. Validation of a prediction model in the same cohort as it was developed in is considered to be methodologically flawed as the developed model will perform better (53,54). The eventual aim is to identify the mortality prediction model that is the most suitable score for a specific patient cohort in clinical practice. The next step to achieve this is to perform a head-to-head comparison in a single external validation cohort.

It would be beneficial to make it mandatory to draft and make a protocol available before the conduct of a study to help achieve the complete development of prediction models. In this protocol, the comprehensive development of the model could be described, and any flaws in methodology could be addressed beforehand. We even believe that research waste can be decreased for most studies by allowing peer-review of these study protocols before the initiation of a study. The same principle applies to method comparison studies that are needed to help validate hemodynamic monitors and for trials on the PGDT intervention or the implementation of monitors in clinical practice. Ultimately, this allows researchers to increase the transparency of study conduct and reporting and may help improve the quality of evidence for interventions performed in patients.

To conclude, future studies on the comparison of hemodynamic monitors or the application of hemodynamic monitors in the PGDT intervention need to properly plan and report on all aspects of study design and conduct. For the PGDT intervention, this means that a protocol is drafted using the proper checklists (55,56), this protocol is peer-reviewed by experts to ensure an as low as possible risk of bias, a feasible theorized intervention effect on relevant outcomes (57,58), and transparency regarding the individual components of the PGDT intervention to limit heterogeneity (7). For method comparison studies of hemodynamic monitors, this also means that a protocol is drafted and peer-reviewed in advance with consideration of the applicable checklists (12,59) to ensure that all components of validation of a novel monitor; precision, accuracy, and trending ability, are addressed according to statistically robust methodology (60,61).

References

1 National Institute for Health and Clinical Excellence. Medical technologies guidance MTG3: CardioQ-ODM oesophageal doppler monitor. 2011.

2 European Society of Anaesthesiology. Perioperative Goal-Directed Fluid Therapy Protocol Summary. p. 7. Available from: http://html.esahq.org/patientsafetykit/resources/downloads/05_Checklists/Various_Checklis ts/Perioperative_Goal_Directed_Therapy_Protocols.pdf

3 Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011.

4 Joosten A, Delaporte A, Ickx B, Touihri K, Stany I, Barvais L, et al. Crystalloid versus Colloid for Intraoperative Goal-directed Fluid Therapy Using a Closed-loop System. Anesthesiology. 2018;128(1):55–66.

5 de Grooth HJ, Postema J, Loer SA, Parienti JJ, Oudemans-van Straaten HM, Girbes AR. Unexplained mortality differences between septic shock trials: a systematic analysis of population characteristics and control-group mortality rates. Intensive Care Med. 2018;44(3):311–22.

6 de Grooth HJ, Parienti JJ, Postema J, Loer SA, Oudemans-van Straaten HM, Girbes AR. Positive outcomes, mortality rates, and publication bias in septic shock trials. Intensive Care Med. 2018;44(9):1584–5.

7 Saugel B, Kouz K, Scheeren TWL. The '5 Ts' of perioperative goal-directed haemodynamic therapy. Br J Anaesth. 2019 Aug;123(2):103–7.

8 Shoemaker WC, Appel PL, Kram HB, Waxman K, Lee TS. Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients. Chest. 1988 Dec;94(6):1176–86.

9 Chong MA, Wang Y, Berbenetz NM, McConachie I. Does goal-directed haemodynamic and fluid therapy improve peri-operative outcomes? Eur J Anaesthesiol. 2018;35(7):469–83.

10 Bainbridge D, Martin J, Arango M, Cheng D. Perioperative and anaesthetic-related mortality in developed and developing countries: A systematic review and meta-analysis. Lancet. 2012;380(9847):1075–81.

11 Kalil AC, Sun J. Bayesian Methodology for the Design and Interpretation of Clinical Trials in Critical Care Medicine. Crit Care Med. 2014 Oct;42(10):2267–77.

12 Montenij LJ, Buhre WF, Jansen JR, Kruitwagen CL, De Waal EE. Methodology of method comparison studies evaluating the validity of cardiac output monitors: A stepwise approach and checklist. Br J Anaesth. 2016;116(6):750–8.

13 Peyton PJ, Chong SW. Minimally Invasive Measurement of Cardiac Output during Surgery and Critical Care. Anesthesiology. 2010 Nov;113(5):1220–35.

14 Jozwiak M, Monnet X, Teboul J-L. Pressure Waveform Analysis. Anesth Analg. 2017;126(6):1.

15 Mantha S, Roizen MF, Fleisher LA, Thisted R, Foss J. Comparing methods of clinical measurement: Reporting standards for bland and altman analysis. Anesth Analg. 2000;90(3):593–602.

16 Berthelsen PG, Nilsson LB. Researcher bias and generalization of results in bias and limits of agreement analyses: A commentary based on the review of 50 Acta Anaesthesiologica Scandinavica papers using the Altman-Bland approach. Acta Anaesthesiol Scand. 2006;50(9):1111–3.

