Continuous glucose and exhaled breath analysis in the Intensive Care Unit

Leopold, J.H.

Citation for published version (APA):
This thesis covers studies on continuous glucose and exhaled breath analysis in the ICU. It is divided into three parts. First, we aimed to compare and test different (continuous) blood glucose measurement methods (part I). Second, we aimed to predict blood glucose levels by analysis of exhaled breath in intubated and ventilated ICU patients (part II). Finally, we aimed to investigate further development of exhaled breath analysis techniques and data analysis methods (part III). We hope that the results of the research in this thesis will enhance the understanding of (continuous) blood glucose measurements and exhaled breath analysis in ICU patients.
Continuous glucose and exhaled breath analysis in the Intensive Care Unit

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 26 oktober 2017, te 14.00 uur
door Jan Hendrik Leopold
geboren te Leiden
Promotiecommissie:

Promotores:  
prof. dr. M.J. Schultz  AMC-UvA  
prof. dr. A. Abu-Hanna  AMC-UvA  

Copromotor:  
dr. L.D.J. Bos  AMC-UvA  

Overige leden:  
prof. dr. N.P. Juffermans  AMC-UvA  
prof. dr. E.H.D. Bel  AMC-UvA  
prof. dr. A.H. Zwinderman  AMC-UvA  
prof. dr. T. Bein  Universität Regensburg  
dr. M.C. Schut  AMC-UvA  
dr. A.W. Boots  Universiteit Maastricht  

Faculteit:  Faculteit der Geneeskunde
Chapter 1  General introduction and outline of this thesis .......... pag 5

Part 1
Chapter 2  Continuous Glucose Monitoring–devices for Use in Intensive Care Units ........................................ pag 15
Chapter 3  Point Accuracy and Reliability of an Interstitial Continuous Glucose Monitoring Device in Critically Ill Patients: A Prospective Study ........................................ pag 31
Chapter 4  Point and Trend Accuracy of a Continuous Intravenous Microdialysis–based Glucose–monitoring Device in Critically Ill Patients – a Prospective Study.................................................. pag 55

Part 2
Chapter 5  Glucose Prediction by Analysis of Exhaled Metabolites – a Systematic Review ........................................ pag 79
Chapter 6  Factors influencing continuous breath signal in intubated and mechanically–ventilated intensive care unit patients measured by an electronic nose ......................................................... pag 99
Chapter 7  Non-invasive breath monitoring with eNose does not improve glucose diagnostics in critically ill patients in comparison to Continuous Glucose Monitoring in blood ......................................................... pag 119

Part 3
Chapter 8  Comparison of classification methods in breath analysis by electronic nose ........................................ pag 133
Chapter 9  Volatile organic compound profiles in outlet air from extracorporeal life-support devices differ from breath profiles in critically ill patients ................ pag 157
Chapter 10  Summary and General Discussion ..................... pag 173
Chapter 11  Nederlandse Samenvatting ............................... pag 181

Appendices
Author Contributions
Portfolio
Chapter 1

General introduction and outline of this thesis

Jan Hendrik Leopold
Decisions about the right treatment for the right patient and individual prognostic trajectories depend on an accuracy diagnosis. The goal of the diagnostic process is to determine an individual’s condition or disease based on relevant variables (such as signs and symptoms) of the patient. Innovations in the measurement of (new) variables and/or their interpretation can lead to important advances in medical diagnosis and hence healthcare in general. In this thesis, we investigate the utility of volatile organic compounds (VOCs) in exhaled breath as predictive variables for diagnostic outcomes in the intensive care unit (ICU). The focus is on an electronic nose (eNose) that can measure the VOCs in exhaled breath continuously.

Below we provide preliminaries on intensive care, blood glucose monitoring, exhaled breath analysis, and others. Next, we state our research questions and provide an outline of the thesis chapters.

**Intensive Care**
Patients are admitted to the intensive care unit if they are critically ill. Patients in the ICU are often intubated and mechanically ventilated, and vital signs are monitored constantly. Ideally, biological signals are also monitored continuously in this patient population, but most measurements can only be taken intermittently. The intermittent character of many biological tests limits their monitoring value on the intensive care unit and makes that this patient group may benefit greatly from technological advancements. The identification of a non-invasive alternative to blood draws is desirable as continuous monitoring requires large of amounts of measurements per patients.

**Blood glucose monitoring**
Alike patients with diabetes mellitus, critically ill patients require frequent monitoring of blood glucose levels. Due to critical illness, glucose metabolism deranges, which is treated with intravenous insulin administration. This potentially deadly intervention requires intensive monitoring of the blood glucose levels to prevent hypoglycemia. Often, glucose is measured intermittently, leaving periods of unmonitored insulin administration and the possibility of hypoglycemia. There is a possible benefit in monitoring glucose continuously to prevent hyper- and hypoglycemia [1,2]. Continuous glucose measurement (CGM) devices have been found safe and beneficial in patients with diabetes mellitus and are frequently used in that population [3]. Use of CGM in the ICU however, is still in its early stages. Therefore, thorough testing of accuracy and reliability of different methods for CGM in ICU patients is important.
Exhaled Breath
Exhaled breath is mainly made up of carbon dioxide, oxygen, nitrogen and water vapor [4]. In addition, a small percentage of breath is composed of thousands of volatile organic compounds (VOCs). In some cases, VOCs can be indicative for disease. The most famous example is the sweet odor of the breath of a person going through an episode of keto-acidosis [5]. In this metabolic process, low blood glucose levels lead to degradation of ketone bodies causing a rising concentration of acetone, which ends up in exhaled breath. While we know the origin of acetone and its connection to changes in glucose status, the origin of most VOCs is unknown. In addition to markers of systemic origin, VOCs can come from a pulmonary source. This could be due to inflammation or local metabolism of the lung, or could have exogenous sources such as inhaled VOCs or resident bacteria [6]. The field of exhaled breath analysis has been expanding quickly over the last years and has been investigating a growing number of health conditions and diseases. These include pulmonary conditions such as asthma and chronic obstructive pulmonary disease [7,8], lung cancer [8-11] and acute respiratory distress syndrome [12,13], and non-pulmonary conditions such as kidney disease [14] and pregnancy [15]. Although numerous studies have been conducted, few had led to technology being used in practice. Numerous breath analysis techniques are currently used. These include gas chromatography and mass spectroscopy (GC-MS) and electronic noses (eNose), which are discussed below.

Gas chromatography and mass spectroscopy
GC-MS is likely the most used breath analysis technique in clinical research and is considered the gold standard. Typically, breath is collected in a storage device (e.g. a bag or a sorbent tube) for a specified amount of time using a gas sampling pump. The breath sample is then introduced into the GC-MS device. Then, the sample is inserted in the gas chromatograph into a long column. Due to the differences in chemical properties, different molecules reach the end of the column after different periods of time. This is called the retention time of the compound. Then, these molecules are ionized and the fragment ions are detected by a mass spectrometer, resulting in a mass/charge ratio. The fragmentation pattern of a molecule in combination with the retention time can be used to identify compounds in the breath sample. Thus, with GC-MS, VOCs can be separated, quantified and identified to get a very detailed snapshot of the content of breath. However, because of the size of GC-MS machines, the time-consuming nature of GC-MS analysis and the expertise needed, GC-MS has a low clinical applicability. This holds especially for intensive care medicine as results are required as soon as possible.
Electronic nose
A typical eNose contains an array of sensors that are sensitive to a combination of molecules \[16\]. While metal oxide sensors are often used, optical sensors, conducting polymer sensors and surface or bulk acoustic wave sensors are also used in eNoses \[16\]. When molecules pass over eNose sensors and bind to them, the resistance of the sensors changes depending on the sensitivity to the molecule. The combination of the reactions of the sensors leaves a certain fingerprint, or breathprint, for each combination of molecules. Using pattern recognition techniques, these breathprints can be linked to certain events. eNoses are used in many fields and are actively investigated for use in clinical settings. While typical eNoses take single measurements, and are often used to diagnose a single event, continuous measurement is possible with adapted devices. In contrary to GC/MS, eNoses can be miniaturized, can provide continuous signals and analysis can be performed at bedside. Electronic noses are especially promising in the ICU setting since many patients in the ICU are intubated and mechanically ventilated, which provides constant and non-invasive access to the exhaled breath.

Continuous blood glucose measurement through eNose in the ICU
Following the clinical recognition of the acetone smell of a diabetic ketoacidosis, many researchers have studies exhaled breath analysis as non-invasive test for glucose control in diabetes patients and healthy controls. In general, these studies showed positive results on the predictive value of exhaled breath markers for blood glucose levels. However, these results have not been translated to the intensive care setting. A possible challenge with continuous exhaled breath analysis in intubated and mechanically ventilated ICU patients is the variance in ventilation modus in these patients, and the risk of noise being introduced to the signal due to the continuous nature of the measurements. Therefore, it must be investigated how continuous breath signals in critically ill patients should be handled with respect to de-noising and normalizing the signal. These cleaned signals should then be compared and correlated to blood glucose measurements.

Analysis challenges
In addition to the challenges in the analysis of continuous eNose signals described in the previous paragraph, it is not clear what the best approach for processing non-continuous eNose signals is. Researchers in the field use many different combinations of pre-processing, dimension reduction, classification and validation techniques. A summary of the used methods followed by an evaluation would shed a light on this topic. More importantly, many investigators do not validate their results. Generalizability of their findings therefore
remains unknown. Stressing the need for external validation by providing examples would stress its importance to the community.

**Analysis of VOCs through the extracorporeal circulation**

In some critically ill patients, pulmonary or cardiovascular conditions require extracorporeal circulation for oxygenation (extracorporeal membrane oxygenation (ECMO)) or CO2 removal (extracorporeal CO2 removal (ECCO2R)). This effectively generates a second circulation in which gas-exchange occurs. In this specific scenario, exhaled breath is not the only source of air that can be analyzed. The two air sources theoretically share VOCs that are produced throughout the body and are normally carried to the lung via the circulation. The main difference between the artificial gas exchange and the lung is that VOCs are also produced locally in the lungs. Comparison of the simultaneous signals from both sources allows for the identification of VOCs and breathprints that are generated locally as compared to systemically.

**HYPOTHESES AND RESEARCH QUESTIONS**

There are several hypotheses that will be tested in this thesis. First, we hypothesized that several devices can be used for continuous glucose monitoring in critically ill ICU patients with adequate accuracy and reliability. Second, we hypothesized that that continuous exhaled breath analysis using an eNose could be used to accurately predict blood glucose levels in intubated, mechanically ventilated patients. Finally, we hypothesized that there is a correlation between VOCs in exhaled breath and VOCs coming from extracorporeal support devices know to be of non-pulmonary origin. To test these hypotheses, we subdivided this thesis into three parts, accompanied by three different aims. First, we aimed to compare and test different (continuous) blood glucose measurement methods (part I). Second, we aimed to investigate the use of exhaled breath to monitor glucose levels (part II). Finally, we aimed to investigate further development of eNose measurement techniques and data analysis methods (part III). Figure 1 illustrates the outline of this thesis.

**Research questions**

The following research questions are used to test the hypotheses:

**RQ1.** What are the different techniques for CGM, and how should accuracy of these devices be assessed?

**RQ2.** What is the accuracy of two different CGM devices developed for use in the ICU?

**RQ3.** Is there an association between VOCs in exhaled breath and blood glucose levels?
RQ4. What are the sources of noise in continuously collected breath signals from ICU patients and how can these sources of noise be reduced?

RQ5. Can continuous exhaled breath analysis using an eNose be used to predict blood glucose levels in intubated ICU-patients?

RQ6. What are the data analysis techniques used in eNose studies and how do they compare?

RQ7. Is there a correlation between VOCs in exhaled breath and VOCs coming from extracorporeal support devices know to be of non-pulmonary origin?

Outline of this thesis

The above-stated research questions are answered in eight chapters, which are shortly summarized here. In chapter 2, an overview is given of the diverse continuous glucose monitoring (CGM) techniques and devices. In chapter 3 and 4 we hypothesized that two different CGM devices were point (and trend) accurate and reliable. In chapter 3, an interstitial CGM device is tested and in chapter 4, an intravenous CGM device is tested. Both studies were carried out in the same setting with a similar study design. In chapter 5 we hypothesized that there is an association between VOCs in exhaled breath and blood glucose levels. It encompasses a systematic review on the matter. The review focusses
on which exhaled metabolites correlate with changes in glucose levels. In chapter 6, we aim to investigate sources of noise in a continuous eNose breath signal measured in a non-controlled clinical setting. In addition, we discuss several approaches to reduce this noise. In chapter 7, we hypothesized that continuous exhaled breath analysis using an eNose could be used to accurately predict blood glucose levels in intubated, mechanically ventilated patients. We use the methods developed in chapter 6 and the data collected in chapter 3 & 4 in this study. In chapter 8, we compared classification methods in breath analysis by eNose used in literature. Our aim was to strengthen eNose research by evaluating which methods work best. These methods were tested on 4 different datasets we obtained from several researchers in the field of exhaled breath analysis. In chapter 9, we compared this data to exhaled breath data. The results described in this study are put into perspective in the discussion (chapter 10). The future directives are also discussed in that chapter.

References


2. Krinsley, J. S.; Preiser, J.-C. Moving beyond tight glucose control to safe effective glucose control. Crit Care 2008, 12, 149.


Chapter 2

Continuous Glucose Monitoring—devices for Use in Intensive Care Units

R.T.M. van Hooijdonk, J.H. Leopold, M.J. Schultz

INTRODUCTION

Many critically ill patients are treated with insulin for shorter or longer periods during their stay in the intensive care unit (ICU) [1]. Intensive monitoring of the blood glucose level is a prerequisite for both efficient and safe insulin titrations in these patients [2]. Presently in the ICU, glucose levels are monitored manually through intermittent measurements of the blood glucose level in central laboratories or using laboratory-based blood gas analyzers and/or glucose strips at the bedside [3]. Intermittent manual glucose monitoring, however, is impractical and expensive, time and blood consuming [4], and could even cause dangerous insulin titration errors in critically ill patients [5]. Glucose monitoring through so-called continuous glucose monitoring (CGM) could overcome some of the shortcomings and drawbacks of intermittent manual glucose monitoring. Specifically, CGM could allow for smoother insulin adjustments based on trends of the glucose level visualized on a monitor [3]. Several CGM devices for use in the ICU are being developed. These all require thorough accuracy testing in diverse cohorts of critically ill patient before they can be implemented in daily ICU practice.

This chapter provides an overview of diverse CGM techniques and CGM devices intended for use in the ICU. This chapter also deals with how point and trend accuracy of CGM systems could be studied in critically ill patients and how the accuracy results could be reported.

Search strategy

We searched MEDLINE (1966–2013) using the following search terms: ('intensive care'[MeSH Terms] OR 'intensive care'[tiab]) OR 'critical care'[MeSH Terms] OR 'critical care'[tiab] OR ('critical illness'[MeSH Terms] OR 'critical illness'[tiab]) AND 'glucose'[tiab] AND ('continuous glucose monitoring'[tiab] OR 'continuous glucose measurement'[tiab] OR 'CGM'[tiab]). Retrieved articles, and cross-referenced studies from those articles, were screened for pertinent information.

Articles were selected if they evaluated a CGM device intended to use in ICU patients. Articles reporting on studies in animals were excluded, as were articles reporting on studies of CGM in other than ICU populations. Revisions and articles that did not report outcomes of interest were also excluded, and in case we found duplicate articles of the same study in abstracts and articles, we report from the most complete data set.

Next, we performed an Internet-search, using similar search terms in GoogleTM. We visited commercial website identified by this search and looked for pertinent information. We also visited websites of medical congresses for information and abstracts of studies that were not yet published.

In August 2013, the two searches identified several CGM devices that were already available for use, as well as devices that were in a developmental phase (table...
1). Studies concerning CGM accuracy in critically ill patients were very limited, and the results of most studies were only available on commercial websites or presented in abstracts presented on medical congresses.

**CGM devices**

Common to all CGM devices is that they measure glucose levels continuously, or intermittently but frequently, though in different body fluids (i.e., whole blood, plasma, dialysate, or interstitial fluid) using dissimilar procedures (i.e., automated blood draws, or no blood draws at all) and distinctive measurement techniques (i.e., based on a chemical reaction, or using fluorescence or spectroscopy) (table 1). While measurement in plasma is considered the ‘gold standard’ for intermittent glucose measurements in the ICU setting, of all CGM devices only one device is reported to measure glucose levels in automated bedside–prepared plasma (OptiScanner). Other devices measure glucose levels in whole blood (GlySure, GluCath, and GlucoClear), dialysate from blood (Eirus and Diramo) or interstitial body fluids (Sentrino, Symphony, and GlucoDay).

CGM devices are reported to measure the glucose level in venous blood via a sensor inserted through a peripheral venous catheter (GluCath) or a central venous catheter (GlySure). Other CGM devices automatically draw venous blood via a central venous catheter (OptiScanner) or via a peripheral venous catheter (GlucoClear). For measurements of the glucose level in subcutaneous tissue, one single sensor or a set of sensors is used (Symphony, Sentrino). Systems that measure glucose levels in dialysate, prepare dialysate in a catheter designed for this purpose and inserted into a central vein (Eirus, Diramo) or into the subcutis (GlucoDay).

CGM devices measure glucose levels by using the glucose oxidase test (Eirus, Diramo, GlucoClear, Symphony, Sentrino and GlucoDay), fluorescence (GlySure and GluCath) or spectroscopy (OptiScanner). The glucose oxidase test is based on an enzymatic reaction, which uses glucose oxidase as a catalyst to bind glucose to water and oxygen to form gluconic acid and hydrogen peroxide. When there is more glucose, more hydrogen peroxide will be released, which can subsequently be measured [6]. The fluorescence technique is based on emission of light by a substance after absorbing light. Fluorescent chemistry is sensitive to glucose. When the glucose level increases, the fluorescent signal increases, which is detected with an optical fiber [7]. The spectroscopy technique is based on the characteristic absorption of vibrational nodes of different molecules, including glucose. Mid–infrared spectroscopy can be used because the glucose spectral peaks are in the mid–infrared region [8].

