

End-tidal CO₂ in the diagnosis of fluid responsiveness – a systematic review

Emilie B. Øberg¹, Bent G. Jørgensen² & Preben G. Berthelsen³

ABSTRACT

INTRODUCTION: An indication of the adequacy of the intravascular volume is of importance in critically ill patients. The status of the intravascular volume can be determined from a fluid challenge test. Most tests involve invasive monitoring. An exception is the capnographic measurement of changes in end-tidal (ET) CO₂ after a fluid challenge. The method is appealing as it rests on solid physiological ground – the Fick principle and the Frank-Starling mechanism. Furthermore, it is non-invasive and convenient. We report the results of a systematic review of the merits of this method.

METHODS: After a registration with PROSPERO, we searched MEDLINE, EMBASE, the Cochrane Library database and trial registers for studies on the diagnostic accuracy of changes in ET-CO₂ in fluid responsiveness testing. Test sensitivity, specificity and area under the receiver operating characteristics curve (AUROC) were the primary outcome measures.

RESULTS: Seven papers met the inclusion criteria. The test was found to have a median sensitivity of 0.75 (range: 0.60-0.91) and a median specificity of 0.94 (range: 0.70-1.00). The median AUROC was 0.82 (range: 0.67-0.94); the diagnostic threshold was an increase in ET-CO₂ of 2 mmHg/5%.

CONCLUSIONS: Monitoring of ET-CO₂ during fluid responsiveness testing provides good diagnostic value with few false negative tests and fewer false positive tests. The included studies have important methodological flaws. It must therefore be acknowledged that the diagnostic value of ET-CO₂ monitoring found in the review is overrated and overrated to an unknown degree. Therefore, at the present state of affairs, implementation of the test cannot be considered evidence-based.

Many methods testing for intravascular volume deficits in critically ill patients are being promoted in the intensive care literature. Almost all methods are invasive or require patients to be ventilated in a non-protective manner. Consequently, most standard methods have a potential to physically harm patients. One exception is the monitoring of end-tidal (ET)-CO₂ changes after a fluid challenge. The method is non-invasive and easily applied as capnography is generally already in place in mechanically ventilated critically ill patients. Furthermore, the method has a sound physiological basis. It

relies on the Frank-Starling mechanism and Fick's 148-year-old – repeatedly validated – hypothesis that cardiac output can be estimated from the volume of carbon dioxide eliminated through the lungs divided by the difference in carbon dioxide content in mixed venous blood and arterial blood. So, if CO₂ production and ventilation are constant, there is no doubt that an increase in preload, causing an increase in cardiac output – and a concomitant decrease in alveolar dead space ventilation – will result in an increase in ET-CO₂. But can our standard CO₂-monitors reliably detect the small increase in ET-CO₂ that is to be expected, considering the limited increase in cardiac output seen after a passive leg raising manoeuvre (PLR) or volume infusion?

To answer this question, we conducted a systematic review of the diagnostic accuracy of changes in ET-CO₂ as an indicator of fluid responsiveness in critically ill adult patients in the intensive care unit (ICU) or the operating room (OR)

METHODS

The review was registered with PROSPERO (CRD42017074232) and the research was conducted as specified in the pre-review registration. The review was reported in accordance with the PRISMA-DTA Statement. Our aim was to include all studies investigating fluid responsiveness and ET-CO₂ monitoring in critically ill adult patients in the intensive care setting or in the OR.

We searched the Medline, EMBASE, Cochrane data-

SYSTEMATIC REVIEW

1) Department of Anaesthesiology, Centre of Abdominal Surgery, Rigshospitalet
2) Department of Anaesthesiology, Næstved Hospital
3) Danish Centre for the History of Anaesthesiology, Denmark

Dan Med J
2019;66(9):A5560

KEY POINTS

- ▶ The assessment of the status of the intravascular volume is essential in critically ill mechanically ventilated patients. The status is usually evaluated with a fluid challenge test. Monitoring of the test result normally requires invasive procedures.
- ▶ Monitoring of changes in end-tidal CO₂ during a fluid challenge test is convenient, non-invasive and theoretically appealing as it is based on solid physiological ground – the Fick principle and the Frank-Starling mechanism.
- ▶ A systematic review of the seven papers investigating the diagnostic accuracy of end-tidal CO₂ monitoring suggests that the method may be useful with only few false negative tests and fewer false positive test results, but methodological flaws in all seven papers must temper the enthusiasm.

TABLE 1

Study characteristics and precision of methods.