17 Jellison S, Nolan J, Vo N, Thai M, Puljak L, Vassar M. Reporting of interventions used in anesthesiology trials: analysis using the Template for Intervention Description and Replication (TIDieR) checklist. J Clin Epidemiol. 2019 Nov;

18 Jospin M, Aguilar JP, Gambus PL, Jensen EW, Vallverdu M, Caminal P. Validation of the qCO cardiac output monitor during Valsalva maneuver. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2012. p. 240–3.

19 Peyton PJ, Wallin M, Hallbäck M. New generation continuous cardiac output monitoring from carbon dioxide elimination. BMC Anesthesiol. 2019;19(1):1–10.

20 Savovic J, Jones HE, Altman DG, Harris RJ, Pildal J, Als-nielsen B, et al. Research and Reporting Methods Influence of Reported Study Design Characteristics on Intervention. Ann Intern Med. 2012;157(6):429–38.

21 Teboul JL, Saugel B, Cecconi M, De Backer D, Hofer CK, Monnet X, et al. Less invasive hemodynamic monitoring in critically ill patients. Intensive Care Med. 2016;42(9):1350–9.

22 Anthon CT, Granholm A, Perner A, Laake JH, Møller MH. No firm evidence that lack of blinding affects estimates of mortality in randomized clinical trials of intensive care interventions: a systematic review and meta-analysis. J Clin Epidemiol. 2018;100:71–81.

23 Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome: systematic review with meta-analysis. Intensive Care Med. 2018;44(10):1603–12.

24 Zhang L, Dai F, Brackett A, Ai Y, Meng L. Association of conflicts of interest with the results and conclusions of goal-directed hemodynamic therapy research: a systematic review with meta-analysis. Intensive Care Med. 2018;44(10):1638–56.

25 Cecconi M, De Backer D, Antonelli M, Beale R, Bakker J, Hofer C, et al. Consensus on circulatory shock and hemodynamic monitoring. Task force of the European Society of Intensive Care Medicine. Intensive Care Med. 2014 Dec 13;40(12):1795–815.

26 De Backer D, Vieillard-Baron A. Clinical examination: a trigger but not a substitute for hemodynamic evaluation. Intensive Care Med. 2019;45(2):269–71.

27 Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, et al. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med. 2019 Feb;45(2):190–200.

28 Elder A, Japp A, Verghese A. How valuable is physical examination of the cardiovascular system? BMJ. 2016 Jul 27;354:i3309.

29 Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC Med Inform Decis Mak. 2016 Dec 3;16(1):138.

30 Krupat E, Wormwood J, Schwartzstein RM, Richards JB. Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors. Med Educ. 2017 Nov;51(11):1127–37.

31 Albanèse J, Leone M, Delmas A, Martin C. Terlipressin or norepinephrine in hyperdynamic septic shock: A prospective, randomized study. Crit Care Med. 2005;33(9):1897–902.

32 Hamzaoui O, Georger J-F, Monnet X, Ksouri H, Maizel J, Richard C, et al. Early administration of norepinephrine increases cardiac preload and cardiac output in septic patients with life-threatening hypotension. Crit Care. 2010;14(4):R142.

33 Scutari M, Auconi P, Caldarelli G, Franchi L. Bayesian Networks Analysis of Malocclusion Data. Sci Rep. 2017 Dec 10;7(1):15236.

34 Chon S-H, Hilgers S, Timmermann F, Dratsch T, Plum PS, Berlth F, et al. Web-Based Immersive Patient Simulator as a Curricular Tool for Objective Structured Clinical Examination Preparation in Surgery: Development and Evaluation. JMIR Serious Games. 2018 Jul 4;6(3):e10693.

35 Chon S-H, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, et al. Serious Games in Surgical Medical Education: A Virtual Emergency Department as a Tool for Teaching Clinical Reasoning to Medical Students. JMIR Serious Games. 2019 Mar 5;7(1):e13028.

36 Salmasi V, Maheshwari K, Yang D, Mascha EJ, Singh A, Sessler DI, et al. Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. Anesthesiology. 2017 Jan;126(1):47–65.

37 Varpula M, Tallgren M, Saukkonen K, Voipio-Pulkki L-M, Pettilä V. Hemodynamic variables related to outcome in septic shock. Intensive Care Med. 2005 Aug 23;31(8):1066–71.

38 Dünser MW, Takala J, Ulmer H, Mayr VD, Luckner G, Jochberger S, et al. Arterial blood pressure during early sepsis and outcome. Intensive Care Med. 2009;35(7):1225–33.

39 Scheer B, Perel A, Pfeiffer UJ. Clinical review: complications and risk factors of peripheral arterial catheters used for haemodynamic monitoring in anaesthesia and intensive care medicine. Crit Care. 2002 Jun;6(3):199–204.

40 Chatterjee A, Depriest K, Blair R, Bowton D, Chin R. Results of a survey of blood pressure monitoring by intensivists in critically ill patients: A preliminary study. Crit Care Med. 2010;38(12):2335–8.

41 Wax DB, Lin H-M, Leibowitz AB. Invasive and Concomitant Noninvasive Intraoperative Blood Pressure Monitoring. Anesthesiology. 2011;115(5):973–8.