**Potential drawbacks**

Glucose levels in plasma are higher than in whole blood, demanding a conversion factor that depends on the hematocrit level [9]. Furthermore, arterial blood
glucose levels are higher compared to peripheral venous glucose levels (difference of ~ 0.2 mmol/L) and central venous glucose levels (difference of ~ 0.3 to 0.4 mmol/L) [10]. Glucose levels in dialysate tend to be slightly lower compared to glucose levels in surrounding fluids from which the dialysate is created [11]. Glucose levels in subcutaneous tissues are dependent on the speed by which glucose diffuses from the blood compartment to the interstitial spaces, as well as the rate at which glucose is taken up by cells in the subcutaneous compartment [12]. Users may take these drawbacks into account when using the GCM devices in daily practice, but researchers certainly will need to correct for this when determining GCM accuracy.

A potential disadvantage of any biosensor is the buildup of body fluid deposits on sensor surfaces, for which repeated calibrations and eventually sensor replacements are needed [13]. Need for repeated replacements of (parts of the) system is not limited to sensor–based devices, though, as all CGM devices need replacement of other parts of the system, such as cartridges, and/or dialysate–membranes. Furthermore, with the exception of CGM using a transdermal sensor (Symphony), all CGM devices must be considered ‘invasive’, and as such could cause infections and/or bleedings. Additionally, all CGM devices that measure the blood glucose level in a vein are at risk of presenting erroneous glucose levels when glucose, or other substances that interfere with the measuring technique, are infused through the same catheter or close to that catheter. Finally, the oxygen level and the pH could affect measurements by both the glucose oxidase test and the fluorescence technique [6, 7]. Drugs could interfere the glucose oxidation reaction through molecules oxidizing with hydrogen peroxide [6] and mid–infrared spectroscopy through spectrum of molecules other than glucose [8]. Users need to be aware of these drawbacks when using the GCM devices in their practice.

**Point and/or trend accuracy**

All CGM devices need accuracy testing in cohorts of patients in which they will be used. Two different types of accuracy can be tested, ‘point accuracy’, and ‘trend accuracy’. Point accuracy is the accuracy of intermittent measurements at a static point. Trend accuracy is the accuracy to detect changes in glucose levels.

Several point accuracy metrics have been used to report accuracy, including correlation coefficients, mean absolute difference (MAD) or mean absolute relative difference (MARD), and Bland–Altman plots [14, 15]. A high correlation coefficient (close to 1 or -1) means that paired glucose measurements (measurement by the device versus measurement by a reference test) lay along any straight line – but this line may not lay along the line of equality were differences between paired measurements are zero. Both MAD and MARD summarize
all paired glucose measurements in a single number, but unfortunately causes loss of important information. Another frequently used metric to demonstrate point accuracy is presenting all collected paired glucose measurements, with bias (the mean overall difference between the paired measurements) and limits of agreement (mean difference ± 1.96 * standard deviation) in Bland–Altman plots [16]. Reports on studies testing the accuracy of home glucose meter commonly use so–called Clark–error grids (CEG) (figure 1). A CEG visualizes information by presenting all collected paired glucose measurements, ‘scoring’ clinical accuracy [17]. For this, a CEG is divided in 5 paired ‘zones’: zones A (measurement within 20% of the reference or glucose levels < 70 mg/dL); zones B (measurement more than 20% different from the reference but still clinically acceptable as they would not change the rate of insulin infusion); zones C (measurement that would lead to unnecessary changes in insulin infusion, i.e., overcorrecting acceptable glucose levels); zones D (potentially dangerous hypo– or hyperglycemic events are missed); and zones E (levels that would lead to a decision opposite to that required, i.e., treatment for hypoglycemia instead of hyperglycemia). General consensus is that 95% of the values should be in zones A and 5% in zones B [14]. It must be noted that the CEG is originally designed for testing accuracy of home glucose meters, not ICU meters. At the moment, it is uncertain whether the CEG zones are useful in the ICU setting. As an alternative to the CEG, an insulin titration–error grid has been proposed [18]. In this grid, very much alike the original CEG, accuracy zones are based on a specific guideline for insulin titration. As guidelines for insulin titration differ (extensively) between ICUs worldwide, it could be difficult to compare results of accuracy testing of CGM devices using these grids.

R–deviation (RD) and absolute R deviation (ARD) have been proposed as rate accuracy metrics [15]. RD is defined as the difference between rates of change of measurements by the device and the reference test, divided by the time interval [15]. The ARD is the absolute value of RD [15]. Unfortunately, like in MAD and MARD, only reporting RD or ARD causes loss of important information.

More recently, the ‘continuous glucose–error grid analysis’ (CG–EGA) has been proposed for testing rate accuracy of CGM devices (figure 2) [19]. The CG–EGA combines point accuracy with rate accuracy though a rate error grid, a point error grid, and an error matrix. The rate error grid plots the rate of change of the glucose level measured by the CGM device and the reference test. A bit similar to original CEG, the rate error grid is divided in 5 paired ‘zones’: zones Ar (rate) (the accurate zone) and Br (the benign error zone), in these zones errors do not cause inaccurate adjustments; zone Cr (over- or underestimation of the
rate of change); zones Dr (reference test detects a change, which is undetected by the CGM device); and zones Er (reference test detects a change, but an opposite change is detected by the CGM device). The point error grid looks like the original CEG, but also takes glucose changes into account. Indeed, in this adjusted grid, zones are defined depending on the speed of change of glucose levels. When there is no significant glucose change, zones are similar to the original CEG, but when reference glucose levels are declining the upper limits change, and when reference glucose levels are increasing the lower limits change. Finally, results from the point and error grids are to be put into an error matrix with three regions, one for hypoglycemic range, one for normoglycemic range and one for hyperglycemic range. The CG–EGA is a complex tool. Also, the creation of a CG–EGA requires (very) frequent sampling to come to meaningful conclusions. However, one should keep in mind that the rate of sampling has an important effect on the results [20].

An alternative for the CG–EGA could be the polar plot, originally developed for testing trend accuracy of cardiac output monitors (figure 3) [21]. A polar plot shows the agreement between measurements by a device and measurements by a reference test as the angle made with the line of identity (where the difference between the measurements is zero) and the magnitude of change as the radian [21]. This way of accuracy testing, however, so far has not yet been used for testing accuracy of CGM devices.

**Reported accuracies of CGM devices**

Studies on point accuracy of CGM devices for use in the ICU are very scarce (table 2). The search in Medline identified only two point accuracy studies in ICU patients (Eirus [11] and GlucoDay [22]). The Internet-search identified several point accuracy studies presented as abstracts (Glysure [23], GluCath [24], Symphony [25] and Sentrino [26]) or on commercial website (OptiScanner [27], Diramo [28] and GlucoClear [29]). Most studies were rather small regarding the number of patients as well as the number of paired measurements. Notably, accuracy was sometimes only tested in 'less severely ill' patient population, i.e., patients in the ICU after (cardiac) surgery [11, 25, 26, 28, 29].

Two studies tested trend accuracy (GlucoDay [22], Symphony [25]). In the study of GlucoDay, a paired sample was obtained in 5 medical ICU patients every 15 minutes. The error matrix of the CG–EGA showed all samples in the hypoglycemic range in zone A, in the hyperglycemic range 88% in zone A and B and in the normoglycemic range 94% in zone A and B [22]. In the study of Symphony in post–cardiac surgery patients, paired samples were obtained only every 30 – 60 minutes. Although not specified for the range of glucose levels, 100% of the samples were in the A and B zone [25].
DISCUSSION

The most frequently suggested potential benefit of CGM in ICU patients is a reduction in time spent by nurses measuring glucose level [30]. Whether CGM truly reduces time spent on glucose monitoring, however, has not yet been demonstrated. CGM devices could indeed reduce the number of manual measurements. However, initiation, repeated manual calibrations and replacement of (parts of) the system could also use up nursing time. Whether time spent with using CGM weighs against the burden of intermittent manual measurement in central laboratories or using laboratory–based blood gas analyzers and/or strips at the bedside could be subject of future studies.

Intermittent manual glucose monitoring is usually seen as expensive [4]. It is questionable, however, whether use of CGM will reduce costs associated with glucose monitoring. Indeed, CGM device will come at a price, as do the disposables used with these devices. Costs for glucose monitoring should never be considered in isolation, but together with potential financial benefits and other healthcare costs (e.g., cost prevented by reducing the incidence of dysglycemia). Therefore, health–economy analyses could accompany future studies of CGM in critically ill patients.

It has been suggested that CGM could prevent dangerous insulin titration errors in critically ill patients [5]. One trial of glucose control confirmed that CGM prevents hypoglycemia, but overall glucose control did not improve [31]. One trial of closed–loop CGM–insulin titration showed improved glucose control, though [32]. Of note, these two trials used a home CGM device and still, frequent intermittent manual glucose measurements were necessary.

As yet, the number of studies assessing accuracy of CGM devices is surprisingly small. In addition, numbers of patients studied in each investigation are low and most studies are performed only in a highly selected ICU population (i.e., patients after cardiac surgery). It could be questioned if accuracy is also good in the ‘more severely ill’ patients, such as patients with severe sepsis or septic shock.

Point accuracy of some CGM devices is low. The question is whether such CGM devices are useless in the ICU setting. One advantage of CGM is that there will be many more glucose readings than with manual intermittent glucose monitoring. Thereby the user could detect trends, and trend accuracy could be more important than point accuracy. An analogy that supports use of CGM devices with poor point accuracy is the comparison between camcorders versus still cameras, as previously pointed out by Kovatchev et al. [19]: ‘Still cameras produce highly accurate snapshots at random sparse points in time, and camcorders generally offer lower resolution of each separate image but capture the dynamics of the action. Thus, it would be inappropriate to gauge the accuracy of still cameras and camcorders using the same static measure of the number
of pixels in a single image. Similarly, it is inappropriate to gauge the precision of [...] devices using the same measures and to ignore the temporal characteristics of the observed process.'

CONCLUSION

Implementation of CGM devices in daily ICU practice is at hand. Several CGM devices, using different body fluids and diverse sample and measuring techniques have been or are currently developed. These all need accuracy testing. The number of studies assessing accuracy of CGM devices is still limited, and most studies so far only included low numbers of highly selective ICU patient population.

References


27. Optiscan, (2013) OptiScan Biomedical Presents Study Results Demonstrating OptiScanner’s Excellent Glucose Monitoring Accuracy in Critically Ill ICU Patients. Press Release 21 March


Figure 1.
The Clarke Error Grid – paired measurements are plotted; measurements are most accurate in zones A and least accurate in zone E, where measurements are erroneous. See text for more details.
Figure 2.
The Continuous Glucose Error Grid with the ‘Rate Error Grid’ (panel A), the ‘Point Error Grid’ (panel B) and the ‘Error Matrix’ (panel C) – the Rate Error Grid is divided in zone Ar – Er, with Ar being the most accurate zone and Er being the least accurate (erroneous) zone; the Point Error Grid has similar zones as the Clark Error grid (figure 1), but the limits depend on rates of change: when there is no significant glucose change, zones are similar to the original CEG; with declining reference glucose levels upper limits change; with increasing reference glucose levels the lower limits change (see arrows in panel B); the results of the point error grid and rate error grid are put into an error matrix (Panel C) with 3 zones; accurate readings , benign errors (/\/) and erroneous readings (=). See text for details.
Figure 3.
A Polar plot. The 4 panels indicate how a polar plot is constructed. Panel A shows 4 paired glucose measurements. Panel B visualizes the same measurements with CGM measurements on the Y–axis and reference test measurements on the X–axis (note that the solid dots and squares in panel A represent the same measurements as solid triangles in panel B); a line is drawn between the consecutive measurements. In panel C, the difference between two consecutive readings (or the rate of change) by the CGM device is plotted on the Y–axis against the difference (or the rate of change) between two readings by the reference test on the X–axis (note how the rates of change make a particular angle with the line of identity, which is the line where the rate of change detected by the CGM device and by the reference test is the same). The radius is calculated as the mean of the rates measured by the CGM device and the reference test (dots in panel C). The angle with the line of identity is one coordinate in the polar plot with the radius being the other coordinate. The transformation to the polar plot is made in panel D, with the black dot representing the same black dot in panel C. Measurements with a large angle, i.e., a large difference between the rate of change measured with CGM and the reference test, are less accurate. Criteria for defining good and poor trend accuracy for the polar plot are uncertain.
<table>
<thead>
<tr>
<th>Device Name</th>
<th>Manufacturer</th>
<th>glucose is measured in:</th>
<th>sampling technique</th>
<th>measurement technique</th>
<th>frequency of measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>OptiScanner</td>
<td>OptiScan Biomedical Corporation</td>
<td>plasma created from central venous whole blood</td>
<td>blood drawn via a central venous catheter</td>
<td>mid-infrared</td>
<td>15 minutes</td>
</tr>
<tr>
<td>GlySure</td>
<td>GlySure Limited</td>
<td>central venous whole blood</td>
<td>blood sensor-based</td>
<td>fluorescence</td>
<td>not reported</td>
</tr>
<tr>
<td>Eirus</td>
<td>Maquet</td>
<td>central venous whole blood</td>
<td>dialysate</td>
<td>glucose oxidase</td>
<td>1 minute</td>
</tr>
<tr>
<td>Diramo</td>
<td>Flowsion AS</td>
<td>central venous or peripheral arterial whole blood</td>
<td>dialysate</td>
<td>glucose oxidase</td>
<td>1 second</td>
</tr>
<tr>
<td>GluCath</td>
<td>GluMetrics Incorporated</td>
<td>peripheral venous or arterial whole blood</td>
<td>blood sensor-based</td>
<td>fluorescence</td>
<td>10 seconds</td>
</tr>
<tr>
<td>GlucoClear</td>
<td>Edwards Lifesciences</td>
<td>peripheral venous whole blood</td>
<td>blood sensor-based</td>
<td>glucose oxidase</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Symphony</td>
<td>Echo Therapeutics</td>
<td>interstitial fluid</td>
<td>transdermal biosensor</td>
<td>glucose oxidase through hydrogel</td>
<td>1 minute</td>
</tr>
<tr>
<td>Sentrino</td>
<td>Medtronic Incorporated</td>
<td>interstitial fluid</td>
<td>fluid sensor-based</td>
<td>glucose oxidase</td>
<td>not reported</td>
</tr>
<tr>
<td>GlucoDay</td>
<td>A. Menarini Diagnostics</td>
<td>interstitial fluid</td>
<td>dialysate</td>
<td>glucose oxidase</td>
<td>3 minutes</td>
</tr>
</tbody>
</table>

Table 1
Overview of CGM–devices intended for use in ICU patients
<table>
<thead>
<tr>
<th>Device</th>
<th>Patients studied</th>
<th>Number of paired samples</th>
<th>Point accuracy</th>
<th>Trend accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>OptiScanner [27]</td>
<td>63 medical or surgical ICU patients</td>
<td>Not reported</td>
<td>CEG analysis: 100% in zones A+B</td>
<td>–</td>
</tr>
<tr>
<td>GlySure [23]</td>
<td>16 medical or surgical ICU patients</td>
<td>296</td>
<td>CEG analysis: 99.6% in zones A+B</td>
<td>–</td>
</tr>
<tr>
<td>Eirus [11]</td>
<td>30 peri–operative cardiac surgery patients</td>
<td>607</td>
<td>CEG analysis: 100% in zones A+B; MARD 5.6%</td>
<td>–</td>
</tr>
<tr>
<td>Diramo [28]</td>
<td>10 post–operative cardiac surgery patients</td>
<td>359</td>
<td>CEG analysis: 100% in zones A+B; MARD 4.1%</td>
<td>–</td>
</tr>
<tr>
<td>GluCath [24]</td>
<td>5 post–operative cardiac surgery</td>
<td>202</td>
<td>MARD 5.5 %</td>
<td>–</td>
</tr>
<tr>
<td>GlucoClear [29]</td>
<td>10 post–operative surgical ICU patients</td>
<td>1393</td>
<td>CEG analysis: 100% in zones A+B; MARD 5.0 %</td>
<td>–</td>
</tr>
<tr>
<td>Symphony [25]</td>
<td>15 peri–operative cardiac surgery</td>
<td>&gt; 600</td>
<td>MARD 12.3%</td>
<td>–</td>
</tr>
<tr>
<td>Sentrino [26]</td>
<td>21 post–operative cardiac surgery patients</td>
<td>864</td>
<td>MARD 12.2</td>
<td>–</td>
</tr>
<tr>
<td>GlucoDay [22]</td>
<td>50 medical ICU patients (5 patients were included for trend accuracy)</td>
<td>N = 820 (for 2–point calibration) N = 555 (for 6–point calibration)</td>
<td>CEG analysis: 95% in zones A+B for 2 point calibration 97% in zones A+B for 6 point calibration; CG–EGA: 99.6% in zones A+B of the error matrix (not specified for which range)</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 2. Overview of studies of CGM–devices in ICU patients. Abbreviations: ICU, intensive care unit; CEG, Clark error grid; MARD, mean absolute relative difference, CG–EGA, continuous glucose error grid; –, no data
Chapter 3

Point Accuracy and Reliability of an Interstitial Continuous Glucose Monitoring Device in Critically Ill Patients: A Prospective Study


Critical Care 2015, 9
ABSTRACT

Introduction: There is need for continuous glucose monitoring in critically ill patients. The objective of this trial was to determine the point accuracy and reliability of a device designed for continuous monitoring of interstitial glucose levels in intensive care unit patients.

Methods: We evaluated point accuracy by comparing device readings with glucose measurements in arterial blood using blood gas analyzers. Analytical and clinical accuracy was expressed in Bland–Altman plots, glucose prediction errors, and Clarke error grids. We used a linear mixed model to determine which factors affect the point accuracy. In addition, we determined the reliability, including duration of device start-up and calibration, skips in data acquisition, and premature disconnections of sensors.