Reference	Type of investigation	Sample size, n: calculation	Fluid responders, n (%)	PLR or fluid infusion	Reference method	Precision/LSC	ET-CO ₂ method	Precision/LSC	Blinded outcome assessors
Lakhal et al, 2017 [3]	Prospective observational ICU	86: -	33 (38)	500 ml crystalloid/colloid	Transthoracic echocardiography Vivid S6/l (GE Healthcare)	9%/-	Mainstream infrared Servo I (Marquet) or Evita 4 (Dräger Medical)	-/-	No
Jacquet-Lagrèze et al, 2016 [4]	Prospective observational OR	40: +	15 (38)	500 ml HAES	Oesophageal Doppler HemoSonic 100 (Arrow International)	-/-	Sidestream infrared Infinity EtCO ₂ Microstream SmartPod (Draeger)	2.2%/3.2%	No
Toupin et al, 2016 [5]	Prospective observational OR	90: +	28 (31)	PLR	Pulmonary artery thermodilution Swan Ganz catheter	-/-	Mainstream infrared (Philips)	-/-	Yes
Xiao-ting et al, 2015 [6]	Prospective observational ICU	48: -	34 (71)	PLR	Transpulmonary thermodilution PICCO (Pulsion Medical Systems)	-/-	Mainstream infrared CO ₂ sensor M2741A (Philips)	-/-	No
Monnet et al, 2013 [7]	Prospective observational ICU	40: -	21 (53)	PLR	Transpulmonary thermodilution PICCO (Pulsion Medical Systems)	12%/-	Mainstream infrared CO ₂ sensor M2741A (Philips)	1.3%/1.8%	No
Young et al, 2013 [8]	Retrospective chart review ICU	34: -	24 (55)	PLR and/or 500 ml crystalloid/colloid	Bioreactance NICOM (Cheetan Medical)	-/-	Mainstream infrared Respirationics NM3 (Philips)	-/-	No
Garcia et al, 2012 [9]	Prospective observational ICU	37: -	21 (57)	500 ml crystalloid/colloid	Oesophageal Doppler CardioQ (Deltex Medical)	2.3%/3.2%	Sidestream infrared M-COVX (Datex-Ohmeda)	1.3%/1.85%	No

ET = end-tidal; HAES = health at every size; ICU = intensive care unit; LSC = least significant change; OR = operating room before surgery; PLR = passive leg raising.

bases and the following trials registries: ClinicalTrials.gov, EU Clinical Trials Register, ISRCTN Register, UMIN Clinical Trials Registry, Australia New Zealand Clinical Trials Registry, Nederland’s Trial Register and PROSPERO for ongoing or completed but unpublished or aborted investigations. (Latest search: April 9, 2019).

We used the search string (end-tidal carbon dioxide and fluid responsiveness) OR (capnography and fluid responsiveness) OR (end-tidal CO₂ and volume responsiveness) OR (end-tidal CO₂ and fluid responsiveness) OR (end-tidal CO₂ and PLR) OR (capnography and PLR) OR (carbon dioxide and PLR).

Titles and abstracts of papers found in the databases were read by all authors. Papers reporting on the sensitivity, specificity and/or the area under the receiver operating characteristics curve (AUROC) of changes in

ET-CO₂ during volume responsiveness testing were selected for further analysis.

Pertinent data from the selected studies were extracted by two of the authors, and disagreements were resolved in conference with the third. We extracted and recorded information on the number of tested patients, OR/ICU, volume infusion or PLR, percentage of fluid responders, study design (prospective, retrospective, calculated sample size, blinding of the investigators), descriptions of the reference and index methods and their precisions, and the sensitivity, specificity and AUROC for ET-CO₂ changes.

The STARD-2015 [1] 30-item score was calculated for each study to quantify the overall quality of the studies, and the QUADAS [2] 14-item tool was used to assess the risk of bias.

The primary outcome measures were the sensitivity,

specificity and the AUROC of changes in ET-CO₂ after a volume challenge.

RESULTS

The search in the databases and trial registries produced 37 potentially interesting items. After reading the abstracts, we rejected 30 of the retrieved papers. Two were in Chinese, one used alternating levels of positive end expiratory pressure and saline infusion to manipulate preload and one paper used ET-CO₂ as the reference test. A total of 26 papers did not inform on the use of ET-CO₂ in fluid responsiveness testing, leaving seven papers for a systematic review [3-9]. Searching the reference lists of the papers and the “Similar Articles” feature of PubMed produced no relevant papers not already retrieved from the databases. We found one letter-to-the-editor (vide infra) in the retrieved papers.