42 Lehman LWH, Saeed M, Talmor D, Mark R, Malhotra A. Methods of blood pressure measurement in the ICU. Crit Care Med. 2013;41(1):34–40.

43 Saugel B, Grothe O, Nicklas JY. Error grid analysis for arterial pressure method comparison studies. Anesth Analg. 2018;126(4):1177–85.

44 Garland A. Arterial lines in the ICU: A call for rigorous controlled trials. Chest. 2014;146(5):1155–8.

45 Vieillard-Baron A, Millington SJ, Sanfillipo F, Chew M, Diaz-Gomez J, McLean A, et al. A decade of progress in critical care echocardiography: a narrative review. Intensive Care Med. 2019;45(6):770–88.

46 Price S, Via G, Sloth E, Guarracino F, Breitkreutz R, Catena E, et al. Echocardiography practice, training and accreditation in the intensive care: document for the World Interactive Network Focused on Critical Ultrasound (WINFOCUS). Cardiovasc Ultrasound. 2008 Dec 6;6(1):49.

47 Narula J, Chandrashekhar Y, Braunwald E. Time to add a fifth pillar to bedside physical examination inspection, palpation, percussion, auscultation, and insonation. JAMA Cardiol. 2018;3(4):346–50.

48 Wong A, Galarza L, Duska F. Critical Care Ultrasound. Crit Care Med. 2019 Mar;47(3):e256–62.

49 Merz TM, Cioccari L, Frey PM, Bloch A, Berger D, Zante B, et al. Continual hemodynamic monitoring with a single-use transesophageal echocardiography probe in critically ill patients with shock: a randomized controlled clinical trial. Intensive Care Med. 2019;45(8):1093–102.

50 Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014 Jan;383(9912):166–75.

51 Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009 Jul;374(9683):86–9.

52 Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. BMJ. 2015 Jun 11;350(1):87.

53 Labarère J, Bertrand R, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med. 2014;40(4):513–27.

54 Granholm A, Perner A, Jensen AKG, Møller MH. Important methodological flaws in the recently published clinical prediction model the REMEMBER score. Crit Care. 2019;23(1):4–5.

55 Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar 23;340(mar23 1):c332–c332.

56 Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014 Mar 7;348(mar07 3):g1687–g1687.

57 Ridgeon EE, Bellomo R, Aberegg SK, Sweeney R Mac, Varughese RS, Landoni G, et al. Effect sizes in ongoing randomized controlled critical care trials. Crit Care. 2017 Dec 5;21(1):132.

58 Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: Version 1.0. Trials. 2017;18(S3):1–50.

59 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007 Oct 20;370(9596):1453–7.

60 Hapfelmeier A, Cecconi M, Saugel B. Cardiac output method comparison studies: the relation of the precision of agreement and the precision of method. J Clin Monit Comput. 2016 Apr;30(2):149–55.

61 Vos JJ, Scheeren TWL. How to "validate" newly developed cardiac output monitoring devices. J Clin Monit Comput. 2016;30(2):147–8.

Nederlandse samenvatting

Dit proefschrift had als eerste doel om het bewijsmateriaal over de toepasbaarheid van hemodynamische monitoring uit te breiden, die gebruikt wordt om hemodynamische instabiliteit tijdens de perioperatieve periode en de opname op de IC te voorkomen en te behandelen. Ten tweede streefden we ernaar de kennis uit te bereiden over het verbeteren van de uitvoering van studies in de perioperatieve en intensive care geneeskunde.

Hemodynamische monitoring gebruikt in perioperatieve goal-directed therapy

Perioperatieve goal-directed therapy is meer dan een enkele interventie

Met behulp van hemodynamische monitoring kan de perioperatieve goal-directed therapy (PGDT) interventie uitgevoerd worden. Deze PGDT-interventie is reeds geëvalueerd in verschillende gerandomiseerde gecontroleerde studies (RCT's). In onze systematische review vonden wij het vanwege klinische heterogeniteit ongepast om een meta-analyse uit te voeren op basis van gegevens verkregen uit deze onderzoeken (***hoofdstuk 2***). Er waren meerdere aspecten van deze uitgebreide heterogeniteit waargenomen in de complexe PGDT-interventie, zoals het gebruik van verschillende hemodynamische variabelen als doel voor interventie en het gebruik van verschillende hemodynamische monitors. Gezien de bestaande normen voor kwalitatief hoogstaand bewijs was het niet gerechtvaardigd om een uniforme conclusie te trekken over het effect van PGDT op patiëntgerichte uitkomsten. We hebben de behoefte aangetoond naar universele en uniforme definities van de PGDT-interventie om een vergelijking tussen verschillende onderzoeken mogelijk te maken. Wanneer de resultaten van individuele RCT's wel worden samengevoegd in een meta-analyse, moet deze inherente heterogeniteit door de lezer worden overwogen voordat conclusies kunnen worden getrokken.