Results: We included 50 patients in whom we used 105 sensors. Five patients from whom we could not collect the predefined minimum number of four consecutive comparative blood draws were excluded from the point accuracy analysis. Therefore, we had 929 comparative samples from 100 sensors in 45 patients (11 [7–28] samples per patient) during 4,639 hours (46 [27–134] hours per patient and 46 [21–69] hours per sensor) for the accuracy analysis. Point accuracy did not meet the ISO14971 standard for insulin dosing accuracy, but improved with increasing numbers of calibrations, and was better in patients who did not have a history of diabetes. Out of 105 sensors, 60 were removed prematurely for a variety of reasons. The device start-up time was 49 [43–58] minutes. The number of skips in data acquisition was low, resulting in availability of real-time data during 95 [89–98]% of the connection time per sensor.

Conclusion: The point accuracy of a device designed for continuous real-time monitoring of interstitial glucose levels was relatively low in critically ill patients. The device had few down times, but one third of the sensors were removed prematurely because of unresolved sensor – or device related problems.

INTRODUCTION

Handheld blood glucose meters or department-based blood gas analyzers are currently the preferred methods to measure blood glucose levels in intensive care unit (ICU) patients [1, 2]. These intermittent glucose monitoring techniques have variable accuracies [3], but foremost lack useful trending because of the interval between consecutive measurements. Continuous glucose monitoring (CGM) is suggested to increase practicalities and safety of insulin titration in ICU patients [4, 1], in particular when targeting normal or near-normal blood glucose levels when hypoglycemic episodes can be expected [5–13]. Glucose oxidase technique-based interstitial CGM devices have been used
before in diabetic patients outside the ICU setting [14]. It is uncertain, however, whether interstitial CGM devices are point accurate in critically ill patients [1]. An altered relationship between blood and interstitial fluid glucose levels during critical illness could affect the point accuracy of interstitial CGM to reflect the blood glucose level [15]. Several interstitial CGM sensor systems originally designed for non–ICU patients have been tested in the ICU setting in recent years [16–28]. Medtronic MiniMed (Northridge, CA) developed the Sentrino Continuous Glucose Management System, an interstitial CGM device that was especially designed for use in critically ill patients. This device was improved from previous models by creating the processor cable and pole-mounted monitor, and by four sensing elements designed to increase responsiveness to glucose changes and to limit the influence from drug interactions. The aim of this study was to test its point accuracy and reliability in a mixed medical–surgical ICU. We hypothesized that the device would provide an accurate reflection of the blood glucose level in ICU patients treated according to a local guideline for blood glucose control targeting blood glucose levels between 90–144 mg/dL. In addition, we determined its reliability, including duration of the device start–up, the need for calibration, skips in data acquisition and number and reasons for premature disconnections.

METHODS

Study design and informed consent
This was an investigator–initiated observational trial. The Institutional Review Board of the Academic Medical Center (Amsterdam, The Netherlands) approved the study protocol (study ID: NL41498.018.12). Medtronic MiniMed provided three devices for the duration of the trial and the necessary sensors, but had no influence on study design or study reporting. Patients or next of kin had to provide written informed consent before start of any study–related procedure.

Study population
Patients were recruited between October 2012 and February 2014 in a 30–bed mixed medical–surgical ICU of a large university hospital (the Academic Medical Center). Patients were eligible for inclusion if they were aged ≥ 18 years and had an anticipated life expectancy > 96 hours. Patients were excluded from participation if they had a platelet count < 30 x 10⁹/L, had participated in a trial testing an investigational product or treatment within the past 30 days, were pregnant, or had a suspected or diagnosed medical condition, which in the opinion of the investigators prevented the patient from completing the study.

Glucose control
ICU nurses performed glucose control with insulin, following a local guideline for blood glucose control targeting a blood glucose level between 90–144 mg/dL [29]. Insulin titration adjustments were based on sliding scales. The local
guideline for blood glucose control dictated nurses to perform blood glucose measurements at least every four hours, and more frequently if blood glucose levels were out of range or were expected to change rapidly. For details, see the electronic supplement.

During conduct of the study, ICU nurses were not allowed to change insulin infusion rate based on the readings by the investigational device. However, they were allowed to perform additional blood glucose measurements if the device suggested rapid changes in the glucose level, or when there was a trend towards hypoglycemia.

**The investigational device**

The disposable glucose sensors of the device were glucose oxidase–based; two probes, each with two sensing elements. The individual measurement results were combined and displayed on the device monitor every minute. The signal was transmitted through the processor cable to the monitor. It was single-patient single-use sensor, which could be used for up to 72 hours. The processor cable was reusable.

The sensor was inserted into the subcutaneous tissue using two parallel introducer needles. The two needles automatically retracted when the introducer hub was pulled away from the sensor base; the sensor probes remained in the subcutis. Each new sensor needed calibration using blood glucose levels after insertion and initialization, and after 1 hour and 2 hours; thereafter, repeated calibrations were performed every 8 hours.

**Study procedures**

Sensors were inserted into the subcutis of the thigh. Successive sensors could be used for 72 hours, depending on length of stay in the ICU, but never longer than for 30 days. Arterial blood glucose levels were measured using RapidLab 1265 blood gas analyzers (Siemens Healthcare Diagnostics, The Hague, The Netherlands), which were used for calibrations of the device. Not only did ICU nurses provide the mandatory calibration blood glucose levels, but also the routinely obtained blood glucose levels (i.e., blood glucose measurements which were not requested by the device for calibrations, but were taken by the nurses as dictated by the local guideline for blood glucose control) were entered into the device as well. Therefore, these measurements were also used for calibrations of the device. If the device displayed a message requesting an additional non-routine calibration to resolve a sensor performance issue (i.e., a ‘Poor Sensor Signal’ alert), the nurses were permitted to disregard manufacturer recommendation and remove sensors rather than enter the requested calibration.

Each day, the place of insertion was photographed and inspected for redness, bruises, and swelling. In case the patient was awake we questioned the patient
whether it was painful. Every item could be scored as ‘none’, ‘minor’ or ‘major’.

Power calculation
We intended to enroll 50 patients to assess accuracy of the CGM device. With 50 patients we expected to have at least 40,000 subcutaneous CGM device results and at least 1,200 blood glucose level measurements with the RapidLab 1265. Based on previous studies testing point accuracy we assumed to have a sufficiently high number of paired samples to enable evaluation of the point accuracy of the device.

Analysis plan
The glucose data collected with each new sensor were downloaded from the device after use in a patient; the arterial blood glucose levels were downloaded from the patient data management system. The arterial blood glucose levels in the patient data management system were compared with the entries for calibrations into the device. In case of an entry error, defined as a difference between the arterial blood glucose level in the patient data management system and the calibration entry > 9 mg/dL, the correct blood glucose level was used in the accuracy analysis. The subsequent pairs, though, were excluded from the accuracy analysis, since these were influenced by the preceding entry. For reporting point accuracy we used analytical and clinical accuracy measures, i.e., Bland–Altman plot with bias and limits of agreement (bias ± 1.96 x standard deviation of the bias) [30], glucose prediction errors, and Clarke error grid analyses [31]. According to ISO criteria, 95% of the paired measurements should be within the glucose prediction error criteria; general consensus is that 95% of the values should be in zones A and 5% in zones B of Clarke error grid analyses. Finally, we expressed the linearity between the device glucose results and blood glucose results by the Pearson correlation coefficient and coefficient of determination, R².

In a post–hoc analysis we also report point accuracy according to the recently published consensus recommendations [1]. In this round–table meeting of ICU experts in blood glucose control it was recommended to always report the mean absolute relative difference (MARD) when testing a CGM device, where MARD values should be < 14%; values > 18% should be considered to represent poor accuracy [32]. We added the MARD as a post–hoc analysis. Furthermore, we analyzed the point accuracy following the recently published surveillance error grid [33]. For more details see the electronic supplement.

We also reported reasons for early disconnection, defined as the removal of a sensor before 72 hours. For details, see the electronic supplement. The time between calibrations using an incorrect glucose value entry and the next calibration was extracted from the total connection time of the device. Definitions of the metrics used to assess device reliability, including those suggested by
recent consensus recommendations [1] are described in the electronic supplement.

**Statistical analysis**

We reported data as means (± SD) or medians [IQR] where appropriate. In order to be considered for the statistical analysis, each patient needed to have at least 4 comparative blood glucose results for accuracy analysis. However, the excluded patients remained included in the reliability analysis.

In a post–hoc analysis, we used a linear mixed model to determine which variables influence the accuracy of the device. In addition, we stratified the accuracy results by diabetic status. For detailed description of this model, see the electronic supplement.

Analyses were performed using R (version: 2.15.1; R Foundation for Statistical Computing, Vienna, Austria).

**RESULTS**

**Patients and sensors**

We included 50 patients. In total, we used 105 sensors (median of 1 [1–3] sensor per patient) with a total connection time of 4,639 hours (median of 46 [27–134] hours per patient and median of 46 [22–69] hours per sensor). Five patients from whom we could not collect the minimum number of four consecutive comparative blood draws were excluded from the point accuracy analysis. A CONSORT diagram is provided in Figure 1. Patient characteristics and metrics of glucose control are shown in Tables 1 and 2.

We could not inspect the insertion site of 3 sensors. Major bruises were observed in 3 out of 102 inspected insertion sensor sites; 10 minor bruises were seen. Major redness of the skin was observed in 7 out of 102 insertion sensor sites, and minor redness in 6 out of 102 insertion sites. Swelling of the skin was never seen, and none of the conscious patients mentioned pain at the sensor insertion site.

**Point accuracy**

We collected 929 comparative samples (11 [7–28] samples per patient). Bland–Altman plot, glucose prediction error grid, and Clarke error grid are presented in Figure 2. The surveillance error grid is presented in figure 3. The Pearson correlation coefficient was 0.81; the R² was 0.65. The MARD was 14.8%.

Fifty–eight percent of the device results were within 12.5% of the arterial blood glucose results (or within 10 mg/dL for results < 99 mg/dL) and 75% were within 20% of the arterial blood glucose results.

In the linear mixed model, only history of diabetes (P = 0.02) and number of calibrations per sensor (P = 0.04) affected the absolute difference between blood
glucose and device result. Per each new calibration the absolute difference decreased by 1.4% (standard error of 0.006%), meaning that the sensor performance increased. The effect of a history of diabetes was larger, since an increase by 34.3% (standard error of 13.0%) in the absolute difference was found when comparing patients with a history of diabetes compared to patients without diabetes. In addition, we stratified the accuracy results by diabetic status, results are shown in figure S1 and table S2. For detailed results of multivariate random intercept model, see the electronic supplement.

Reliability of the CGM device
Start–up time after placing a new sensor was 49 [43–58] minutes. The number of skips in data acquisition was low, resulting in availability of real-time data during 95 [89–98]% of the connection time per sensor. Table 3 summarizes reliability metrics of the investigational device.
Out of 105 sensors, 60 were removed before 72 hours; reasons and connection times of sensors are shown in table 4. Out of 105 sensors, 42 were removed before 72 hours after insertion for other reasons than ICU discharge or death; and 36 sensors were removed because of an unresolved ‘Poor Sensor Signal’–alert or a device error – 19 with no attempt to resolve.

DISCUSSION
We determined the point accuracy and reliability of a device specifically designed for continuous real–time monitoring of interstitial glucose levels in critically ill patients. The analytic point accuracy of the device was low in a typical cohort of patients from a mixed medical–surgical ICU, according to ISO criteria and consensus recommendations. The clinical point accuracy was low according to Clarke error grid analysis, but better according to surveillance error grid analysis. The device had few down–times, but one–third of the sensors were removed prematurely because of sensor– or device–related problems. The present findings are in line with results from a previous trial testing the same device in cardiac surgery patients [34]. In that study the mean absolute relative difference was 12.2% with 95% real-time data. Similar results come from studies testing other devices for interstitial glucose monitoring that were originally designed for use in non–critically ill patients. Those studies were performed in cardiac surgery patients [21, 24, 35], surgery patients [26], patients with neurologic emergencies [27] and non-surgical patients [16, 22, 25], with only two reporting more favorable accuracy results [21, 22]. Taken together, these data suggest that point accuracy of interstitial glucose monitoring cannot replace blood glucose level measurements.
In contrast to our findings, a previous publication by Brunner et al. suggests a better point accuracy of another interstitial CGM device in critically ill patients [18]. This report combined data of two separate trials in medical ICU patients.
The tested device in that study was from the same manufacture, though not specifically designed for use in critically ill patients. In addition, the sensor was used for up to 72 hours and never replaced. One important difference with the present study was that the sensors were placed exclusively under the skin of the abdomen in patients included in these two trials. In most other trials sensors were inserted under the skin of the abdomen [16, 18, 22, 24, 26, 28], thigh [25, 26] or shoulder [21]. Reported point accuracies do not suggest superiority of one of these sites. Certainly, there could be other unknown and unreported factors that could have resulted in the differences in performance.

We performed a mixed linear model to determine which factors could have influenced the point accuracy of the tested sensor. Rank order of measurement and presence of a history of diabetes affected the accuracy. The finding that rank order of measurement improved sensor performance is not new [18], and certainly not surprising: more calibrations may always increase accuracy of a sensor. A history of diabetes was the most important variable influencing point accuracy, which deteriorated sensor performance with 34%. As yet, this effect remains unexplained. It could be that microcirculation alteration in diabetic patients affects interstitial glucose level. However in previous studies with interstitial devices, diabetes was not found to be significantly associated with poor sensor accuracy in critically ill and cardiac surgery patients [16, 18, 28]. Moreover, in a recent study in cardiac surgery patients, an impaired microcirculation did not affect accuracy of two interstitial glucose sensors from two different manufactures [28]. The difference found between patients with and without diabetes might also be related to glucose variability. Patients with diabetes will have more glucose variability compared to patients without diabetes. Thereby when focusing on percent difference a greater disparity could be found when comparing variability differences.

It should be stressed that we compared interstitial glucose measurements with glucose levels in arterial blood samples, which is far from comparable. Indeed, the interstitial glucose level is dependent on several factors other than the blood glucose level, such as the speed of glucose diffusion from blood to interstitial spaces, as well as the rate of glucose uptake by subcutaneous cells [37]. Importantly, these factors are not constant, particularly in critically ill patients. Furthermore, there is a time lag between interstitial glucose and blood glucose measurement [37]. Studies suggested that the interstitial glucose level decreases before the blood glucose decreases [37, 38], though this was not confirmed in other studies [39]. It is probably very difficult, if not impossible, to correct for factors causing a difference between interstitial and arterial blood glucose levels. Moreover, it is unknown whether differences between arterial and interstitial glucose levels are physiological.

Nevertheless, subcutaneous glucose monitoring could have advantages. One
potential advantage is that continuous monitoring of interstitial glucose levels which enables detection of trends in the blood glucose level [32]. This could allow earlier responses to a rise or a decline of the blood glucose level. In both cases, knowledge of the direction of the trend may be more valuable than the exact blood glucose level.

It is clear that the tested device can never replace blood glucose measurements. First, initial calibrations are always necessary, as are calibrations every 8 hours thereafter. As nurses were allowed to perform additional blood glucose measurements, and as we asked them to insert the values into the investigational device monitor where they were used for additional calibrations, the number of calibrations in this study was higher than mandated. In fact, this could have improved the accuracy of the investigational device: it is possible that with fewer calibrations point accuracy becomes worse.

Our trial knows several strengths and weaknesses. Strengths include the fact that we were able to use the sensors for several days in the participating patients. Moreover, we used accurate blood gas analyzer measurements for comparisons, as well as for the calibrations. Furthermore, we were able to test the device in a typical mixed medical–surgical ICU. Weaknesses include the small sample size and the single–center design of the trial. Furthermore, we did not collect as many samples as we expected. A more important limitation of our trial, though, is the fact that the vast majority of blood glucose levels were in a narrow range, in particular preventing us to draw firm conclusions regarding accuracy in the hypoglycemic range. While the ICU nurses were not allowed to change insulin infusion rates, they could have anticipated hypoglycemia by performing new blood glucose measurements earlier than dictated by the local guideline for blood glucose control, allowing them to respond earlier to e.g., hypoglycemia. Still, some hypoglycemic events occurred, probably because not all nurses were paying attention to the readings of the investigational device. In addition, nurses could have noted that its point accuracy was not always good, so they could have mistrusted the device readings. Finally, we cannot exclude the possibility that hypoglycemia can occur even with the use of CGM. The later will be subject of planned trial. An accuracy analysis limitation was that the assessment focused on percent difference comparisons between the continuous sensor and discrete reference points, evaluated to standards meant for discrete measurements for dosing. Another important limitation is that trend accuracy was not evaluated. Trending is the most interesting endpoint, but mandates very short intervals (i.e., a short as 15 minutes) between blood glucose reference measurements [32, 40]. Trend accuracy should and will be evaluated in future studies.

Notably, length of stay in ICU and sensor connection time was far from similar. This was caused by the fact that sensors could not be used before informed
consent was obtained. Thus we may have missed an important phase of glucose control (i.e., the first day or days of stay in the ICU). In addition, one third of the sensors were removed before sensor life ended, because of sensor – or device related factors. This is an important problem for the reliability of the device. However, nurses did not always attempt to solve sensor – or device related problems that could have been solved. During conduct of the trial they were always allowed to remove the sensor because of ‘Poor Sensor Signal’-alerts or recurrent alarms. With increasing device-specific experience it could be that there are fewer, early removals.

CONCLUSIONS

The point accuracy of a device designed for continuous real-time monitoring of the interstitial glucose level did not meet the ISO15197 standard or the recent consensus guidance for discrete glucose measurement for dosing when used on critically on critically ill patients admitted to a mixed medical–surgical ICU. While this device is not a replacement for current blood gas analyzer–measurements, a real–time system may be used for trend guidance on timely reference measurement for insulin adjustment. The device had few down times, but one third of the sensors were removed prematurely because of unresolved sensor – or device–related problems.

References


FIGURES

Patients screened
N = 790

Excluded:
- discharge < 24 hours N = 531
- No Informed Consent N = 61
- life expectancy < 96 hours N = 42
- missed N = 40
- platelet count < 30 N = 21
- doctor exclusion N = 16
- not willing to participate N = 15
- in other trial N = 5
- pregnant N = 3
- no monitor available N = 2
- age < 18 years N = 2
- readmission already included N = 20

Included patients
N = 50

Comparative samples < 4 N = 5

Included in the point accuracy analysis
N = 45

Figure 1.
CONSORT diagram of the study.