Study characteristics of the seven included studies are presented in **Table 1**. Five studies investigated critically ill patients in the ICU [3, 6-9] and two [4, 5] were performed in cardiac surgery patients after induction of anaesthesia, but before surgery. In the four studies [5-8] using PLR to increase preload, the change in ET-CO₂ was recorded one minute after leg elevation. In the three studies [3, 4, 9] in which preload was augmented by fluid infusion, ET-CO₂ was determined as soon as the infusions had been completed.

The sensitivity/specificity of changes in ET-CO₂ and the true positive, true negative, false positive and false negative values are shown in **Figure 1**. The median sensitivity was 0.75 (range: 0.60-0.91 (n = 5)) and the median specificity 0.94 (range: 0.70-1.00 (n = 5)).

The diagnostic accuracy of changes in ET-CO₂ and the STARD and QUADAS scores are shown in **Table 2**. The median AUROC was 0.82 (range: 0.67-0.94 (n = 7)). The median 30-item STARD-2015 score was 21 (range: 16-26) and the median 14-item QUADAS score 12 (range: 9-13).

Post hoc calculations (by the authors of the in-

cluded papers) indicated that an increase in ET-CO₂ of 2 mmHg (5%) or more was diagnostic of fluid responsiveness. None of the seven studies were industry sponsored or initiated, and none of the scientists involved in the studies reported having relevant conflicts of interest. No patient harm was inflicted by the performance of the index or reference methods.

DISCUSSION

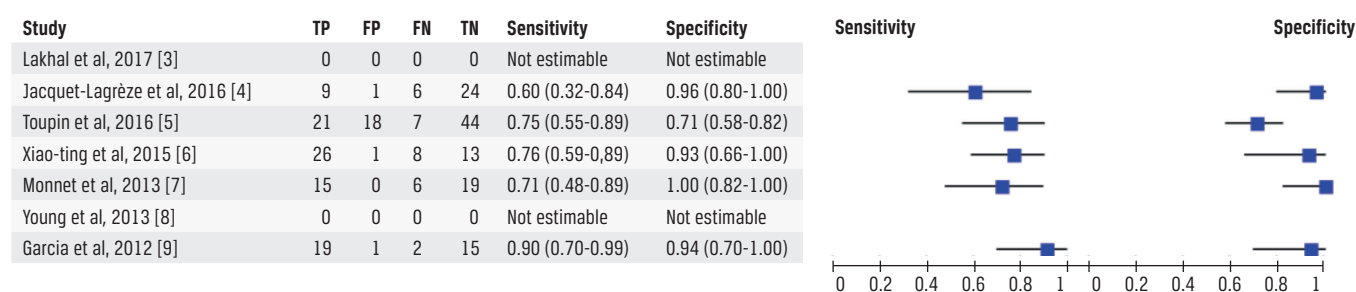
The results of the review suggest that monitoring ET-CO₂ after a fluid load has a good sensitivity (few false negative tests) and a high specificity (very few false positive tests) for prediction of fluid responsiveness in patients on mechanical ventilation, in the ICU or OR. The median AUROC of 0.82 (range: 0.67-0.94) also indicates that measurement of ET-CO₂ is of good diagnostic value (AUROC: 0.75-0.90) [10]. It seems that the many different commercially available and routinely used capnographs are able to detect the changes in ET-CO₂ expected after a volume challenge in hypovolaemic patients (Table 1).

A false negative test may mean that some patients will not receive an adequate volume substitution right away. Fortunately, there are other clinical signs of hypovolaemia (heart rate, blood pressure, diuresis), that may bail out the patient (and the clinician). More significant and essential is the very low risk of a false positive test as a false test result would suggest infusing volume in a patient who is already overhydrated (on or over the top of the Frank-Starling curve) – a scenario that may be harder to diagnose clinically.

It is of interest that cardiac arrhythmias (atrial fibrillation), mode of ventilation and/or the use of inotropes/vasopressors did not invalidate the method [3, 4, 7]. None of the studies evaluated the usefulness of ET-CO₂ monitoring during surgery. It is quite likely that the rapid haemodynamic changes seen perioperatively may invalidate monitoring of CO₂ changes as an indicator of fluid responsiveness.

FIGURE 1

Median values (95% CI) of sensitivity and specificity for a change in end tidal CO₂ after volume replenishment.



CI = confidence interval; FN = false negative; FP = false positive; TN = true negative; TP = true positive.

 **TABLE 2**

Diagnostic accuracy of changes in end-tidal CO₂ during fluid responsiveness testing.