Hoewel we hebben aangetoond dat er geen kwalitatief hoogstaand bewijs beschikbaar was voor de PGDT-interventie, oftewel een meta-analyse die gegevens uit onderzoeken combineert met een laag risico op bias, moet een combinatie van verschillende aspecten worden overwogen om ervoor te zorgen dat de juiste patiënten profiteren van de PGDT-interventie. Ten eerste moet de PGDT-interventie worden toegepast bij hoog risicopatiënten die een hoog risico operatie ondergaan. Ten tweede moet de initiatie van de PGDT-interventie vroeg in de perioperatieve periode worden gestart. Ten derde moet het PGDT-algoritme flow optimalisatie, oftewel cardiac output of slagvolume, combineren met dynamische preloadvariabelen, bijvoorbeeld slagvolumevariatie (SVV), om de vloeistofrespons te beoordelen (***hoofdstuk 3***).

Hemodynamische monitoring in de intensive care

Gegevens verkregen uit de prospectieve observationele Simple Intensive Care Studies-I (SICS-I) werden gebruikt om het beschikbare bewijs met betrekking tot hemodynamische monitoring in de intensive care te vergroten. De primaire hemodynamische beoordeling wanneer een circulatoire shock wordt vermoed, kan worden beschreven als een stapsgewijze benadering waarbij aanvankelijk klinisch onderzoek wordt uitgevoerd, gevolgd door een meting van cardiac

output met behulp van echocardiografie, en afgesloten met geavanceerdere hemodynamische monitoring wanneer een complexe circulatoire shocktoestand dit vereist. In dit proefschrift hebben we het bewijsmateriaal voor elk van deze stappen uitgebreid.

Lichamelijk onderzoek is de eerste stap maar dient verbeterd te worden

We beoordeelden de huidige standaard van lichamelijk onderzoek bij kritisch zieke patiënten en brachten een Bayesiaans netwerk in kaart. Uit dit netwerk bleek dat artsen de aanwezigheid van een hoge dosis noradrenaline infusie en een verlengde capillaire refill tijd of mottling van de knie beschouwen als voorwaardelijke afhankelijkheden van een lage hartpompfunctie (***hoofdstuk 4***). Dit was de eerste keer dat een dergelijk Bayesiaans netwerk gebouwd werd voor dit doeleinde en daarom waren rigoureuze methoden en validatie noodzakelijk. Onze bevindingen geven inzicht in het mogelijke denkproces van artsen bij het uitvoeren van het lichamelijk onderzoek en tonen aan dat verbeteringen met betrekking tot de diagnostische nauwkeurigheid van het lichamelijk onderzoek voor hemodynamische beoordeling nodig zijn.

Verschillende methoden om bloeddruk te meten zijn niet uitwisselbaar

Bloeddrukmetingen verkregen met een non-invasieve oscillometrische bovenarm manchet vertoonden een te brede grens van overeenstemming in vergelijking met invasieve arteriële bloeddrukmeting. Error grid analyse toonde aan dat bij één op de vier patiënten die behandeld werd met noradrenaline er minstens een laag risico bestaat op nadelige veranderde behandeling indien gebruik gemaakt zou zijn van de niet-invasieve methode (***hoofdstuk 5***). Bloeddrukmetingen verkregen met een invasieve arteriële katheter zijn momenteel de klinische standaard bij patiënten met circulatoire shock, en onze bevindingen ondersteunen niet dat een non-invasieve meting als vervanging van deze klinische standaardmethode gebruikt kan worden. Wij zijn van mening dat het in de klinische praktijk essentieel is om kritisch te blijven op de uitwisselbaarheid van methoden voor bloeddrukmeting. De analyses die werden gebruikt om de twee bloeddrukmeetmethoden te vergelijken bieden beperkte gegevens over meetverschillen tussen verschillende methoden bij individuele patiënten. Voor de klinische praktijk betekent dit dat bij elke individuele patiënt de optimale toepasbare meetmethode moet worden bepaald op basis van de voor- en nadelen.

Experts blijven noodzakelijk om de beelden verkregen met echocardiografie juist te interpreteren

Geneeskundestudenten zijn in staat om echocardiografie beelden van voldoende kwaliteit te verkrijgen bij kritisch zieke patiënten in de intensive care, maar experts blijven noodzakelijk voor een juiste interpretatie van deze beelden (***hoofdstuk 6***). De kwaliteit van de beelden verkregen door de geneeskundestudenten was van voldoende kwaliteit, maar de metingen van cardiac output die door de geneeskundestudenten werden gemaakt op basis van deze beelden toonden te brede grenzen van overeenstemming. Dat wil zeggen dat er sprake was van onaanvaardbare nauwkeurigheid van de meetverschillen. De metingen werden gemaakt door de geneeskundestudenten en deze werden vergeleken met metingen door experts die gebruik maakten van dezelfde afbeeldingen. Het kunnen verkrijgen van de benodigde afbeeldingen en het kunnen uitvoeren van de cardiac output berekening is van essentieel belang, omdat het

gebruik van echocardiografie in de intensive care zal blijven toenemen. Onze bevindingen zijn interessant omdat het kunnen berekenen van cardiac output met behulp van echocardiografie beschouwd wordt als een geavanceerde vaardigheid en de geneeskundestudenten geen eerdere ervaring hadden met het gebruik van echocardiografie. Voor toekomstig onderzoek en klinische toepassing van echocardiografie is het mogelijk om beginners de beelden te laten verkrijgen die vervolgens door experts kunnen worden geïnterpreteerd, waardoor tijd en middelen worden bespaard.