Figure 2.
Bland–Altman plot with bias and limits of agreement (bias ± 1.96 x standard deviation of the bias)[30], glucose prediction errors, and Clarke error grid analyses[31].
Figure 3.
Surveillance error grid with risk scores.
### TABLES

<table>
<thead>
<tr>
<th></th>
<th>N=50 All included patients</th>
<th>N=45 Patients included in the point accuracy analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age – years, median [IQR]</td>
<td>65 [56–72]</td>
<td>65 [55–72]</td>
</tr>
<tr>
<td>Male gender, number (%)</td>
<td>25 (50%)</td>
<td>24 (53%)</td>
</tr>
<tr>
<td>Race, number (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>–Caucasian</td>
<td>45 (90%)</td>
<td>40 (89%)</td>
</tr>
<tr>
<td>–Black</td>
<td>4 (8%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>–Asian</td>
<td>1 (2%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>BMI – kg/m², median [IQR]</td>
<td>24.7 [22.4–27.6]</td>
<td>24.4 [22.2–27.3]</td>
</tr>
<tr>
<td>Admission diagnosis, number (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>–Medical</td>
<td>31 (62%)</td>
<td>26 (58%)</td>
</tr>
<tr>
<td>–Emergency surgery</td>
<td>11 (22%)</td>
<td>11 (24%)</td>
</tr>
<tr>
<td>–Planned surgery</td>
<td>8 (16%)</td>
<td>8 (18%)</td>
</tr>
<tr>
<td>Planned admission, number (%)</td>
<td>10 (20%)</td>
<td>9 (20%)</td>
</tr>
<tr>
<td>History of Diabetes, number (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>–No diabetes</td>
<td>39 (78%)</td>
<td>34 (76%)</td>
</tr>
<tr>
<td>–Diabetes, unknown treatment</td>
<td>2 (4%)</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>–Diabetes treated with insulin</td>
<td>4 (8%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>–Diabetes treated with oral agents</td>
<td>5 (10%)</td>
<td>5 (12%)</td>
</tr>
<tr>
<td>ICU mortality, number (%)</td>
<td>11 (22%)</td>
<td>10 (22%)</td>
</tr>
<tr>
<td>Hospital mortality, number (%)</td>
<td>15 (30%)</td>
<td>14 (31%)</td>
</tr>
</tbody>
</table>

**Table 1.**

Patient Characteristics. Abbreviations: APACHE, Acute Physiology and chronic Health Evaluation; BMI, body–mass index; IQR, interquartile range; ICU, intensive care unit; LOS, length of stay; SAPS, Simplified Acute Physiology Score;
<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of measurements</td>
<td>929</td>
</tr>
<tr>
<td>Mean blood glucose level per patient – mg/dL (median [IQR])</td>
<td>132 [125–148]</td>
</tr>
<tr>
<td>Standard deviation of blood glucose level per patient – mg/dL (median [IQR])</td>
<td>24 [16–33]</td>
</tr>
<tr>
<td>Number of measurements per patient (median, [IQR])</td>
<td>11 [7–29]</td>
</tr>
<tr>
<td>Severe Hypoglycemia ≤ 40 mg/dL – measurements, number (%)</td>
<td>3 (0.3%)</td>
</tr>
<tr>
<td>Severe Hypoglycemia ≤ 40 mg/dL – patients, number (%)</td>
<td>2 (4.4%)</td>
</tr>
<tr>
<td>Mild hypoglycemia 41 – 70 mg/dL – measurements, number (%)</td>
<td>15 (1.6%)</td>
</tr>
<tr>
<td>Mild hypoglycemia 41 – 70 mg/dL – patients, number (%)</td>
<td>7 (15.6%)</td>
</tr>
<tr>
<td>Mild hyperglycemia 150 – 179 mg/dL – measurements, number (%)</td>
<td>163 (17.5%)</td>
</tr>
<tr>
<td>Mild hyperglycemia 150 – 179 mg/dL – patients, number (%)</td>
<td>35 (77.8%)</td>
</tr>
<tr>
<td>Severe hyperglycemia &gt; 180 mg/dL – measurements, number (%)</td>
<td>111 (11.9%)</td>
</tr>
<tr>
<td>Severe hyperglycemia &gt; 180 mg/dL – patients, number (%)</td>
<td>19 (42.2%)</td>
</tr>
</tbody>
</table>

**Table 2.**
Measurements. Data considers all paired measurements and result of blood gas analyzer is shown; Abbreviations: IQR, interquartile range.
<table>
<thead>
<tr>
<th></th>
<th>Per patient</th>
<th>Per sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sensors used</td>
<td></td>
<td>105</td>
</tr>
<tr>
<td>Number of sensors used, median [IQR]</td>
<td>1 [1–3]</td>
<td>–</td>
</tr>
<tr>
<td>Total connection time in hours, median [IQR]</td>
<td>46.2 [26.8–134.2]</td>
<td>45.8 [21.1–69.1]</td>
</tr>
<tr>
<td>Initialization time in minutes, median [IQR]</td>
<td>Median 34 [34–34.5]</td>
<td>34 [34–35]</td>
</tr>
<tr>
<td>Real-time data in hours</td>
<td>42.3 [23.1–130.3]</td>
<td>41.4 [20.6–64.0]</td>
</tr>
<tr>
<td>Percentage of real-time data, median [IQR]</td>
<td>94.1 [88.9–97.1]</td>
<td>94.6 [88.7–97.9]</td>
</tr>
<tr>
<td>Time of skips in data acquisition in hours, median [IQR]</td>
<td>4.3 [1.2–9.1]</td>
<td>2.6 [0.6–5.4]</td>
</tr>
<tr>
<td>Time of skips in data acquisition in hours caused by poor sensor signal, median [IQR]</td>
<td>0 [0–1.0]</td>
<td>0 [0–0.2]</td>
</tr>
<tr>
<td>Time of skips in data acquisition in minutes caused by other reasons, median [IQR]</td>
<td>3.3 [0.9–8.4]</td>
<td>2.0 [0.4–3.7]</td>
</tr>
<tr>
<td>Percentage of time of skips in data acquisition, median [IQR]</td>
<td>5.9 [2.9–11.1]</td>
<td>5.4 [2.1–11.3]</td>
</tr>
<tr>
<td>Percentage of time of skips in data acquisition caused by poor sensor signal, median [IQR]</td>
<td>0 [0–0.7]</td>
<td>0 [0–0.3]</td>
</tr>
<tr>
<td>Percentage of time of skips in data acquisition caused by other reasons, median [IQR]</td>
<td>4.2 [2.3–8.0]</td>
<td>3.8 [1.5–8.0]</td>
</tr>
<tr>
<td>Number of calibrations, median [IQR]</td>
<td>14 [9–34]</td>
<td>12 [7–16]</td>
</tr>
<tr>
<td>Number of mandated calibrations, median [IQR]</td>
<td>8 [4–20]</td>
<td>6 [4–8]</td>
</tr>
</tbody>
</table>

**Table 3.**
Device reliability. Abbreviations: IQR, interquartile range
<table>
<thead>
<tr>
<th>Total number of sensors used</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sensors removed &lt; 72 hours</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of sensors</th>
<th>Percentage of sensors removed &lt;72 hours</th>
<th>Percentage of total number of sensors used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient–related factors:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Discharge &lt; 72 hours after insertion</td>
<td>14</td>
<td>(23%)</td>
</tr>
<tr>
<td>– Death, 72 hours after insertion</td>
<td>4</td>
<td>(7%)</td>
</tr>
<tr>
<td>Sensor– or device–related factors:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Accidental removal of sensor</td>
<td>6</td>
<td>(10%)</td>
</tr>
<tr>
<td>– Poor sensor signal (19 had no attempt to resolve)</td>
<td>34</td>
<td>(57%)</td>
</tr>
<tr>
<td>– Device error</td>
<td>2</td>
<td>(3%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration of sensors in place [hours], median [IQR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>– All sensors</td>
</tr>
<tr>
<td>– Sensor that were removed &lt; 72 hours</td>
</tr>
<tr>
<td>– Sensors that were removed &lt; 72 hours because of patient–related factors</td>
</tr>
<tr>
<td>– Discharge &lt; 72 hours after insertion</td>
</tr>
<tr>
<td>– Death, 72 hours after insertion</td>
</tr>
<tr>
<td>– Sensors that were removed &lt; 72 hours because of patient–related factors</td>
</tr>
<tr>
<td>– Accidental removal of sensor</td>
</tr>
<tr>
<td>– Poor sensor signal</td>
</tr>
<tr>
<td>– Device error</td>
</tr>
</tbody>
</table>

**Table 4.**
Sensors removed < 72 hours
SUPPLEMENTARY INFORMATION

The local guideline for glucose control
ICU nurses performed glucose control with insulin, following a local guideline for blood glucose control targeting a blood glucose level between 90–144 mg/dL [29].
According to this local guideline, insulin infusion was started when the blood glucose level was > 144 mg/dL. Insulin titration adjustments were made based on sliding scales. The guideline advised to stop insulin infusion and to give boluses of dextrose only when the blood glucose level declined to < 61 mg/dL. Insulin infusion was exclusively given intravenously and continuously; boluses of insulin were only allowed when the blood glucose level was > 360 mg/dL; subcutaneous insulin boluses were never allowed.
The local guideline also dictated to perform blood glucose measurements at least every four hours, but more frequently if blood glucose levels were out of range or rapidly changing. Typically, blood glucose levels were measured more frequently at the start of insulin titration, and in cases of an increased risk of hypoglycemia. Blood glucose levels used for insulin adjustment were measured in arterial blood samples using RapidLab 1265 blood gas analyzers (Siemens Healthcare Diagnostics, The Hague, The Netherlands), located in the ICU. The results were automatically downloaded to the patient data management system (MetaVision®, iMDsoft, Tel Aviv, Israel), which was present at every ICU bed.

Methods to calculate point accuracy
For reporting point accuracy we used glucose prediction errors, defined as [blood glucose – device glucose result]. The percentage of data points that fell within ± 15 mg/dL of the blood glucose results for blood glucose results < 75 mg/dL and within 20% of the blood glucose results for blood glucose results ≥ 75 mg/dL were reported according to the current International Standards Organization standard (ISO15197) [32]. We also used Clarke error grid analyses to show the percentage of paired data values falling within each zone of the Clarke error grid [30], and Bland–Altman plot [31]. The Clarke error grid is divided in 5 paired ‘zones’: zones A (measurement within 20% of the reference or glucose levels < 70 mg/dL); zones B (measurement more than 20% different from the reference but still clinically acceptable as they would not change the rate of insulin infusion); zones C (measurement that would lead to unnecessary changes in insulin infusion, i.e., overcorrecting acceptable glucose levels); zones D (potentially dangerous hypo- or hyperglycemic events are missed); and zones E (levels that would lead to a decision opposite to that required, i.e., treatment for hypoglycemia instead of hyperglycemia). General consensus is that 95% of the values should be in zones A and 5% in zones B.
The Bland–Altman plot is presented with bias (mean difference between the device glucose results and blood glucose results) and limits of agreement (bias ± 1.96 x standard deviation of the bias) to analyze the agreement between the device glucose results and blood glucose results.

In a post–hoc analysis we also determined point accuracy according to the recently published consensus recommendations [1]. For this, the percentage of data points that fell within 12.5% of the blood glucose results, or within 10 mg/dL for readings < 99 mg/dL were reported. In a round the table meeting of ICU experts it was recommended to report the mean absolute relative difference and values should be <14%; values >18% were considered to represent poor accuracy [32]. Furthermore, we analyzed the accuracy following the recently published the surveillance error grid [33].

**Definitions of metrics for device reliability**

The following metrics and definitions were used to assess device reliability, including those suggested by recent consensus recommendations [1]

- **Connection time** – time between first device glucose results and last glucose result
- **Start–up time** – time between the start of initialization of sensor and first device glucose result after calibration, including blood glucose measurement time and time for nurse to enter value into the device.
- **Initialization time** – time between initialization of sensor and ready for calibration
- **Real-time data** – time when device glucose results were available
- **Percentage of real-time data** – percentage of time device glucose results were available divided by total connection time
- **Skips in data acquisition all causes** – total time when the monitor gave no results
- **Percentage of skips in data acquisition all causes** – percentage of time when the monitor gave no results divided by the connection time
- **Skips in data acquisition poor sensor signal** – percentage of time when the monitor gave no results caused by poor sensor signal
- **Percentage of skips in data acquisition poor sensor signal** – percentage of time of skips in data acquisition caused by poor sensor signal divided by the connection time minus the time of skips in data acquisition caused by other reasons
- **Skips in data acquisition other reasons** – time of skips in data acquisition caused by other reasons than poor sensor signal
- **Percentage of skips in data acquisition other reasons** – percentage of time of skips in data acquisition caused by other reasons divided by the connection time minus time of skips in data acquisition caused by poor sensor signal
sensor signal

- ‘Poor Sensor Signal’ – a device alert indicating that the sensor may be experiencing decreased performance. This alert removes the real time sensor glucose value display until a requested reference calibration value is entered to recover sensor performance.

FACTORS THAT AFFECT POINT ACCURACY

Background
The aim of the primary study was to test the point accuracy and reliability of an interstitial CGM device in a mixed medical–surgical ICU. We found a low point accuracy of an interstitial CGM device in a mixed medical–surgical ICU. We were interested if this was dependent on particular variables. Therefore, we performed a post-hoc analysis to determine which variables influence the accuracy of the device.

Methods
We used a linear mixed model to determine which variables influence the accuracy of the device. For this, patient and sensor were used as random intercepts to account for repeated measurements. The absolute difference between the arterial blood glucose level and device glucose level was the dependent variable. The absolute difference was logarithmically transformed (using the natural logarithm) to obtain a normal distribution. The following variables were chosen based on clinical relevance and previous trials testing other CGM devices [18, 19, 28]: demographic variables including gender, age, body mass index and history of diabetes; disease severity variables including the APACHE II score and the circulation score of the Sequential Organ Failure Assessment (SOFA) Score on the day of measurements; in addition, we added time between calibrations (as shorter time between calibrations could improve accuracy) and the rank order of the paired glucose results (as more calibrations could improve accuracy) [18]. All variables were added to the model without considering further model reduction strategies. Visual inspection of residuals was done. Correlation between covariates was assessed to investigate collinearity. The effect of covariates on the absolute difference was reported as the percentage of change in the absolute difference with the standard error.

Results (table S1)
We performed a linear mixed-effects model with a fit by maximum likelihood. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Pearson correlation coefficients were all under 0.5 showing no collinearity.
In the linear mixed model only history of diabetes (P = 0.02) and number of calibrations per sensor (P = 0.04) affected the absolute difference between blood glucose and device result. Per each new calibration the absolute difference decreased with 1.4% (standard error of 0.006%), meaning that the sensor performance increased. The effect of a history of diabetes was bigger, though, since diabetes increased the absolute difference with 34.3% (standard error of 13.0%). Therefore we stratified our accuracy metrics by diabetic status (see figure S1 and table S2).

The formula for the final mixed model was:
\[
\text{Log(Absolute difference)} = 2.419 + \text{random intercept per patient} + \text{random intercept per sensor} + 0.295 \times \text{Diabetes} - 0.014 \times \text{rank order of measurement} + 0.011 \times \text{Sofa Circulation Score} + 0.025 \times \text{BMI} - 0.073 \times \text{Gender} - 0.003 \times \text{Age} - 0.009 \times \text{APACHE II} + 0.0001 \times \text{time between calibration}
\]

RELIABILITY ANALYSIS

Background
In the present study we found that more than half of the sensors had to be removed before 72 hours. We wanted to know reasons for disconnection and when this happened. Therefore we did a post–hoc analysis to investigate reasons for early disconnection.

Methods
Early disconnection was defined as the removal of a sensor before 72 hours, which could be caused by:

- Poor sensor signal – sensor performance issue, in which the systems requests additional calibrations to solve. Nurses were able to remove the sensor when the monitor gave a poor sensor signal alarm without attempt to solve.
- Accidental removal of the sensor
- Device error – the device monitor had an technical failure

Furthermore the connection time was calculated (time between first device glucose results and last glucose result) for sensors, which were removed before 72 hours. The time between calibrations using an incorrect glucose value entry and the next calibration was extracted from the total connection time of the device.
Figure S1.
Bland–Altman plot with bias and limits of agreement (bias ± 1.96 standard deviation of the bias), glucose prediction errors, and Clarke error grid analyses stratified by diabetic status.
<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>0.036</td>
<td>0.189</td>
</tr>
<tr>
<td>Sensor ID</td>
<td>0.047</td>
<td>0.217</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Value</td>
<td>Standard error</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.419</td>
<td>0.489</td>
</tr>
<tr>
<td>History of Diabetes</td>
<td>0.295</td>
<td>2.408</td>
</tr>
<tr>
<td>Rank order</td>
<td>-0.014</td>
<td>0.007</td>
</tr>
<tr>
<td>Sofa circulation score</td>
<td>0.011</td>
<td>0.026</td>
</tr>
<tr>
<td>BMI</td>
<td>0.025</td>
<td>0.015</td>
</tr>
<tr>
<td>Male gender</td>
<td>-0.073</td>
<td>0.109</td>
</tr>
<tr>
<td>Age in years</td>
<td>-0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>APACHE II score</td>
<td>-0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>Time between calibration in minutes</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Table S1.**

<table>
<thead>
<tr>
<th></th>
<th>Diabetic</th>
<th>Non-diabetic</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of paired samples</td>
<td>337</td>
<td>592</td>
<td>929</td>
</tr>
<tr>
<td>Mean absolute relative difference</td>
<td>16.0</td>
<td>14.2</td>
<td>14.8</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.84</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>R2</td>
<td>0.70</td>
<td>0.50</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Consensus recommendations

- percentage of measurements within 12.5% blood glucose results (or within 10 mg/dL for results < 99 mg/dL)

- percentage of measurements within 20% blood glucose results

<table>
<thead>
<tr>
<th></th>
<th>Diabetic</th>
<th>Non-diabetic</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>55</td>
<td>59</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>77</td>
<td>75</td>
</tr>
</tbody>
</table>

**Table S2.**
Chapter 4

Point and Trend Accuracy of a Continuous Intravenous Microdialysis–based Glucose–monitoring Device in Critically Ill Patients – a Prospective Study


Annals of intensive care 2016, 6
ABSTRACT

Introduction: Microdialysis is a well-established technology that can be used for continuous blood glucose monitoring. We determined point and trend accuracy, and reliability of a microdialysis–based continuous blood glucose-monitoring (CGM) device (EIRUS®) in critically ill patients.