Reference	AUROC, median (95% CI)	Diagnostic threshold	QUADAS-score	STARD-score
Lakhal et al, 2017 [3]	0.82 (0.73-0.90)	≥ 1 mmHg	12	21
Jacquet-Lagrèze et al, 2016 [4]	0.82 (0.67-0.97)	≥ 2 mmHg	13	26
Toupin et al, 2016 [5]	0.80 (0.70-0.90)	≥ 2 mmHg	11	20
Xiao-ting et al, 2015 [6]	0.85 (0.74-0.93)	≥ 5%	12	21
Monnet et al, 2013 [7]	0.93 (0.81-0.99)	≥ 5%	12	22
Young et al, 2013 [8]	0.67 (0.48-0.80)	-	9	16
Garcia et al, 2012 [9]	0.94 (0.82-0.99)	≥ 5%	12	21

AUROC = area under the receiver operating characteristics curve, CI = confidence interval; QUADAS = Quality Assessment of Studies of Diagnostic Accuracy; STARD = Standards for Reporting Diagnostic Accuracy Studies.

None of the studies reported patient-important outcomes. The studies have this in common with all other investigations, using diverse methods, to detect intravascular fluid deficits.

Bias, heterogeneity and other problems

It must be recognised that all of the included studies are biased. In only one study [4] were the results of the reference method and ET-CO₂ measurements interpreted without knowledge of the results of the other test; in only two [4, 5] studies was an a priori sample size calculation reported, and in none of the seven included studies was the number of discarded test results or loss of participants stated. These methodological deficiencies are all known to inflate estimates of diagnostic accuracy.

Furthermore, publication bias cannot be excluded. We searched many pre-trial registries without finding registrations of aborted or unpublished investigations. However, as long as it is not mandatory to register observational studies, we cannot be sure that investigations on the merits of ET-CO₂ have been conducted, but not published.

One of the studies reported rather discrepant and unfavourable results [8] – an AUROC of only 0.67 (95% confidence interval: 0.48-0.80); this is also the study with the lowest STARD and QUADAS scores (Table 2). The study is a retrospective database interrogation; and the reference method for measuring cardiac output – bioreactance (NICOM, Cheetah Medical) – is of questionable merits [11, 12]. In our view, the results of the study must be viewed with scepticism.

The validity of the 2015 study by Xiao-ting and colleagues [6] has recently been challenged in a letter by Mallat [13]. Mallat noticed that the pre-challenge haemodynamic data and SDs are identical in the three different scenarios where fluid responsiveness was tested (PLR and two infusion protocols). Xiao-ting has not responded to the letter and the paper has not been retracted.

The precision and accuracy of the reference method

is of importance. It is therefore a methodological shortcoming that only three studies [3, 7, 9] reported some measure of the reliability of their cardiac output measurements (Table 1). It may also present a problem that four different reference techniques with two different criteria for fluid responsiveness – a 10% or 15% increase in cardiac output/stroke volume – have been used as the precision and accuracy of the methods are known to differ to a clinically significant degree.

Ideally, volume responsiveness should be determined by the haemodynamic effect of a fluid infusion. In four of the studies [5-8], a PLR manoeuvre was instead used. The volume effect of PLR is not quantifiable, so it would have been ideal if the effect of the PLR manoeuvre had been verified by an infusion of fluid.

The PLR manoeuvre requires a repositioning of the patient which may influence pulmonary function and carbon dioxide excretion. In none of the four studies using PLR have the pulmonary effects of the manoeuvre been determined. To eliminate these confounding factors, we advocate that future studies should use volume infusion and not a PLR manoeuvre to increase preload.

The post hoc determined diagnostic thresholds for an increase in ET-CO₂ indicating fluid responsiveness were remarkably similar despite the heterogeneity of the studies and the fact that they were arrived at by different methods/calculations. Lakhal and colleagues [3] showed that an increase in ET-CO₂ of 1 mmHg had a positive likelihood ratio of > 5.0 for indicating volume responsiveness.

In other papers [4, 6, 7], the authors selected the ET-CO₂ increase that resulted in the minimal number of false negative and false positive tests. The last study [9] found that an increase in ET-CO₂ > 2 mmHg “during PLR was associated with a positive response to fluid infusion in all cases”.

The precision and least significant change that could be detected reliably during ET-CO₂ monitoring was determined in three of the studies [4, 7, 9]. The least detectable differences were in all cases well below

the post hoc defined diagnostic threshold of 5% (Table 1).