Ongekalibreerde polsgolfanalyse om cardiac output te meten is niet geïndiceerd bij patiënten met circulatoire shock

Cardiac outputmetingen verkregen met een apparaat dat gebruik maakt van ongekalibreerde polsgolfanalyse toonden een te brede grens van overeenstemming en klinisch onaanvaardbaar trending vergeleken met echocardiografie bij patiënten met circulatoire shock (***hoofdstuk 7***). Op dit moment wordt de ongekalibreerde polsgolfanalysemethode voor het meten van cardiac output niet aanbevolen bij patiënten met circulatoire shock en onze bevindingen ondersteunen dit. Vanwege de gekozen methoden was het niet mogelijk om directe superioriteit of non-inferioriteit van de ene methode ten opzichte van de andere te bewijzen, omdat echocardiografie geen gouden standaard referentietechniek, en die is vereist voor een formele methodevergelijking. Vergelijkingen die gedaan worden in prospectieve observationele studies kunnen mogelijk de klinische praktijk beter weerspiegelen en maken validatie van deze methoden mogelijk in populaties van patiënten die aanvankelijk niet de doelpopulatie waren.

Initiatieven om methodologie en uitvoer van studies te verbeteren

Externe validatie van onderzoeksbevindingen in onafhankelijke cohorten is nodig

We presenteerden een overzicht van 43 geïdentificeerde mortaliteit predictiemodellen voor het voorspellen van sterfte bij willekeurige populaties van kritisch zieke patiënten. Deze modellen hadden een variërende kwaliteit van methodologie, en de validatie van

de individuele modellen verschilde onderling en was meestal onvolledig (***hoofdstuk 8***). Externe validatie van de mortaliteit predictiemodellen ontbreekt vaak, en een-op-een-vergelijkingen zijn nodig om uit de beschikbare mortaliteit predictiemodellen het best presterende model te identificeren voor het begeleiden van klinische zorg en onderzoek onder verschillende omstandigheden en in verschillende populaties.

Verplichte publicatie en peer-review van onderzoeksprotocollen

Peer-review van onderzoeksprotocollen door middel van publicatie voorafgaand aan de start van een studie kan de transparantie van uitvoer en verslaglegging van onderzoek vergroten (***hoofdstuk 9***). Wij zijn van mening dat een gestructureerde bespreking van de essentiële onderdelen voor start van een onderzoek, de discussie achteraf over de opzet van het onderzoek en de interpretatie van de onderzoeksresultaten kan beperken. Protocollen kunnen worden beoordeeld aan de hand van de checklist die bij het onderzoek ontwerp past, en wijzigingen kunnen worden aangebracht op basis van vastgestelde regels met een hoge validiteit. Opmerkelijke gebreken en weglatingen

kunnen worden geïdentificeerd, waardoor correcties op tijd mogelijk zijn. Er moet rekening worden gehouden met mogelijke vertragingen bij het begin van de studie en innovatieve ideeën staan open en kunnen door andere onderzoekers worden overgenomen. Transparantie en optimalisatie van onderzoeksopzet en uitvoer van studies komen niet alleen de onderzoekers, maar met name de patiënten ten goede, en dit is het uiteindelijke doel van al het uitgevoerde onderzoek.

List of publications

In this thesis

Kaufmann T, Clement RP, Scheeren TWL, Saugel B, Keus F, van der Horst ICC. Perioperative goal-directed therapy: A systematic review without meta-analysis. *Acta Anaesthesiol Scand. 2018 Nov 6;62(10):1340–55*.

Kaufmann T, Granholm A, Keus F, Myburgh J, Perner A, Møller MH, van der Horst ICC. Foresight over hindsight: Mandatory publication of clinical research protocols prior to conduct. *Acta Anaesthesiol Scand. 2019 Feb;63(2):267–9*.

Kaufmann T, Saugel B, Scheeren TWL. Perioperative goal-directed therapy – What is the evidence? *Best Pract Res Clin Anaesthesiol. 2019 Jun;33(2):179–87*.

Kaufmann T, Castela Forte J, Hiemstra B, Wiering MA, Grzegorczyk M, Epema AH, van der Horst ICC. A Bayesian Network Analysis of the Diagnostic Process and its Accuracy to Determine How Clinicians Estimate Cardiac Function in Critically Ill Patients: Prospective Observational Cohort Study. *JMIR Med Informatics. 2019 Oct 30;7(4):e15358*.

Kaufmann T, Clement RP, Hiemstra B, Vos JJ, Scheeren TWL, Keus F, van der Horst ICC. Disagreement in cardiac output measurements between fourth-generation FloTrac and critical care ultrasonography in patients with circulatory shock: a prospective observational study. *J Intensive Care. 2019 Dec 11;7(1):21*.