Methods: Prospective study involving patients with an expected intensive care unit (ICU) stay of ≥ 48 hours. Every 15 minutes, device readings were compared with blood glucose values measured in arterial blood during blocks of 8 hours per day for a maximum of three days. The Clarke Error grid, Bland–Altman plot, mean absolute relative difference and glucose prediction error analysis were used to express point accuracy, and the Rate Error Grid to express trend accuracy. Reliability testing included aspects of the device and the external sensor, and the special central venous catheter (CVC) with a semipermeable membrane for use with this device.

Results: We collected 594 paired values in 12 patients (65 [26–80; 8 – 97] median [IQR; total range]) paired values per patient). Point accuracy: 93.6% of paired values were in zone A of the Clarke error grid, 6.4% were in zones B; bias was 4.1 mg/dL with an upper limit of agreement of 28.6 mg/dL and a lower level of agreement of –20.5 mg/dL in the Bland Altman analysis; 93.6% of the values ≥ 75 mg/dL were within 20% of the reference values in the glucose prediction error analysis; the mean absolute relative difference was 7.5%. Trend accuracy: 96.4% of the paired values were in zone A, and 3.3% and 0.3% were in zones B and zones C of the rate error grid. Reliability: out of 16 sensors, 4 had to be replaced prematurely; out of 12 CVCs, two malfunctioned (one after unintentional flushing by unsupervised nurses of the ports connected to the internal microdialysis chamber, causing rupture of the semipermeable membrane; one for an unknown reason). Device start–up time was 58 [56–67] minutes; availability of real–time data was 100% of the connection time.

Conclusions: In this study in critically ill patients who had no hypoglycemic episodes and a limited number of hyperglycemic excursions, point accuracy of the device was moderate to good. Trend accuracy was very good. The device had no downtimes, but 4 out of 16 external sensors and 2 out of 12 CVCs had practical problems.

INTRODUCTION

Most, if not all critically ill patients receive intravenous infusion of insulin for blood glucose control at some point during stay in the intensive care unit (ICU) [1]. This strategy requires frequent blood glucose measurements for the guidance of insulin titrations, but this is both time and blood–consuming [2]. Automation of blood sampling and glucose measurement through continuous glucose monitor (CGM) devices could reduce this burden, and has the potential
to improve overall blood glucose control [3,4]. Microdialysis offers the opportunity to sample blood analytes with high accuracy but without the need for drawing blood samples. EIRUS® (Maquet Critical Care AB, Solna, Sweden), a microdialysis–based device that can measure blood glucose and lactate levels, has been tested and validated previously in studies in surgical patients, where it has been found to be safe and accurate [4-7]. To date, its accuracy with regard to blood glucose monitoring, and reliability have not yet been tested extensively in ICU patients [4,6]. We hypothesized the EIRUS® system to be point and trend accurate, and to be reliable in ICU patients. To test this hypothesis, we used this CGM device in a series of critically ill patients, to compare device readings with frequently measured arterial blood glucose values. Along the study, we determined reliability of the device and the special central venous catheter (CVC) with a semi-permeable membrane designed for use with this device.

METHODS
Study design and population
This investigator–initiated prospective study was conducted in the mixed medical–surgical ICU of the Academic Medical Center, Amsterdam, The Netherlands. The Institutional Review Board of the Academic Medical Center approved the study protocol, and informed consent was obtained from all patients or their legal representatives before start of the study. Maquet Critical Care AB provided the CGM device and its disposables free of charge. Maquet Critical Care AB had neither influence on the design of this study, nor on reporting of the results. The study was registered at the Netherlands Trial Register (NTR4527). Patients were eligible for participation if they were at least 18 years old, were expected to stay in the ICU for ≥ 48 hours, had an arterial catheter in place and were in need of a (new) CVC. Patients were excluded if they participated in another investigational drug or device study, or were known to be pregnant.

Blood glucose control
ICU nurses followed a local guideline aiming at a blood glucose level between 90–144 mg/dL (5–8 mmol/L) as part of standard care. This guideline mandated nurses to measure blood glucose every four hours, or more frequently when glucose levels were out of range or when rapid changes were expected. Infusion of insulin was started when glucose levels were over 144 mg/dL and stopped when glucose was lower than 61 mg/dL. Adjustments of insulin titration were based on sliding scales. More details can be found in the online supplement. In addition, details on how nurses were trained can also be found in the online supplement. During the study, ICU nurses were not allowed to change insulin infusion rate
based on the readings by the device. The ICU nurses, however, had access to device readings and additional arterial blood glucose measurements were allowed if the device suggested rapid changes in the glucose level, or when there was a trend towards hypoglycemia. In addition, the ICU nurses could also adjust insulin infusion rates based on reference blood glucose values obtained during study observation periods (see below).

**The study device**
For intravenous microdialysis–based glucose monitoring a special CVC with a semi–permeable membrane (Maquet Critical Care AB, Solna, Sweden) is needed. This CVC has five lumens, three ‘normal’ ports for intravenous administration of fluids or medication, and two ‘special’ ports for transport of normal saline alongside the semi–permeable membrane, which should not be flushed and cannot be used for intravenous administration of fluids or medication. The ‘afferent’ port is connected to a saline–filled syringe placed in the syringe pump of the device. The ‘efferent’ port is connected to the disposable sensor. Small metabolites such as glucose pass through the semi–permeable membrane creating equilibrium between blood and the dialysate. The dialysate is pumped over the sensor in a continuous fashion, where the glucose oxidase method is used to measure the glucose level [4,5]. The device can be used for a maximum of 96 hours per sensor. Reference measurements are needed for calibration of the device, which is performed at start–up and every 8 hours thereafter.

Of note, because the dialysate needs to be transported to the sensor outside the patient, where measurements are performed, there is a delay in time of 5 minutes between dialysate formation and the actual measurements.

**Study procedures**
In three blocks of eight hours per day, and for a maximum of three days, every 15 minutes an arterial blood sample of 200 µl was drawn through an existing arterial line. Blood glucose levels were measured using a blood gas analyzer (RAPIDLab 1265, Siemens Healthcare Diagnostics, The Hague, The Netherlands). Definitions of the metrics used to assess device reliability, including those suggested by recent consensus recommendations [8], are described in detail in the online supplement, and included the percentage of real–time data, skips in data acquisition, failures to calibrate, sensor failures and CVC failures.

**Power calculation**
Based on previous studies [9,10] we chose to collect approximately 1,000 paired measurements or to connect the device to a minimum number of 11 patients. Inclusion of patients was restricted by the time the device was available for
this study, and the number of disposable CVCs and sensors provided by the manufacturer.

**Analysis plan**

Patient characteristics were reported as means, medians or percentages, where appropriate. Because of the delay between dialysate formation and the actual measurements of the blood glucose level, we subtracted 5 minutes from the timestamp of the values of the CGM device; as such reference blood glucose values matched with the moment dialysate was formed. Subsequently, device and reference measurements were merged. Paired measurements were used for determining point and trend accuracy of the device. To be considered for the statistical analysis, each patient needed to have at least multiple samples with at most 30 minutes in between. However, patients excluded for statistical analysis remained included in the reliability analysis. While each paired sample was included in the point accuracy analysis, only the samples with a gap of at most 30 minutes to the next sample were included in the trend accuracy analysis. In addition, when the device was calibrated within the daily 8–hour block of intense sampling, the calibration sample and the subsequent sample were not considered for trend accuracy analysis. This way, large changes in trend due to the calibration were excluded from the analysis.

Point accuracy was expressed using a Clarke error grid, a Bland–Altman plot, the glucose prediction error analysis, and the mean absolute relative difference (MARD). To be considered point accurate, at least 95% of values must be in zones A, a maximum of 5% can be in zones B, and no values are allowed in zones C to E of the Clarke error grid [11]. Also, the MARD should be below 14%; a value above 18% represent poor accuracy [3].

Trend accuracy was expressed using rate error grid Analysis (R–EGA) [12]. Values outside zones A and B of the R–EGA corresponding to values in zones A and B of the Clarke error grid were considered benign errors. On the other hand, values outside zones A and B of the R–EGA corresponding to values outside zones A and B of the Clarke error grid were considered erroneous readings [12].

**Post–hoc analysis**

Point accuracy was also expressed using the recently published surveillance error grid [13]. Two of the CVCs were malfunctioning. In one case it was immediately clear that the CVC was defect, and no additional measurements were performed. In another case this was not immediately clear, and only after reviewing the readings it became clear that the CVC started to malfunction from a certain timepoint. We chose to perform a post–hoc analysis excluding the data from that patient.
RESULTS

A total of 12 patients were included in whom 598 paired measurements were available. Figure 1 shows the CONSORT diagram. One patient was excluded from the point and trend accuracy analyses because no comparative samples could be obtained while the device was connected due to calibration problems. In one patient, four arterial blood samples had to be discarded as they were diluted during sampling. Thus we had 594 samples (65 [26 – 80; 8 – 97] (median [IQR; total range]) paired values per patient) for determining point accuracy of the CGM device. For trend accuracy analysis, 482 samples were used. Patient characteristics are shown in Table 1. Metrics of glucose control are shown in Table 2.

Point and trend accuracy

The Clarke error grid, Bland–Altman plot, and glucose prediction error grid are presented in Figure 2. Bias in the Bland–Altman plot was 4.1 mg/dL with an upper limit of agreement of 28.6 mg/dL and a lower limit of agreement of -20.5 mg/dL. Glucose prediction error analysis showed that 93.6% of the values ≥ 75 mg/dL within twenty percent of the values measured by the blood gas analyzer were within range. The MARD was 7.5%. The rate error grid is presented in Figure 3, consisting of 99.7% accurate readings and 0.3% benign errors.

Reliability

Table 3 shows reliability results. Start–up time was 58 [56 – 67; 48 – 112.8] (median [IQR; total range]) minutes. In three patients, the initial sensor could not be calibrated at start–up, and a second sensor was needed. In two patients, the CVC malfunctioned after some hours. In one patient, this was due to improper handling by one of the trained ICU nurses. This and other details on reliability are discussed in more detail in the online supplement.

Post–hoc analysis.

The Surveillance error grid is presented in Figure S1 in the supplement. Results of the post–hoc analysis excluding the data from the patient mentioned above in whom the CVC was malfunctioning for unknown reasons is presented in online supplement Figures S3, S4 and S5.

DISCUSSION

In this study in a cohort of critically ill patients, point accuracy of a microdialysis–based CGM device developed was moderate to good. Trend accuracy was very good. Reliability was moderate, seen as 4 out of 16 external sensors could not be used and 2 out of 12 CVCs had practical problems. Point accuracy in the present study was less than the point accuracy reported
from two previous studies in patients after cardiac surgery [4,6]. In these study, all paired values were in zones A and B, with 97% and 99% of values in zones A of the Clark error grid, and the MARD was only 5.6%. and 5%, respectively. Both those studies, and the present study used arterial blood gas analyzers as a reference standard. The present study, however, was conducted in patients that were more severely ill than cardiac surgery patients, reflected by a longer length of stay in the ICU stay (15 vs. 3 days) and hospital (20 vs. 8 days). Thus, these two studies included completely different patients, which could, at least in part explain the differences. The results of the present study, however, are very similar to a pilot study in abdominal surgery patients [5], in which all paired values were in zone A and B, with 94% of values in zones A of the Clark error grid.

According to a recent consensus on blood glucose monitoring, 95% of paired values need to be in zones A of the Clarke error grid to qualify a device as point accurate [11]. In contrast, a more recent consensus amongst a panel of ICU experts, the MARD should be < 14% [3]. While the studied device did not meet the first criteria, it did meet the last. There are no generally accepted criteria for trend accuracy of CGM devices in the ICU setting [3]. Nonetheless, we believe EIRUS to be very accurate, as only one value was in the benign error range[12]. In addition, it should be noted that the paired measurement in zone C mentioned above came from the patient in whom the special CVC was malfunctioning. Since both glucose and lactate measured by the device decreased rapidly and non–physiologically, we suspect that the semipermeable membrane of that CVC broke.

Both the afferent and efferent ports of the CVC, connected to the dialysate chamber, were labeled with tags mentioning not to flush these ports. Unfortunately, unsupervised nurses thought there was backflow of blood in the afferent port because of the deep–red/purple color, and flushed it with normal saline immediately. This resulted in a rupture of the delicate semi–permeable membrane, and thus malfunctioning of the CVC: saline pumped into the chamber disappeared into the circulation, and the efferent port stopped producing dialysate. After this we continued using the special CVC as a normal CVC, with two stops at the extra ports. The manufacturer changed the color of the lumen and its connector to prevent this incident after this study. These problems however, did raise some concerns. However, we do not believe that these problems were caused by an insufficient introduction of the study device in the unit since we organized multiple training sessions for nurses and instructed nurses individually when a patient was included in the study and monitored by the device.

Our study has several limitations. First and most importantly, no hypoglycemic periods were captured during the study, and the number of hyperglycemic
events was small. While the device proved point accurate in the hypoglycemic range in one study in animals [14], we remain uncertain on hypoglycemic performance in ICU patients. The absence of hypoglycemia might be explained by the fact that reference measurements were performed very frequently, and because nurses had access to the device readings. Nurses were allowed to use the reference measurements, and thus could improve blood glucose control (i.e., prevent hypoglycemic events). Even the device readings could have helped nurses to prevent dangerous excursions of the blood glucose level, even though they knew that this was an investigational device. The local Institutional Review Board did not accept blinding the nurses for the reference measurements and the device readings. In addition, the fact that we only actively collected paired measurements during day–time hours means that we might have missed possible interesting data overnight. More paired samples, also outside working hours, could have yielded more hypoglycemic events. To make a more conclusive statement on device accuracy in the hypoglycemic range, other methods for capturing hypoglycemic and hyperglycemic events have to be further explored. One recently suggested way to improve the execution of accuracy testing of investigational devices in the clinical setting includes data mining of electronic medical records [15,16]. Data mining is a technique that uses large quantities of data in search for certain events, in this case hypoglycemia and hyperglycemia. Comparison between consecutive measurements of the blood glucose level by means of a CGM device and comparative measurements in a central laboratory then could be used to determine the accuracy in these extreme situations. This approach certainly increases the number of hypoglycemic and hyperglycemic events that can be used for accuracy testing, but of course requires extensive use in one of more intensive care units. Finally, as of September of 2015, shortly after analyzing the data before reaching our goal of 1,000 paired samples, we had to stop the study prematurely. The data allowed for a sufficiently narrow interval of confidence on the point and trend accuracy of the machine and therefore we did not consider it ethically justified to include more patients, seen the potential burden and risks of obtaining blood samples every 15 minutes.

There were also several strengths to this study. This is the first study to date to investigate trend accuracy of a CGM device in critically ill patients. In addition, the investigated microdialysis CGM device had not been tested in a mixed ICU before. This makes the results of this study more clinically applicable as this is indeed the patient population in which glucose monitoring is most relevant. Finally, we used precise blood gas analyzers for reference measurements, and we corrected for the 5–minute delay between formation of the dialysate and the measurement at the sensor side.
CONCLUSION

The point and trend accuracy of the tested microdialysis–based CGM device was moderate to good in patients who were stable with regard to their blood glucose levels. Trend accuracy was very good. The device had no downtimes, but 4 out of 16 external sensors and 2 out of 12 CVCs had practical problems.

References

FIGURES

Patients screened
N = 2794

Excluded:
- Expected stay < 48 h  N = 1141
- Missed or wrong CVC  N = 688
- No CVC  N = 618
- No arterial line  N = 317
- Age < 18  N = 20
- In other trial  N = 4
- No informed consent  N = 3
- Pregnant  N = 1

Included
N = 12

No paired measurements  N = 1

Included in accuracy analysis
N = 11

Figure 1.
CONSORT diagram.
Figure 2.
Measures of point accuracy. Bland–Altman plot (upper left panel), glucose prediction error grid (lower left panel) and Clarke error grid (right panel).