Future research

A randomised controlled study of ET-CO₂-guided clinical management with patient-important end-points (death, renal failure, length of stay ICU and days on mechanical ventilation) should be the next step forward. Such a study will require the inclusion of many patients, the cooperation of multiple centres and be time-consuming.

In the meantime, to increase the grade of evidence for a recommendation of using ET-CO₂-monitored fluid replenishment, there is a need for a prospective, adequately powered study with the best available reference method (thermodilution; transpulmonary or pulmonary artery cardiac output), with masked assessment of all measurements and a full disclosure of discarded measurements and excluded patients.

Furthermore, it may be of interest to determine if ET-CO₂ monitoring outperforms the other methods of detecting intravascular volume deficits. There is currently no registration of such studies in the trial registries.

One of the studies, included in this review [3], secondary analyses showed statistically and clinically significantly lower diagnostic accuracy of commonly used methods like pulse pressure variation, heart rate variation, blood pressure changes and changes in femoral artery blood flow.

CONCLUSIONS

The seven studies all have important methodological problems that exaggerate the diagnostic value of changes in ET-CO₂ during fluid responsiveness testing. It is not possible to quantify this exaggeration, but as the diagnostic value is quite high (AUROC: 0.82) and the test based on solid physiological ground, it is not unlikely that monitoring of ET-CO₂ during a fluid responsiveness test will be of value with few false nega-

tive tests – incorrectly curtailing volume substitution – and fewer false positive tests erroneously suggesting the need for volume infusion.

The physiological basis of the test is appealing, but it must, nevertheless, be acknowledged that implementation of the test today would not be evidence-based.

CORRESPONDENCE: Preben G. Berthelsen.

E-mail: p.g.berthelsen@dadlnet.dk

ACCEPTED: 3 June 2019

CONFLICT OF INTEREST: none. Disclosure forms provided by the authors are available with the full text of this article at Ugeskriftet.dk/dmj

LITERATURE

1. Bossuyt PM, Reitsma JB, Bruns DE et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
2. Withing P, Rutjes AWS, Reitsma JB et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Met* 2003;3:25.
3. Lakkhal K, Nay MA, Kamel T et al. Change in end-tidal carbon dioxide outperforms other surrogates for change in cardiac output during fluid challenge. *Br J Anaesth* 2017;118:355-62.
4. Jacquet-Lagrèze M, Baudin F, David JS et al. End-tidal carbon dioxide variation after a 100- and a 500-ml fluid challenge to assess fluid responsiveness. *Ann Intensive Care* 2016;6:37-45.
5. Toupin F, Clairoux A, Deschamps A et al. Assessment of fluid responsiveness with end-tidal carbon dioxide using a simplified passive leg raising maneuver: a prospective observational study. *Can J Anaesth* 2016;63:1033-41.
6. Xiao-ting W, Hua Z, Da-wei L et al. Changes in end-tidal CO₂ could predict fluid responsiveness in the passive leg raising test but not in the mini-fluid challenge test: a prospective and observational study. *J Crit Care* 2015;30:1061-6.
7. Monnet X, Bataille A, Magalhaes E et al. End-tidal carbon dioxide is better than arterial pressure for predicting volume responsiveness by the passive leg raising test. *Intensive Care Med* 2013;39:93-100.
8. Young A, Marik P, Sibole S et al. Changes in end-tidal carbon dioxide and volumetric carbon dioxide as predictors of volume responsiveness in hemodynamically unstable patients. *J Cardiothor Vasc Anesth* 2013;27:681-4.
9. Garcia MIM, Cano AG, Romero MG et al. Non-invasive assessment of fluid responsiveness by changes in partial end-tidal CO₂ pressure during a passive leg-raising maneuver. *Ann Intensive Care* 2012;2:9-18.
10. Ray P, Le Manach Y, Riou B et al. Statistical evaluation of a biomarker. *Anesthesiology* 2010;112:1023-40.
11. Kupersztynch-hagege E, Teboul JL, Artigas A et al. Bioreactance is not reliable for estimating cardiac output and the effects of passive leg raising in critically ill patients. *Br J Anaesth* 2013;111:961-6.
12. Lamia B, Kim HK, Severyn DA et al. Cross-comparisons of trending accuracies of continuous cardiac output measurements: pulse contour analysis, bioreactance, and pulmonary artery catheter. *J Clin Monit Comput* 2018;32:33-43.
13. Mallat J. A comment on "Changes in end-tidal CO₂ could predict fluid responsiveness in the passive leg raising test but not in the mini-fluid challenge test: a prospective and observational study". *J Crit Care* 2016;31:273.