Keuning BE, **Kaufmann T**, Wiersema R, Granholm A, Pettilä V, Møller MH, Christiansen CF, Castela Forte JN, Snieder H, Keus F, Pleijhuis RGH, van der Horst ICC. Risk scores for mortality in the adult critically ill: A scoping review. *Acta Anaesthesiol Scand. 2020 Apr 26;64(4):424–42*.

Kaufmann T, Cox EGM, Wiersema R, Hiemstra B, Eck RJ, Koster G, Scheeren TWL, Keus F, Saugel B, van der Horst ICC. Non-invasive oscillometric versus invasive arterial blood pressure measurements in critically ill patients: a post hoc analysis of a prospective observational study. *J Crit Care. 2020 Jun;57:118–23*.

Koster G, **Kaufmann T**, Hiemstra B, Wiersema R, Dijkhuizen D, Vos ME, Hummel YM, Keus F, van der Horst ICC. Feasibility of cardiac output measurements in critically ill patients by medical students. *Ultrasound J. 2020 Dec 8;12(1):1*.

Not in this thesis

Wiersema R, Castela Forte JN, **Kaufmann T**, de Haas RJ, Koster G, Hummel YM, Koeze J, Franssen CFM, Vos ME, Hiemstra B, Keus F, van der Horst ICC. Observational Study Protocol for Repeated Clinical Examination and Critical Care Ultrasonography Within the Simple Intensive Care Studies. *J Vis Exp. 2019 Jan 16;(143):e58802*.

Hiemstra B, Koster G, Wiersema R, Hummel YM, van der Harst P, Snieder H, Eck RJ, **Kaufmann T**, Scheeren TWL, Perner A, Wetterslev J, de Smet AMGA, Keus F, van der Horst ICC. The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. *Intensive Care Med. 2019 Feb;45(2):190–200*.

de Keijzer IN, **Kaufmann T**, Scheeren TWL. Which type of fluid to use perioperatively? *J Emerg Crit Care Med. 2019 Oct;3(2):51–51*.

Hiemstra B, Eck RJ, Wiersema R, **Kaufmann T**, Koster G, Scheeren TWL, Snieder H, Perner A, Pettilä V, Wetterslev J, Keus F, van der Horst ICC. Clinical Examination for the Prediction of Mortality in the Critically Ill. *Crit Care Med. 2019 Jul;47(10):1301–9*.

Wiersema R, Koeze J, Eck RJ, **Kaufmann T**, Hiemstra B, Koster G, Franssen CFM, Vaara ST, Keus F, Van der Horst ICC. Clinical examination findings as predictors of acute kidney injury in critically ill patients. *Acta Anaesthesiol Scand. 2019 Sep 30;aas.13465*.

Koster TM, Wetterslev J, Gluud C, Jakobsen JC, **Kaufmann T**, Eck RJ, Koster G, Hiemstra B, van der Horst ICC, Keus F. Apparently conclusive meta-analyses on interventions in critical care may be inconclusive – a meta-epidemiological study. *J Clin Epidemiol. 2019;114:1–10*.

Saugel B, Critchley LAH, **Kaufmann T**, Flick M, Kouz K, Vistisen ST, Scheeren TWL. Journal of Clinical Monitoring and Computing end of year summary 2019: hemodynamic monitoring and management. *J Clin Monit Comput. 2020 Apr 14;34(2):207–19*.

Kaufmann T, van der Horst ICC, Scheeren TWL. This is your toolkit in hemodynamic monitoring. *Curr Opin Crit Care. 2020 Apr 22*.

Cox EGM, Koster G, Baron A, **Kaufmann T**, Eck RJ, Veenstra C, Hiemstra B, Wong A, Kwee TC, Tulleken JE, Keus F, Wiersema R, van der Horst ICC. Should the ultrasound probe replace your stethoscope? A SICS-I sub-study comparing lung ultrasound and pulmonary auscultation in the critically ill. *Crit Care. 2020 Dec 13;24(1):14*.

Wiersema R, Jukarainen S, Eck RJ, **Kaufmann T**, Koeze J, Keus F, Petilla V, van der Horst ICC, Vaara ST. Different applications of the KDIGO criteria for AKI lead to different incidences and mortality rates in critically ill patients: a post-hoc analysis from the prospective observational SICS-II study. *Crit Care. 2020 Dec 21;24(1):164*.

Submitted

van Beek DEC, Wiersema R, **Kaufmann T**, Hiemstra B, Eck RJ, Keus F, Scheeren TWL, van der Horst ICC. Validation of serum albumin as a prognostic marker in the intensive care unit: a SICS-I sub-study

Boekel MF, Venema S, **Kaufmann T**, van der Horst ICC, Vos JJ, Scheeren TWL. The effect of compliance to goal-directed therapy on outcome after high-risk surgery: a before-after study

Wiersema R, **Kaufmann T**, van der Veen H, de Haas R, Franssen C, Koeze J, van der Horst ICC, Keus F. Diagnostic accuracy of arterial and venous renal Doppler assessment for acute kidney injury in critically ill patients: a prospective study.