Continuous Glucose - Error Grid Analysis

Figure 3.
Rate Error–Grid of the Continuous glucose–error grid analysis. This grid is divided into similar zones as the Clarke error grid. Perfectly trend accurate values are the dashed line in the middle.
### TABLES

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, median [IQR]</td>
<td>65 [60 – 79]</td>
</tr>
<tr>
<td>Male gender, number (%)</td>
<td>6 / 50%</td>
</tr>
<tr>
<td>Race, number (%)</td>
<td></td>
</tr>
<tr>
<td>- Caucasian</td>
<td>10 (83.3 %)</td>
</tr>
<tr>
<td>- Black</td>
<td>1 (8.3%)</td>
</tr>
<tr>
<td>- Asian</td>
<td>1 (8.3%)</td>
</tr>
<tr>
<td>BMI in kg/m², median [IQR]</td>
<td>23 [20 – 27]</td>
</tr>
<tr>
<td>Admission type, number (%)</td>
<td></td>
</tr>
<tr>
<td>- Medical</td>
<td>6 (50%)</td>
</tr>
<tr>
<td>- Emergency Surgery</td>
<td>3 (25%)</td>
</tr>
<tr>
<td>- Planned Surgery</td>
<td>3 (25%)</td>
</tr>
<tr>
<td>History of diabetes, number (%)</td>
<td></td>
</tr>
<tr>
<td>- No diabetes</td>
<td>11 (92%)</td>
</tr>
<tr>
<td>- Diabetes, unknown treatment</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>- Diabetes treated with insulin</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>- Diabetes treated with oral agents</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>APACHE II, median [IQR]</td>
<td>21 [18 – 26]</td>
</tr>
<tr>
<td>SAPS II, median [IQR]</td>
<td>44 [37 – 53]</td>
</tr>
<tr>
<td>ICU LOS, days, median [IQR]</td>
<td>15 [7 – 17]</td>
</tr>
<tr>
<td>Hospital LOS, days, median [IQR]</td>
<td>20 [18 – 35]</td>
</tr>
<tr>
<td>ICU mortality, number (%)</td>
<td>7 (58%)</td>
</tr>
<tr>
<td>Hospital mortality, number (%)</td>
<td>8 (67%)</td>
</tr>
</tbody>
</table>

**Table 1. Patient characteristics**

IQR, Interquartile range; BMI, body mass index; APACHE II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Sepsis-related Organ Failure Assessment score II; ICU, Intensive Care Unit; LOS, Length of stay.
Total number of measurements: 594
Mean blood glucose level per patient, mg/dL, median [IQR; total range]: 133 [118 – 140; 112 – 162]
Standard deviation of blood glucose level per patient, mg/dL, median [IQR; total range]: 15 [11 – 18; 1 – 49]
Number of measurements per patient, median, [IQR; total range]: 65 [26 – 80; 8 – 97]

| Mild hyperglycemia 150 -179 mg/dL in measurements, number (%) | 62 (10) |
| Mild hyperglycemia 150-179 mg/dL in patients, number (%) | 10 (91) |
| Severe hyperglycemia >180 mg/dL in measurements, number (%) | 29 (5) |
| Severe hyperglycemia >180 mg/dL in patients, number (%) | 3 (27) |
| Severe hypoglycemia ≤40 mg/dL in measurements, number (%) | 0 |
| Severe hypoglycemia ≤40 mg/dL in patients, number (%) | 0 |
| Mild hypoglycemia 41-70 mg/dL in measurements, number (%) | 0 |
| Mild hypoglycemia 41-70 mg/dL in patients, number (%) | 0 |

Table 2.
Metrics of glucose control. IQR, Interquartile range

| Total number of sensors used | 16 |
| Number of sensors used, median [IQR; total range] | 1.0 [1.0–2.0; 1 – 2] |
| Total connection time, median [IQR; total range] | 50.8 [13.5–54.7; 2.5 – 55.8] hours |
| Real–time data, median [IQR; total range] (%) | 50.8 [13.5–54.7; 2.5 – 55.8] hours (100%) |
| Time of skips in data acquisition | 0.0 hours |
| Percentage of time skips in data acquisition | 0.0 hours |
| Initialization time, median [IQR; total range] | 42 [42–43; 40.8 – 62.4] minutes |
| Total start-up time, median [IQR; total range] | 58 [56–67; 48 – 112.8] minutes |
| Number of calibrations needed before start, median [IQR; total range] | 1.0 [1.0–1.3; 1 – 3] |
| Number of calibrations during duration of measurement, median [IQR; total range] | 6.5 [2.0–7.0; 1 – 8] |
| Number of failed calibrations during duration of measurement, median [IQR; total range] | 0.5 [0.0–2.0; 0 – 3] |

Table 3.
Reliability and safety of the CGM device
SUPPLEMENTAL INFORMATION

Guideline for glucose control
A local guideline aiming at a blood glucose level between 90–144 mg/dL (5–8 mmol/L) was followed as part of standard care with insulin infusion being initiated when glucose levels rose over 144 mg/dL. When glucose levels were below 61 mg/dL insulin infusion was stopped and boluses of dextrose were given. Sliding scales were used for the adjustment of insulin titration. Insulin was only infused intravenously and in a continuous manner and boluses were exclusively administered when blood glucose levels rose over 360 mg/dL. The guideline mandated nurses to measure blood glucose at least every four hours, or more frequently when glucose levels were out of range or when rapid changes were expected. Blood gas analyzers (RAPIDLab 1265, Siemens Healthcare Diagnostics, The Hague, The Netherlands) were used to analyze the samples. Data was stored in the patient data management system.

Training of the nurses and the use of the device
Before the study device was introduced into the unit, nurses were trained on how to use and calibrate the device. When a patient was eligible for inclusion in the study and after placement of the special CVC was placed the device was connected by the researchers who performed the first calibrations. Nurses were instructed on how to use the CVC, in particular not to flush the two special ports used by the CGM device. In addition, both ports were labeled with adhesive tags clearly showing the following text: ‘DO NOT FLUSH’ (in Dutch: ‘NIET FLUSHEN’).

Metrics for device reliability
- Total number of sensors used – Total number of sensors used between all patients.
- Number of sensors used – Number of sensors used per patient.
- Total connection time – Total connection time the system was connected to a patient.
- Real-time data – Total time the data was displayed in real-time.
- Time of skips in data acquisition – Time the data was not displayed.
- Percentage of time skips in data acquisition – Percentage of the time the data was not displayed.
- Initialization time – Total time from connecting the device, to being ready for calibration.
- Total start-up time - Total time from connecting the device, to displaying the first glucose value.
- Number of calibrations needed before start – Number of calibrations that
were necessary before first glucose value was displayed.

- Number of calibrations during duration of measurement – Total number of calibrations that were performed during the time the device was connected to the patient.
- Number of failed calibrations during duration of measurement – Number of calibrations that failed during the time the device was connected to the patient.

Reliability and practical problems

In three patients, the initial sensor could not be calibrated at start-up, and a second sensor was needed. In one patient, the device gave repeated calibration problems, also after replacing the sensor. No measurements could be taken in this patient. This is the aforementioned patient that was excluded from the point and trend accuracy analysis. Calibrated sensors had no down-times and displayed values for 100% of the connection time. In one patient, a non-supervised but trained ICU nurse unintentionally flushed the afferent port of the CVC, causing an abrupt rupture of the semipermeable membrane, and consequently this CVC could no longer be used for continuous blood glucose monitoring (data from this patient were excluded from the point and trend accuracy analysis). In another patient the CVC malfunctioned for an unknown reason. While device data showed a stop in flow, possibly caused by membrane rupture, the nurse denied flushing the special ports of the CVC. Figure S2 shows that the trend in blood glucose levels was comparable to that of blood lactate levels. The non-physiologic drop in lactate allowed us to identify malfunctioning of the CVC, for yet unknown reasons.

POST-HOC ANALYSIS EXCLUDING DATA FROM ONE PATIENT WITH A MALFUNCTIONING CVC

Methods

To investigate if the special CVC was malfunctioning in this patient, we analyzed both continuous glucose and lactate data as measured by EIRUS® system and plotted the values against reference values from our blood gas analyzer. To investigate the point and trend accuracy of EIRUS® system without the patient in whom the special CVC was malfunctioning, data from the aforementioned patient was excluded. Thereafter, the same instruments were used to analyze point – and trend accuracy. Point accuracy was expressed using a Clarke error grid, a Bland–Altman plot, the glucose prediction error analysis, the mean absolute relative difference (MARD) and the Surveillance error grid. Trend accuracy was expressed using rate error grid Analysis (R–EGA)
Results
As can been seen in figure S2, both glucose and lactate values as measured by the EIRUSTM drop significantly within 15 minutes (133 mg/dL to 90 mg/dL and 23 mg/dL to 17 mg/dL, respectively). We consider this fast change to be implausible. In addition, the monitor intermittently displayed the message that it was flushing the line. Therefore, we suspect that the CVC was not performing correctly, possibly after being flushed.
A total of 582 paired measurements in 10 patients were analyzed. The Clarke error grid, Bland–Altman plot, and glucose prediction error grid of the post-hoc analysis are presented in Figure S3. Bias in the Bland–Altman plot was 4.0 mg/dL with an upper limit of agreement of 28.0 mg/dL and a lower limit of agreement of -19.9 mg/dL. Glucose prediction error analysis showed that 94.3% of the values ≥ 75 mg/dL within twenty percent of the values measured by the blood gas analyzer were within range. The MARD was 7.3%. The rate error grid is presented in Figure S4. The Surveillance error grid is presented in figure S5.
Figure S1
Surveillance error grid with risk scores.
Figure S2.
Blood glucose, CGM and lactate values of patient with failing CVC
Figure S3.
Measures of point accuracy. Bland-Altman plot (upper-left panel), glucose prediction error grid (lower-left panel) and clarke error grid (right panel).

Figure S4.
Rate Error—Grid of the Continuous glucose—error grid analysis This grid is divided into similar zones as the clarke error grid. Perfectly trend accurate values are the dashed line in the middle.
Figure S5.
Post-hoc Surveillance error grid with risk scores.
Part 2
Glucose Prediction by Analysis of Exhaled Metabolites – a Systematic Review


BMC anesthesiology 2014, 14
ABSTRACT

**Background:** In critically ill patients, glucose control with insulin mandates time- and blood-consuming glucose monitoring. Blood glucose level fluctuations are accompanied by metabolomic changes that alter the composition of volatile organic compounds (VOC), which are detectable in exhaled breath. This review systematically summarizes the available data on the ability of changes in VOC composition to predict blood glucose levels and changes in blood glucose levels.

**Methods:** A systematic search was performed in PubMed. Studies were included when an association between blood glucose levels and VOCs in exhaled air was investigated, using a technique that allows for separation, quantification and identification of individual VOCs. Only studies on humans were included.

**Results:** Nine studies were included out of 1041 identified in the search. Authors of seven studies observed a significant correlation between blood glucose levels and selected VOCs in exhaled air. Authors of two studies did not observe a strong correlation. Blood glucose levels were associated with the following VOCs: ketone bodies (e.g., acetone), VOCs produced by gut flora (e.g., ethanol, methanol, and propane), exogenous compounds (e.g., ethyl benzene, o-xylene, and m/p-xylene) and markers of oxidative stress (e.g., methyl nitrate, 2-pentyl nitrate, and CO).

**Conclusion:** There is a relation between blood glucose levels and VOC composition in exhaled air. These results warrant clinical validation of exhaled breath analysis to monitor blood glucose levels.

Keywords: Glucose, Monitoring, Volatile organic compound, Breath

BACKGROUND

Many, if not all, critically ill patients are treated with insulin at some point during their stay in the intensive care unit (ICU) [1]. Intensive monitoring of the blood glucose level is a prerequisite for both efficient and safe insulin titration in these patients [2]. Current practice in the ICU holds that glucose levels are monitored manually through intermittent measurements of the blood glucose level in central laboratories or using laboratory-based blood gas analyzers and/or glucose strips at the bedside [3]. Intermittent manual glucose monitoring however, is expensive and time consuming [4]. Moreover, intermittent glucose monitoring lacks the ability to detect temporal trends, potentially causing dangerous insulin titration errors in critically ill patients [5]. Glucose is a central molecule in metabolism [6,7]. Indeed, metabolic pathways are activated to maintain normoglycemia when the concentration of glucose changes [6,8]. Changes in the activity of these pathways could result in changes in production of volatile metabolites. These so-called volatile organic compounds (VOCs) can be detected in exhaled breath [9].
We hypothesize that there is an association between VOCs in exhaled breath and blood glucose levels. Previous excellent reviews focused on the correlation between glucose and exhaled breath condensate (thus soluble markers) [10] in diabetes [11,12], but none compared all available literature or discussed the implications for the ICU population. The specific aim of this systematic review is to provide an overview of the available data on the association breath VOCs and blood glucose levels and to discuss techniques for VOC detection.

METHODS

This systematic review was done according to standard methodology [13,14]. Medline was searched through Pubmed using the following search terms: (“Blood Glucose”[Mesh] OR “Glucose”[MeSH Terms] OR glucose[tiab]) AND (“Exhalat*”[MeSH Terms] OR “Volatile Organic Compounds”[Mesh] OR exhal*[tiab] OR Volatile Organic Compound*[tiab] OR Volatile Compound*[tiab] OR electronic nose[tiab] OR breath[tiab]. The search was conducted on the 3rd of January 2014. No limits were used for year of publication and language. Only human studies were included, with no restriction on subject health, age, gender or study setting.

Two independent researchers (JHL, LDB) selected articles for full-text assessment when the title and abstract suggested investigating the use of exhaled breath to measure or estimate blood glucose levels. Articles were only included if an association between blood glucose levels and VOCs in exhaled air was investigated. Also, VOC compositions of exhaled air had to be measured by an analytical technique that allows for separation, quantification and identification of individual VOCs, including gas chromatography and mass spectrometry (GC–MS), ion mobility mass spectroscopy (IMS), ion–molecule reaction mass spectrometry (IMR–MS), proton transfer reaction (time of flight) mass spectrometry (PTR(−TOF)-MS) and/or selected ion flow tube mass spectrometry (SIFT–MS).

Data from included studies were extracted and methodological quality was assessed independently by two researchers (JHL, LDB) using the QUADAS–2 tool for quality assessment [15]. The tool was adapted to be more relevant to the included literature. Disagreement between the two reviewers on inclusion of studies was resolved by consensus. The adjusted version of QUADAS–2 is presented in Additional file 1. Extracted data included: 1) characteristics of the study (design, year of publication and country of study conduction); 2) characteristics of the study population (including age, sex distribution and health status); 3) characteristics of the index test (including technique and included VOCs); 4) characteristics of the reference standard (blood glucose); 5) characteristics of the outcome (including main results and correlation coefficient between exhaled breath and glucose); 6) statistical validation technique used.
RESULTS

Search results
The literature search in Pubmed yielded 1041 titles (figure 1). After reading titles and abstracts, 1012 articles were excluded because the topic was outside of the scope of this review and 29 articles were retained for full-text assessment. After the exclusion of 20 papers (5 reviews/non-original studies, 13 on unrelated topics, 2 index test not compliant with inclusion criteria), 9 articles were included in the analysis. Characteristics of selected articles are presented in Table 1. Five studies included healthy non-diabetic subjects, two studies included Type 1 Diabetes Mellitus (T1DM) subjects, one study included Type 2 Diabetes Mellitus (T2DM) subjects and one study included both healthy and T1DM subjects.

Results of the quality assessment using the QUADAS–2-tool are presented in Table 2. The risk of bias was considered high for all studies; none of the studies used a random sample of patients, all using a pre-specified target group such as T1DM or T2DM patients. The use of blood gas measurements or central lab measurements was considered to be the correct reference standard [16]. The adequate reference standard was used in four studies. Four studies used finger prick measurement, which increases the possibility of incorrect insulin titration in clinical practice [16]. Comparing these measurements to exhaled breath could lead to biased results. However, none of these studies were excluded from our review.

Point correlation
Authors of seven out of nine studies found a strong correlation between one or more metabolites in exhaled breath and blood glucose levels, with a mean linear regression coefficient of 0.82 [range: 0.08-0.98] (Table 1). However, none of the included studies validated their results internally (e.g. cross-validation) or externally (e.g. in an separate validation cohort). A total of ten metabolites have been reported to correlate with blood glucose levels, including exhaled acetone, VOCs produced by gut flora (ethanol, methanol, and propane), exogenous compounds (Ethyl benzene, o-xylene, and m/p-xylene) and VOCs that reflect oxidative status (methyl nitrate, 2-pentyl nitrate, and carbon monoxide (CO)) (Table 3). Authors of two studies did not observe a strong correlation. The first one of those did not find a significant correlation between a single measurement of breath acetone and blood glucose in T2DM subjects. Authors of the second study were unable to demonstrate a strong correlation between glucose levels and exhaled CO in healthy subjects. Researchers in one of the studies that did show a strong correlation between breath metabolites and glucose levels, only observed this after overnight fast, showing a weak correlation after consuming a meal [17].
Temporal association
Researchers in seven out of nine studies performed multiple measurements with an interval ranging from 2.5 to 40 minutes. Two studies had a cross-sectional design and only performed a single measurement, or two unpaired measurements. None of the authors of the included studies reported on the possibility of predicting glucose trend.

DISCUSSION
This systematic review identified nine studies that investigated the ability of exhaled breath to measure or estimate blood glucose levels. A significant correlation between VOCs in exhaled breath and blood glucose levels was found in seven studies. These results indicate that there is an association between the two, although not all studies are consistent. Researchers in one of these seven studies only found a strong correlation after overnight fasting of the subjects and were unable to replicate the results after a meal [17]. Authors of two negative studies did not find a strong correlation, possibly due to a different study design [18] and the VOC (e.g. CO) that was studied [19]. Authors of the study that included subjects with T2DM did not show a significant correlation between exhaled VOCs and blood glucose levels. This study also had a different (cross-sectional) design. The analytical technique used for VOC detection did not modify the reported correlation. None of the studies monitored breath continuously. Also the glucose trend, thus the temporal association between glucose and exhaled VOCs, was not explicitly investigated. However, the data from three longitudinal studies [20-22] suggest that trends in glucose levels could possibly be monitored using exhaled breath when measurements are taken more frequently.

Index tests: exhaled breath analysis
A significant correlation between metabolites in exhaled breath and blood glucose levels was found using GC-MS, SIFT-MS, PTR-TOF-MS, a nano-sensing film-based sensor, and an electrochemical analyzer as analytical method. GC-MS is considered to be the gold standard for VOC detection and has shown to have a high sensitivity to identify single VOCs [23]. Therefore, GC-MS is suitable to accurately quantify a number of different VOCs in a cross-sectional study. However, the time-consuming nature of the technique limits use of the device for real-time and continuous measurements, which hampers clinical application. Other analytical techniques such as SIFT-MS [22,23] and PTR-TOF-MS [23-25] can also identify single VOCs and can be used for real-time continuous measurements. Disadvantages include possible selection bias [26] and the limitation to the concentration range that can be detected [25]. The electrochemical analyzers used in selected studies are two different Smo-
kerlyzer Micro (Bedfont, UK) devices. These devices measure the amount of CO in exhaled breath. However, there is a cross sensitivity to hydrogen [19]. While a correlation between exhaled CO and glucose levels was found by the researchers of one study [27], researchers of another study [19] could not reproduce these findings. Contrasting results may be due to the high cross sensitivity to hydrogen in the electrochemical analyzer used previously [27], which was less apparent using a newer device [19]. This exemplifies the importance of an adequate analytical technique that suits the aim of the study.

An important limitation of the techniques used in included studies is that none of them was used to continuously monitor exhaled breath. Continuous analysis of the exhaled breath was previously described by means of IMR-MS [28], PTR-MS [24], and electronic nose [29,30]. After a training phase, electronic noses learn to recognize specific disease states and can therefore be used for classification. The devices cannot identify and quantify single VOCs, but they do give a rapid, bedside diagnosis, which, from a clinical perspective, renders this device attractive. The electronic nose has been used to discriminate between patients with and without diabetes [31]. One could postulate that the ability to diagnose diabetes is partly due to the metabolomic alterations because of higher blood glucose levels in diabetic patients. Therefore, electronic nose analysis may complement mass-spectrometry based techniques for the monitoring of blood glucose levels in clinical practice, providing signals based on probabilistic training and validation. Alternatively to semi-selective recognition, nanosensors also rely on specific recognition of certain VOCs [30]. In one study, an acetone-selective nanomaterial-based sensor was used alongside PTR-TOF-MS and showed a strong correlation between acetone, glucose and the sensor [17]. Small size of devices using nanomaterial-based sensors as compared to spectrometry-based methods facilitates clinical application.

Mechanisms related to the association between VOCs and glucose levels can be found in Table 3. Acetone appeared to be associated with blood glucose levels [17,20-22,32]. As a result of increased synthesis of acetone and degradation of ketone bodies, acetone is expected to be higher in diabetics [33]. On the other hand, healthy humans only have elevated levels of ketone bodies when fasting or exercising [11]. Therefore, it is more likely to find a correlation between exhaled acetone and glucose levels after fasting compared to finding a correlation after lunch [17]. Acetone possibly is a good marker for glucose levels in ICU-patients. However, the large variation in breath acetone levels between subjects [18,34-36] may result in low accuracy when using acetone cross-sectionally.

VOCs such as ethanol [17,20,22,32], propane [22] and methanol [17,22] are likely to reflect gut flora activity, since the metabolism of gut bacteria is responsive
to glycemic fluctuations [22,32]. However, we cannot exclude that other biochemical pathways also contribute to the production of these compounds. In critically ill patients on the ICU, the quantity and composition of the gut microbiome are changing over time and therefore the amount of VOCs they produce may not be stable [37]. Therefore, these markers are less likely to predict glucose levels in ICU-patients.

Ethyl benzene [20,22], o-xylene [20] and m/p-xylene [20] are gasses that are inhaled, partially metabolized by the liver and subsequently exhaled at lower concentrations [20]. Rapid-onset hyperglycemia likely suppresses hepatic metabolism, thus causing peak concentrations of these compounds in exhaled air. Recent evidence suggests that cyclic hydrocarbons such as ethyl benzene and xylene are emitted by the ventilator and tubing [38]. Given that exhaled air is readily accessible for measurements in mechanically ventilated ICU patients, use of exhaled air for the prediction of glucose levels is therefore plausible.

An isomer of methyl nitrate [22,39] is formed when methanol reacts with nitric oxide, which in turn reacts with superoxide ion (O$_2^{-}$), a by-product of oxidative reactions [39]. Furthermore, 2-pentyl nitrate [22] is generated through pathways involving organic peroxy radical (RO$_2^-$) and NO or NO$_2$. This could be modulated by acute changes in systematic oxidative status [22]. Changes in CO [27] in exhaled breath are possibly related to oxidative stress. When glucose levels rise, particularly in diabetic patients, this can lead to oxidative stress. As a protective response, heme oxygenase is activated, leading to the positive modulation of CO on insulin secretion [27]. For critically ill patients on the ICU however, markers of oxidative stress will be non-specific for high blood glucose, as they increase with any form of oxidative stress such as sepsis, high inspired-oxygen fraction and acute respiratory distress syndrome [40].

**Study design**

The observed correlation between blood glucose levels and exhaled VOCs may be due to the inclusion of T2DM patients and/or a cross-sectional study design. First, T2DM influences the responsiveness of the body to changes in blood glucose levels [41]. This is typically characterized by insulin resistance but may also influence the formation of ketone bodies and the induction of liver enzymes. Second, breath acetone levels tend to differ between T1DM, T2DM, and healthy subjects [18,21,35]. Therefore, a decrease in blood glucose levels may not induce the same rise in breath acetone levels with different baseline values and in the context of different co-morbidities. Finally, in line with the previous point, correction for baseline differences between subjects cannot be accomplished with a cross-sectional study design. This is further acknowledged by the fact that the predictive algorithm requires calibration for every subject in several studies [20,22]. Since the relation between exhaled breath metabolites
and blood glucose levels shows high inter-person variation, a cross-sectional design may not be ideal for predicting glucose levels using breath metabolites. The possibility of using a single breath maneuver to estimate blood glucose levels thus seems implausible. Future studies may therefore focus on longitudinal measurements in the same subject.

Five included studies used a clamp study design and 2 studies used an oral glucose tolerance test (OGTT). Clamp studies and OGTT result in a more or less predictable course of blood glucose levels. Although a clamp design is ideal for research purposes and enables comparability between studies, clinical practice is often very different and less predictable. The transition of the results of these studies to the clinical setting will be a major challenge for the field of blood glucose estimation by exhaled breath analysis.

**Strengths and limitations**

We used a standardized systematic review approach, combining all evidence available. All VOCs that are linked to changes in glucose levels are discussed and their most likely biochemical pathways are described. In addition, we carefully assessed the quality of the included studies. This systematic review also has an important limitation. The included studies were highly heterogeneous with respect to patient selection, exhaled breath sampling and analysis and blood glucose measurement, limiting the comparability of the studies. Therefore, we decided to describe the results separately. Most of the included studies had a relatively high risk of bias and we found that included studies did not validate their results. Possibly, this is inevitable in the early stages of biomedical research but it hinders strong conclusions. Furthermore, models can possibly be overfit, yielding overoptimistic results. Our search only identified one negative study. Negative studies are often not published leading to publication bias.

None of the studies investigated ICU-patients, while glucose fluctuations are large and frequent in this population [42]. Therefore, we cannot draw firm conclusions on the use of these methods in ICU-patients. We did try to identify potential pitfalls for the implementation of these methods in ICU patients by reviewing the biochemical pathways for the formation of VOCs. Finally, the use of exhaled breath to monitor glucose trends was not discussed in any of the articles. Monitoring glucose trends (in ICU patients) however, has several potential advantages over using single values. First, trend has a better predictive value compared to single glucose levels; recent trend can be used to predict future levels. In ICU patients, this can lead to improved insulin titration. Second, because outliers can be filtered out, trend is less susceptible to random noise. Third, possible bias (constantly predicting values too high/low) will be constant throughout the trend, having a smaller effect.
ges of using glucose trend are possible lag in the signal, and the potential of amplification of errors.

CONCLUSION

In conclusion, a significant association between VOCs in exhaled breath and blood glucose levels was found in the majority of studies included in this systematic review. Acetone, carbon monoxide, ethanol, ethyl benzene, M/P-xylene, methanol, O-xylene, and propane were correlated with blood glucose levels. Several potential effect modifiers were identified for ICU-patients. The included studies were performed under highly controlled circumstances, which limit generalizability. Our results warrant clinical validation of exhaled breath analysis for the monitoring of blood glucose levels in critically ill ICU-patients.

Abbreviations

CO, Carbon monoxide; GC–MS, Gas chromatography and mass spectrometry; ICU, Intensive care unit; IMS, Ion–mobility spectroscopy; OGTT, Oral glucose tolerance test; PTR-(TOF)-MS, Proton transfer reaction (time of flight) mass spectrometry; SIFT-MS, Selected ion flow tube mass spectrometry; T1DM, Type 1 Diabetes Mellitus; T2DM, Type 2 Diabetes Mellitus; VOC, Volatile organic compound

References

ANALYSIS IN THE INTENSIVE CARE UNIT

CONTINUOUS GLUCOSE AND EXHALED BREATH ANALYSIS IN THE INTENSIVE CARE UNIT


FIGURES

Articles identified through database searching (n = 1041)

Articles excluded based on titles (n = 1012)

Full-text articles assessed for eligibility (n = 29)

Full-text articles excluded (n = 20)
- Review / non-original research: 5
- Different topic: 13
- Index test not compliant with inclusion criteria: 2

Studies included in Review (n = 9)

Figure 1.
<table>
<thead>
<tr>
<th>First Author</th>
<th>Setting</th>
<th>Patients</th>
<th>Age</th>
<th>Sex distribution</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Main Results</th>
<th>Mean correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Righettoni [17]</td>
<td>Healthy subjects sampled after overnight fast and after lunch.</td>
<td>8</td>
<td>22-55 years</td>
<td>7 male, 1 female</td>
<td>PTR-TOF-MS: Acetone, ethanol, methanol, isoprene</td>
<td>Finger prick measurement with Bayer Contour Blood Glucose Meter</td>
<td>After overnight fast a high correlation between sensors and glucose, and acetone, ethanol, methanol and isoprene was found. These high correlations were not found after lunch.</td>
<td>Morning: PTR-TOF-MS: Acetone: 0.98, Ethanol: 0.9, Methanol: 0.93, Isoprene: 0.00. Afternoon: PTR-TOF-MS: Acetone: -0.08, Ethanol: -0.11, Methanol: -0.16, Isoprene: -0.40. Nano sensing films: 0.96.</td>
</tr>
<tr>
<td>Storer [18]</td>
<td>T2DM subjects not asked to fast but to refrain from eating. Cross-sectional study</td>
<td>38, T2DM</td>
<td>32-76 years, median age 62</td>
<td>13 male, 25 female</td>
<td>SIFT-MS: Acetone</td>
<td>Finger prick measurement with Abbot Optium Xceed</td>
<td>No strong correlation found between blood glucose and breath acetone. Breath acetone was found to be significantly higher in men.</td>
<td>r = 0.003</td>
</tr>
</tbody>
</table>
Minh [22] Clamp study. Overnight fast. T1DM subjects were asked not to take long-acting insulin. Healthy: 28 ± 1 years, T1DM: 25.8 ± 1.7 years, 11 male, 14 female. GCMS: Group A (Ethanol, acetone, methyl nitrate, ethyl-benzene) Group B (2-pentyl nitrate, propane, methanol, ethanol), Room samples collected.

Turner [21] Clamp study. Overnight fast. T1DM subjects. Healthy: r = 0.8325, T1DM: r = 0.935. No strong correlation at baseline. Linear correlation between acetone and blood glucose values. Breath acetone decreased when blood glucose decreased. In healthy volunteers the opposite was seen: Low blood glucose values yield high acetone values.
<table>
<thead>
<tr>
<th>First Author</th>
<th>Setting</th>
<th>Patients</th>
<th>Age</th>
<th>Sex distribution</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Main Results</th>
<th>Mean correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee [20]</td>
<td>Clamp study. Healthy subjects admitted to lab after overnight fast.</td>
<td>10</td>
<td>26 ± 4 years</td>
<td>5 male, 5 female</td>
<td>GCMS: Ethanol, Acetone, Methyl nitrate, Ethylbenzene, o-oxylene, m/p-xylene. Room samples collected.</td>
<td>IV catheters in antecubital veins; Beckman Glucose analyzer II</td>
<td>Best 4 gas model: Ethanol, acetone, methyl nitrate, ethyl benzene (mean r of 0.913(0.698-0.977)) 9 samples per patient</td>
<td>r = 0.913 (0.698-0.977)</td>
</tr>
<tr>
<td>Fritsch [19]</td>
<td>OGTT. Healthy volunteers admitted after 10 hours fast.</td>
<td>6</td>
<td>24-32 years</td>
<td>5 male, 1 female</td>
<td>Electro-chemical analyzer, laser spectrometer, and breath hydrogen: Carbon monoxide measured with Micro smokerlyzer.</td>
<td>Finger prick measurement, Accu check Aviva.</td>
<td>No strong correlation between glucose and carbon monoxide</td>
<td>None</td>
</tr>
<tr>
<td>Novak [39]</td>
<td>Clamp study. T1DM subjects admitted after eating light breakfast. Patients on insulin followed normal regimen.</td>
<td>10, T1DM</td>
<td>13.8 ± 0.5 years</td>
<td>7 male, 3 female</td>
<td>GCMS: Methyl nitrate. Room samples collected.</td>
<td>IV lines in arms, Blood samples every 30 min. Beckman glucose analyzer II</td>
<td>Methyl nitrate had strongest correlation with blood glucose levels. Correlation increased with 30-minute lag time. Ethanol and Acetone DID NOT correlate with glucose</td>
<td>One subject mentioned, r = 0.99</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Participants</td>
<td>Procedure</td>
<td>Analysis</td>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galassetti [32]</td>
<td>OGTT. Healthy subjects admitted to research center in morning after overnight fast.</td>
<td>10</td>
<td>27.4 ± 3.1 male, 5 female</td>
<td>GCMS: Ethanol and acetone. Room samples collected. IV catheter. Determined with a quantitative enzymatic measurement.</td>
<td>Multiple linear regression analysis with ethanol and acetone gave an average r of 0.70.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paredi [27]</td>
<td>OGTT in 5 patients, CO and glucose measured. Only CO measured in larger cohort</td>
<td>5</td>
<td>33±4 years 3 male, 2 female</td>
<td>Micro smokerlyzer: Carbon monoxide Finger prick measurement, RefloluX S.</td>
<td>The maximal glucose increase was associated with a significant increase in exhaled CO concentration. Both parameters returned to the baseline at 40 min after glucose administration.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.
<table>
<thead>
<tr>
<th>Study</th>
<th>Risk Of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minh (2011) [22]</td>
<td>High</td>
<td>High, Reference Standard High, Flow And Timing High, Index Test High, Reference Standard High, Comments: Clamp study design possibly lowers clinical relevance because of lack of generalizability. Test review bias because reference standard is used for index test.</td>
</tr>
<tr>
<td>Study</td>
<td>Bias</td>
<td>Recall</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Paredi (1999) [27]</td>
<td>?</td>
<td>High</td>
</tr>
</tbody>
</table>

Clamp study design possibly lowers clinical relevance because of lack of generalizability. Test review bias because reference standard is used for index test. Possible reporting error, results of only one subject mentioned in detail.

Small sample size. OGTT study design possibly lowers clinical relevance because of lack of generalizability. Test review bias because reference standard is used for index test. Possible verification bias because of incorrect reference standard.

Small sample size. OGTT study design possibly lowers clinical relevance because of lack of generalizability. Possible verification bias because of incorrect reference standard.

Table 2.
Results of QUADAS-2 tool
<table>
<thead>
<tr>
<th>VOC</th>
<th>Mechanism(s)</th>
<th>Pathway(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-pentyl nitrate [22]</td>
<td>Generated through pathways involving organic peroxy radical (RO2) with NO or NO2. Could be modulated by acute changes in systematic oxidative status [22]</td>
<td></td>
</tr>
<tr>
<td>Acetone [20–22, 32]</td>
<td>Derived from acetoacetate and is produced by synthesis and degradation of ketone bodies and is therefore related to blood glucose levels [32].</td>
<td>Glycolysis / Pyruvate metabolism</td>
</tr>
<tr>
<td>Cabon monoxide [27]</td>
<td>Possibly due to activation of HO by glucose, and the positive modulation of CO non insulin secretion [27].</td>
<td></td>
</tr>
<tr>
<td>Ethanol [20, 22, 32]</td>
<td>Not produced by mammalian cells. Likely due to alcoholic fermentation of glucose by gut bacteria and yeast [32].</td>
<td>Glycolysis / Gluconeogenesis</td>
</tr>
<tr>
<td>Ethyl benzene [20, 22]</td>
<td>Inhaled and partly metabolized by liver, then exhaled at lower concentration. Rapid-onset hyperglycemia likely suppressed hepatic metabolism causing peaks in exhaled air [20].</td>
<td></td>
</tr>
<tr>
<td>M/P-xylene [20]</td>
<td>Inhaled and partly metabolized by liver, then exhaled at lower concentration. Rapid-onset hyperglycemia likely suppressed hepatic metabolism causing peaks in exhaled air [20].</td>
<td></td>
</tr>
<tr>
<td>Methanol [22]</td>
<td>Reflects gut flora activity and therefore responsive to glycemic fluctuations [22]</td>
<td></td>
</tr>
<tr>
<td>Methyl nitrate [22, 39]</td>
<td>A small fraction of superoxide ion (O2−), a byproduct of oxidative reactions, reacts with nitric oxide which in turn can react with methanol to eventually form an isomer of Methyl nitrate [39].</td>
<td></td>
</tr>
<tr>
<td>O-xylene [20]</td>
<td>Inhaled and partly metabolized by liver, then exhaled at lower concentration. Rapid-onset hyperglycemia likely suppressed hepatic metabolism causing peaks in exhaled air [20].</td>
<td>N-4 fatty acid, Peroxidation, Protein oxidation</td>
</tr>
</tbody>
</table>
CONTINUOUS GLUCOSE AND EXHALED BREATH ANALYSIS IN THE INTENSIVE CARE UNIT
Factors influencing continuous breath signal in intubated and mechanically–ventilated intensive care unit patients measured by an electronic nose


Sensors 2016, 8
ABSTRACT

Introduction: Continuous breath analysis by electronic nose (eNose) technology in the intensive care unit (ICU) may be useful in monitoring (patho)physiological changes. However, the application of breath monitoring in a non-controlled clinical setting introduces noise into the data. We hypothesized that the sensor signal is influenced by (1) humidity in the side-stream; (2) patient–ventilator disconnections and the nebulization of medication; (3) changes in ventilator settings and the amount of exhaled CO₂. We aimed to explore whether the aforementioned factors introduce noise into the signal, and discuss several approaches to reduce this noise.

Methods: Study in mechanically-ventilated ICU patients. Exhaled breath was monitored using a continuous eNose with metal oxide sensors. Linear (mixed) models were used to study hypothesized associations.

Results: 1251 hours of eNose data was collected. First, the initial 15 minutes of the signal was discarded. There was a negative association between humidity and sensor 1 (Fixed-effect β: -0.05 [95%-confidence interval: -0.06 – -0.05] and a positive association with sensors 2-4 (Fixed-effect β: 0.12 [95%-confidence interval: 0.11 – 0.12]); the signal was corrected for this noise. Outliers were most likely due to noise and therefore removed. Sensor values were positively associated with end-tidal CO₂, tidal volume and the pressure variables. The signal was corrected for changes in these ventilator variables after which the associations disappeared.