Wolters J, de Haas RJ, Eck RJ, **Kaufmann T**, Koster G, Koeze J, Franssen CFM, Wong A, Keus F, Wiersema R, van der Horst ICC. The practicality of measuring renal perfusion by ultrasound in a large cohort of critically ill patients

Curriculum Vitae

Thomas Kaufmann was born on August 11th, 1991, in Venlo and grew up in Roermond, in the southern part of the Netherlands. He attended the Stedelijk Lyceum in Roermond from 2003 until 2009 and then moved to the north to study medicine at the University of Groningen until his graduation in 2016. During his studies, Thomas participated in various organizational committees of the Medical Faculty Association Panacea and enjoyed a board year at Stichting Sociëteit Volonté, the medical faculty pub.

His interest in anesthesiology and intensive care medicine emerged due to medical internships at the University Medical Center Groningen (UMCG) and Medisch Spectrum Twente in Enschede. Towards the end of his studies, he became intrigued by clinical research thanks to a scientific clerkship with the Simple Intensive Care Studies research group at the Department of Critical Care of the UMCG under the supervision of Dr. Keus and Dr. van der Horst.

After his graduation in 2016, Thomas started working at the Department of Critical Care of the UMCG. He then worked as a researcher at the Department of Anesthesiology of the UMCG under the supervision of Prof. Dr. Scheeren from 2018 until the end of 2019, helping set up observational research studies and performing clinical trials. During this time, Thomas continued his research, which evolved into a formal Ph.D. trajectory. His area of interest is hemodynamic monitoring techniques and their clinical application.

Thomas presented several posters and had oral presentations at large international conferences such as the International Symposium on Intensive Care & Emergency Medicine (ISICEM) and Euroanaesthesia. He became an invited speaker at the annual conference of the European Society of Intensive Care Medicine (ESICM) and helped organize its annual Hemodynamic Monitoring Masterclass.

At present, Thomas has started his residency to become an anesthesiologist at the UMCG.

Dankwoord

De afgelopen jaren heb ik met veel plezier gewerkt aan dit proefschrift. Wat begon als een extra onderzoek klusje naast mijn stage wetenschap is uitgegroeid tot een promotietraject waarin ik met veel mensen heb mogen samenwerken en veel heb kunnen leren. Een aantal betrokken personen wil ik graag in het bijzonder bedanken.

Allereerst mijn promotoren, Prof. dr. T.W.L. Scheeren en Prof. dr. J.C.C. van der Horst.

Prof. dr. T.W.L. Scheeren, hartelijk dank voor de fijne samenwerking en de mogelijkheid om mijn onderzoek voort te kunnen zetten in een promotietraject. Daarnaast wil ik u bedanken voor de kansen en ruimte om me verder te ontplooien in de wetenschapswereld. Dankzij u heb ik kunnen deelnemen aan internationale congressen en de organisatie van de Hemodynamic Monitoring Masterclass. Ik heb altijd erg genoten van onze gezamenlijke reizen naar congressen of onderzoek overleggen en ik verheug me op onze verdere samenwerking.

Prof. dr. J.C.C. van der Horst, beste Iwan, wat heb ik een mazzel gehad dat ik bij jou en bij de SICS terecht gekomen ben. De inspiratie voor dit proefschrift en de afronding ervan heb ik aan jou te danken. Je hebt me laten zien dat je ook met weinig middelen mooie resultaten kunt bereiken als je maar goed nadenkt over wat je doet, samenwerkt, en plezier hebt. Dank dat je altijd toegankelijk bent voor overleg en voor je tientallen ideeën waarvan er altijd wel eentje tot succes kon worden gebracht. Ook op afstand is één overleg met jou voor mij motiverender en productiever geweest dan twee weken zelf werken. Ik had me geen betere promotor kunnen wensen.

De beoordelingscommissie bestaande uit Prof dr. J. Bakker, Prof. dr. W.F. Bühre, en Prof. dr. P.H.J. van der Voort wil ik hartelijk bedanken voor het beoordelen van dit proefschrift.

Het is fantastisch om deel uit te mogen maken van de SICS-onderzoeksgroep en ik wil graag het hele SICS-team bedanken: Roos Bleijendaal, Wouter Bult, Yasmin Cawale, Ramon Clement, Eline Cox, Devon Dijkhuizen, Ruben Eck, Jose Alves Castela Forte, Anja Haker, Bart Hiemstra, Casper Hilbink, Yoran Hummel, Britt Keuning, Eric Keus, Martiene Klasen, Manon Klaver, Jacqueline Koeze, Geert Koster, Thijs Koster, Laura Schokking, Victor Sikkens, Madelon Vos, Justin Woerlee en Renske Wiersema. We hebben met het team mooie mijlpalen bereikt en dat komt met name door de teamspirit, maar ook door ieders unieke inzet en expertise.

Een aantal van de bovenstaande personen wil ik in het bijzonder bedanken. Eric, met jouw pragmatische kijk op wetenschap heb je me geleerd om kritisch te zijn en goed na te denken over de uitvoer van onderzoek. Bedankt voor je begeleiding tijdens mijn stage wetenschap en voor de verdere ondersteuning van mijn onderzoek. Renske, dankzij jouw toewijding en aansturing heeft de SICS verder uit kunnen groeien en succesvol kunnen blijven. Het was heel fijn om in hetzelfde schuitje te zitten qua onderzoek en ik heb veel gehad aan je enthousiasme. Bart en

Ruben, SICS'ers van het eerste uur, dank voor jullie kritische blik op mijn artikelen. Eline, jouw onuitputtelijke energie was een steun in de rug om door te blijven werken. Geert, dank voor de lessen over echocardiografie. Ik ben vereerd om paranimf bij jouw promotie te mogen zijn.

De artsen en verpleegkundigen van de Intensive Care Volwassenen van het UMCG wil ik bedanken voor het ondersteunen van de SICS. Dankzij jullie behulpzaamheid konden er zoveel patiënten geïncludeerd worden. Wim Dieperink, Marisa Onrust en het ICV-researchteam wil ik bedanken voor de samenwerking.
Gedurende mijn stage wetenschap, semi-arts stage, en ANIOS-periode op de ICV was het heel plezierig om onderdeel uit te mogen maken van het team.

Ik heb een fantastische tijd gehad op de onderzoekskamer van de anesthesiologie en daar wil ik graag mijn medeonderzoekers en kamergenoten voor bedanken: Mariana da Gaya Costa, Gerbrich Galema, Michele Introna, Ilonka de Keijzer, Sowmya Muchukunte Ramaswamy, Hong Su, Celien Tigchelaar, Remco Vellinga, en Maud Weerink. Celien, mijn vaste maatje op de onderzoekskamer, dank je voor de gezelligheid en voor het sparren over onderzoek. Ik heb onze openhartige gesprekken altijd enorm gewaardeerd.

Verder wil ik nog een aantal personen van de afdeling anesthesiologie van het UMCG bedanken. Froukje Knotnerus, Martine Oostra, Sabine Reckel, Rob Spanjersberg, en Jaap Jan Vos. Froukje, samen hebben we een zeer ingewikkelde klinische studie voorbereid en mogelijk gemaakt. Ik vond het altijd heel fijn samenwerken, soms zelfs tot diep in de nacht als er weer eens een operatie uit liep en wij nog drie uur lang testjes en bloedafnames moesten doen.

Mijn bestuursgenoten van Bestuur 2011 der Stichting Sociëteit Volonté en tevens goede vrienden wil ik graag bedanken. Jefta van Dijk, Jorn Kalsbeek, Pieter Steinkamp, Marieke van der Velde, en Laura van Wijk. In onze studententijd heb ik met jullie een fantastisch bestuursjaar gehad waarin we niet altijd even serieus hoefden te zijn en alles skitterend of juist heel aal was. Gezien het feit dat we allemaal nog niet zo veel konden toen is het best bijzonder dat we allemaal met onderzoek bezig zijn gegaan. En over een paar jaar kunnen we zelfs met drie anesthesiologen en drie chirurgen mooi onze eigen kliniek openen.

Mijn ploeggenoten van Altijd Hard, de beste competitieroeiploeg die A.G.S.R Gyas ooit gekend heeft, wil ik graag bedanken. Henk Aarnink, Freek Apperloo, Mark Kloosterhuis, Harmen Gietema, Jeroen de Groot, Reitse Nijland, Jelle Vos, en Harmen Sjoerd Westra. Nu moet ik natuurlijk wel eerlijk noemen dat ik zelf niet heel veel geroeid heb, maar de borrels en feesten waren altijd fantastisch. Sinds 2009 hebben we trouw nog steeds iedere donderdag ploegavond met een standaard ongezonde maaltijd. Nu we burgers en arbeiders zijn hebben we onze smaak wat verfijnd en in dat kader wil ik ook de Happy Italy bedanken voor de vele vierkante meters pizza.

Mijn paranimfen heb ik te danken aan bovengenoemde vriendengroepen en graag zou ik ze nog in het bijzonder willen noemen. Pieter, direct vanaf het begin waren we dikke vrienden (toen ook letterlijk). Al snel kwamen we erachter dat we ook maar bestuursgenoten en huisgenoten moesten worden, en uiteindelijk hebben we dan ook in drie verschillende huizen samengewoond. Dank je voor je heerlijke humor. Mark, ik ben dankbaar voor onze vriendschap. Afspreken voor een borrel of een hapje eten is altijd heel ontspannend voor mij geweest. Jouw ontspannen kijk op allerlei dingen is altijd heerlijk relativerend. Fantastisch dat jullie straks aan mijn zijde zullen staan bij de verdediging van dit proefschrift.

Ik wil graag mijn ouders bedanken. Dankzij jullie heb ik alle kansen en volledige vrijheid gekregen om te worden wie ik ben geworden. Jullie onvoorwaardelijke steun en vertrouwen is altijd een enorme hulp geweest. Mijn broer, Jelle, wil ik graag bedanken voor de gezelligheid hier in het hoge noorden. Ik vind het ongelooflijk dapper dat je de stap hebt gezet om hier te komen studeren en ik ben heel erg trots op jou. Opa en oma, Wim en Jeanne, het is heel fijn om een hechte familie te hebben.

Nadine, jij bent het beste wat me ooit overkomen is. Ik hou van jou.