Conclusion: Variations in humidity, ventilator disconnections, nebulization of medication and changes of ventilator settings indeed influenced exhaled breath signals measured in ventilated patients by continuous eNose analysis. We discussed several approaches to reduce the effects of these noise inducing variables.

INTRODUCTION

Introduction into breath analysis
Analysis of exhaled breath has received increasing attention over the last years as a potential diagnostic tool for a variety of diseases. Several technologies are available for breath analysis [1]. Gas chromatography and mass-spectrometry (GC-MS) is an often used technique to obtain a detailed snapshot of the volatile organic compounds (VOCs) in breath. Therefore, GC-MS is well-suited for biomarker discovery and discrimination between disease-states at a single time-point. However, for clinical practice monitoring of dynamic processes is frequently more desired than a singular measurement representing a stable state. Monitoring of several specific VOCs has been successfully attempted with proton-transfer reaction time-of-flight mass-spectrometry (PTR-TOF-MS) [2]. The PTR-TOF-MS device however is large, with limited possibility for mini-
aturization, expensive and data analysis is elaborate. Another possibility for continuous analysis of VOCs is by cheaper and more portable electronic nose (eNose) technology [3-12]. Many eNose studies have shown its potential [13-16]. eNoses contain an array of cross-reactive sensors. The physical properties (such as electrical conductivity) of these sensors change upon exposure to certain VOCs. The sensors are typically very sensitive for a wide range of VOCs but show remarkable cross-reactivity. Therefore, the analysis of eNose data relies on pattern recognition.

Ventilated ICU patients
Intensive care unit (ICU) patients are per definition severely ill and their physiology is easily disturbed, which makes their clinical condition highly unstable. Therefore, there is a need for continuous monitoring, as is standard of care for several physiological parameters such as pulse oxymetry and end-tidal CO2 monitoring in exhaled air [17,18]. The latter is an example that illustrates the advantage of breath analysis over other bio-monitoring techniques, such as frequent blood draws and expensive blood gas analyses. Therefore, breath analysis may be desirable as bio-monitoring technique in the ICU and, because of constant access to breath in intubated and mechanically-ventilated patients, continuous breath analysis is highly feasible. Breath can be collected and analyzed fully non-invasively in these patients [19], but so far continuous measurements have not been investigated. Therefore, several challenges have to be faced before safe and meaningful continuous analysis of the exhaled breath is possible in intubated and ventilated ICU patients. Continuous exhaled breath analysis should be safe and should never interfere with any of the clinical activities, such as ventilatory support and monitoring. However, the application in a non-controlled, non-laboratory setting is likely to introduce noise into the data. The aim of this study was to evaluate noise inducing variables in exhaled breath signals as obtained through an eNose sensor in a non-controlled setting and to discuss several approaches to reduce noise. We hypothesized that the sensor signal was influenced by (1) the humidity in the side-stream, as the sensor are cross-reactive with water; (2) patient-ventilator disconnections and the nebulization of medication; (3) changes in ventilator settings and/or exhaled CO2.

METHODS
Study design and population
Intubated and mechanically-ventilated ICU patients, expected to remain mechanically-ventilated for at least 48 hours, were eligible for this study. Additionally, patients had to be older than 18 years. Patients were excluded when they were not expected to survive for a considerable amount of time. The
Institutional Review Board of the Academic Medical Center, Amsterdam, the Netherlands concluded that the legislation on human participation in research was not applicable because of the non-invasive, non-interventional nature of the study.

**Standard of care**

Standard of ventilatory support for intubated and mechanically-ventilated patients on the ICU of the Academic Medical Center in Amsterdam included, but was not limited to: Pressure Support or Pressure Controlled mode of ventilation, tidal volume of 6-8 mg/kg predicted body weight, a level of Positive End Expiratory Pressure (PEEP) ≥ 5 cm H2O, nebulization of acetylcysteine and salbutamol, and a pulmonary toilet including suction of secretions every 6 hours. In addition, a heat-moist exchanger was placed in the circuit and active humidification was not used routinely. Mechanical ventilators from several vendors were used. Arterial blood gas measurements were taken on indication, with a minimum of 3 measurements per day. These factors, and other aspects of clinical care were not influenced by study procedures.

**Study procedures and data collection**

Exhaled breath was monitored in patients of the mixed medical/surgical ICU of the Academic Medical Center in Amsterdam, the Netherlands using a continuous eNose adapted for clinical use in the intensive care unit (Comon Invent, Delft, the Netherlands). The sensor array contained 4 different metal oxide sensors (Figaro, Japan), which were chosen for their stability, clinical potential, performance and because they were widely used [14,20,21]. Tin dioxide was used as sensing material and the metal oxide sensors could operate between -40 and +70 °C. The device was similar to that described by De Vries et al. [14], with the following adaptions: (i) it consists of a single sensor array, (ii) it has a roller-pump to continuously supply exhaled breath at a flow of 35 mL/min, (iii) it has a plastic outer body to allow for thorough cleaning and (iv) it has an offline modus that disabled mobile connectivity to prevent interference with the mechanical ventilator and other devices [22]. The eNose was connected to the expiratory tube of the ventilator using a T-piece to create a side-stream. This is illustrated in figure 1. The flow of 35 mL/min to the eNose was not enough to trigger a ventilator alarm in our experiment. Data of the metal oxide sensors and a humidity sensor were stored every minute. Ventilator data were automatically stored in the Patient Data Management System (PDMS).

**Data analysis**

Data analysis was performed using R version 3.2.4.
Noise inducing variables
Noise inducing variables in the data were identified by plotting data for visual inspection by the investigators. Changes occurring concurrently in both sensor signals and variables were noted. The major noise inducing variables are plotted in figure 2. These include time delay to reach steady state, changes in humidity in the side-stream connector, disconnections, nebulization of medication and changing ventilator settings. Each of these causes of noise will be discussed in detail below.

Time delay to reach steady state
The sensors of the eNose have to adapt to the changing substances in the air when connecting an eNose to a subject. The left panel in Figure 2A shows this schematically. Therefore, the first measurements after connecting the eNose cannot be relied on and must be deleted from the data. Different time periods were investigated and the period that resulted in stabilization in all patients was chosen.

Changes in humidity
The humidity in side-stream connector leading to the eNose may cause a change in the sensor response, unrelated to the actual VOC profile (Figure 2, panel B). The association between humidity and sensor response was investigated by Pearson's correlation coefficient. A linear regression model was fitted with the sensor response as dependent variable and the humidity as independent variable. Standardized residuals of these regression models were used to replace sensor variables as these values are corrected for the variance imposed by changes in humidity.

Outlier removal and smoothing
After correcting for humidity, outliers, particularly those due to intermittent disconnections and nebulization of medication, were removed. As short periods of extreme values are most likely to be erroneous, we chose to discard the top and bottom 2.5% of the measurements. This removes the most prominent peaks and dips. Then, a LOESS smoother with default settings was calculated, after which the relative error between the signal and the smoother was computed. When this relative error was above a set threshold at a point in the data, this data point was replaced by the value of the LOESS smoother. Finally, to illuminate the worst jitter, the signal was smoothed using a LOESS smoother, with a span of 30 observations.

Changes in ventilation settings
Changes in ventilation settings could have a big influence on the eNose signal.
It can be imagined that a change in tidal volume, PEEP or minute volume may influence the abundance of measured VOCs. Ventilation settings stored in the PDMS were used to correct for this phenomenon. A similar strategy as used for correcting for changes in humidity was followed. However, instead of a linear regression model, a mixed model with the patient number as random effect was used. We investigated the available relevant variable, which are: changes in minute volume, changes in end tidal CO2 (ETCO2), changes in tidal volume, inspiratory pressure, peak pressure and PEEP. Backwards variable selection was used to eliminate non-significant effects. Finally, a LOESS smoother with a span of 15 observations was used once more to correct for jitter introduced after ventilator variable correction.

**RESULTS**

Between October of 2012 and July of 2015, 1251 hours of eNose data was collected in 23 different patients. Patient characteristics are shown in table 1.

Noise reduction

**Steady state**

The first fifteen minutes of sensor signal was discarded to allow the sensors to adapt to the new circumstances. A shorter period did not result in a steady state in all patients. The middle and right panel in Figure 2A illustrate this.

Changes in humidity

Figure 3 illustrates the mean correlation between sensors and relative humidity (RH) of all included patients. Influence of changes in humidity differed among patients, but it indicates that humidity indeed introduced noise into the signal. In figure 2B, influence of humidity, and signal after correcting for it are found. Correlation matrixes and plots for each individual patient can be found in supplement 1. The fixed effect of the mixed effects model for each sensor can be found in Table 2; there was a negative association between humidity and sensor 1 (-0.05 [-0.06 - -0.05] and a positive association with sensors 2-4 (0.12 [0.11 – 0.12]). The correlation coefficient r between humidity and eNose sensors for each patient can be found in Figure 4.

Outlier removal and smoothing

The middle and right panel of Figure 2C illustrate one of the sensor signals before and after outlier removal. As the example in the right panel demonstrates, large influential peaks and dips caused by outliers were removed from the signal. By removing these peaks, a less aggressive LOESS smoother was necessary to remove the remaining outliers and jitter from signal. This process is illustrated in the right panel of Figure 2D. Plots for all other patients are found in supplement 1.
Changes in ventilation settings

A matrix with mean correlation values between ventilator readings and sensors of all patients in our study is shown in Figure 5. The fixed effect of the mixed effects model for each sensor can be found in Table 3. Since ventilation settings varied greatly between patients, there was not one setting that seemed highly correlated with sensor signals in every patient (Figure 5, Figure 6). Nonetheless, when plotting raw sensor values and ventilator settings, change in ventilator settings seems to influence sensor values (Figure 2E). In addition, changes in settings were more likely to occur when patients were monitored for a longer period of time. After backwards variable selection, none of the variables was eliminated and sensors were corrected for all pre-defined variables. Table 4 shows the fixed effects of the mixed effects models after correcting for ventilator variables. Sensor values were positively associated with end-tidal CO2, tidal volume and the variables. Figure 7 illustrates the signal in one patient after each pre-processing step.

DISCUSSION

The presented results suggest that humidity, ventilator disconnections, nebulization of medication and ventilator settings indeed influenced exhaled breaths signal measured in ventilated patients by continues eNose analysis. We described several approaches to reduce these types of noise. This implies that direct translation of breath analysis technology developed for singular measurements is impossible; the signal should be corrected in a multi-step fashion to remove noise that correlates with variations in patient care that are not directly linked to the (patho)physiological process of interest.

This influence on the signal is explained by the following considerations. First, the commonly used metal oxide sensors that we also used in this study are known to be influenced by humidity [20,23]. Therefore, the eNose is connected behind the heat-moist exchanger of the ventilation circuit to minimize the influence of moist. The influence of moist was further diminished by correcting the signal for the remaining fluctuations in humidity. Second, disconnections of the eNose or ventilation circuit had a major effect on the signal because of a sudden inlet of ambient air. This leads to lower concentration of exhaled compounds, but an increase in concentration of, for example, ethanol. This has a large impact on sensor values. Third, frequent nebulization of medication such as acetylcysteine and salbutamol can influence sensor signals. When nebulized medication is not completely absorbed by the lungs and is consequently partially exhaled, it could possibly bind the eNose sensors and inflict a change in sensor signal. Finally, changes in ventilator settings were
associated with sensor readings. Increased inspiratory and end-expiratory pressure for example, may cause parts of the lung that were previously collapsed to open, thereby influencing the exhaled VOC mixture [24]. Also, Increased minute volume, while everything else is constant, decreases the concentration of systemically produced VOCs, in accordance with end-tidal CO2 [25].

Several strengths of this study should also be noted. First, correcting for confounding factors was possible because we recorded data systematically in a database at one-minute intervals. Therefore, distilling the actual underlying eNose signal was possible. Second, the long observation periods per patient ensured that we measured a large number of possible sources of noise. Third, the sensor array that has been used was shown to discriminate between asthmatics, COPD patients, lung cancer patients and healthy controls [14]. Therefore, it could be argued that it is a valid tool to use in our investigation. There are also several weaknesses. Since metal oxide sensors that are used in eNosbases are very cross-reactive, the analysis of this type of data relies on pattern recognition. Therefore, changes in individual VOC concentrations cannot be identified. Although this is not a limitation of the eNose as a measurement instrument, it does hamper us in identifying or quantifying individual VOCs in this study. While this is the first study to use cross-reactive sensors for continuous breath analysis in intubated and ventilated ICU-patients, PTR-MS has been studied in this setting [2]. Contrary to eNosbases, PTR-MS allows for breath-by-breath measurements of the concentration of specific VOCs. However, the large size and high costs of PTR-MS machines currently limit the application as a bedside test. Breath-by-breath responsiveness was not obtained with the used sensor array. Therefore, it only facilitates monitoring of changes over hours, not minute by minute. However, biological phenomena like changing glucose levels do not require this high frequency of measurements. Finally, we cannot be certain that some of the signal of interest is influenced by our methods for noise reduction.

The described steps to remove noise inducing variables from eNose signals is a first step towards continuous breath monitoring in a clinical setting. In addition, continuous monitoring of biological markers allows for trend analysis, which is impossible with infrequent blood draws. Next to monitoring systemic markers, investigating molecular processes in the lung can possibly be simplified by monitoring exhaled breath. Currently, this is only possible by performing a bronchoalveolar lavage, which is considered to be very invasive. Diagnostic accuracy however, should be studied and will be reported separately. Our findings help other researchers in their analysis and interpretation of their results are beneficial to developers of eNose technology.
To conclude, changing humidity influenced eNose sensors and sensor signals were corrected. After outlier removal and smoothing, the signal was corrected for changes in ventilator settings. Pre-processing is the first step toward using continuous monitoring of exhaled breath.

References

Figure 1. eNose and Side stream of the that is created to connect the eNose. In the upper panel, the eNose that is used in our study is pictured: (1) is the sensor block containing the metal oxide sensors. (2) Is the pump that is used to pump exhaled breath over the metal oxide sensors. The lower panel illustrates the side stream of the that is created to connect the eNose. A T-piece is connected distal of the heat moist exchanger in the ventilation circuit.
Figure 2.
Noise inducing variables in continuous breath signals of ventilated intensive care unit patients. From left to right, panel 1: schematic representation of type of noise. Panel 2: example of the noise in the data. Blue dashed lines indicate humidity in B and an example of a ventilator signal (end tidal CO2) in E. Panel 3: example after correction for the noise type. Green line indicates sensor signal after noise reduction step, red line indicates discarded signal values. In D, the grey line represents the initial values of one of the eNose sensors, the green line corresponds to the LOESS smoother and the red dots are values that are above the set threshold and are therefore considered outliers.
Figure 3.
Correlation between relative humidity (RH) and sensors 1 through 4. Values indicate fixed effect regression coefficient with the 95% confidence interval between brackets. Blue indicates positive correlation while red indicates negative correlation. Darker more elliptical shapes indicate greater correlation.
Figure 4.
Mean correlation coefficient $R$ between sensors and relative humidity. Upper and lower “hinges” correspond to first and third quartile, whiskers indicate 95% confidence interval.
Figure 5. Correlation between ventilator variables and sensors 1 through 4. Values indicate fixed effect regression coefficient with the 95% confidence interval between brackets. Blue indicates positive correlation while red indicates negative correlation. Darker more elliptical shapes indicate greater correlation.
Figure 6.
Mean correlation coefficient R between sensors and ventilator variables. Upper and lower “hinges” correspond to first and third quartile, whiskers indicate 95% confidence interval.
Figure 7.
eNose signal for one sensor in one patient after each pre-processing step.
TABLES.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, median [IQR]</td>
<td>67 [62 – 75]</td>
</tr>
<tr>
<td>Male gender, number (%)</td>
<td>9 39%</td>
</tr>
<tr>
<td>Admission type, number (%)</td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>21 91%</td>
</tr>
<tr>
<td>Planned Surgery</td>
<td>2 9%</td>
</tr>
<tr>
<td>APACHE II, median [IQR]</td>
<td>22 [19 – 28]</td>
</tr>
<tr>
<td>SAPS II, median [IQR]</td>
<td>52 [42 – 65]</td>
</tr>
<tr>
<td>ICU LOS, days, median [IQR]</td>
<td>12 [9 – 15]</td>
</tr>
<tr>
<td>ICU mortality, number (%)</td>
<td>10 43%</td>
</tr>
<tr>
<td>Measurement duration in hours, median [IQR]</td>
<td>51 [41 – 65]</td>
</tr>
</tbody>
</table>

Table 1.
Patient characteristics. IQR, Interquartile range; BMI, body mass index; APACHE II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Simplified Acute Physiology Score II; ICU, Intensive Care Unit; LOS, Length of stay.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Relative humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-0.056 [-0.058 – -0.054]</td>
</tr>
<tr>
<td>S2</td>
<td>0.116 [0.115 – 0.117]</td>
</tr>
<tr>
<td>S3</td>
<td>0.119 [0.112 – 0.120]</td>
</tr>
<tr>
<td>S4</td>
<td>0.118 [0.117 – 0.119]</td>
</tr>
</tbody>
</table>

Table 2.
Fixed effects regression coefficients of relative humidity vs sensors. Values indicate fixed effect regression coefficient with the 95% confidence interval between brackets.
<table>
<thead>
<tr>
<th>Sensor</th>
<th>Minute Volume</th>
<th>etCO2</th>
<th>Tidal Volume</th>
<th>Inspiratory pressure</th>
<th>Peak pressure</th>
<th>PEEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-0.292 [-0.335--0.250]</td>
<td>0.028 [0.019–0.038]</td>
<td>28.147 [26.785–29.510]</td>
<td>0.164 [0.117–0.210]</td>
<td>-0.093 [-0.147–0.042]</td>
<td>0.066 [0.047–0.085]</td>
</tr>
<tr>
<td>S2</td>
<td>-0.177 [-0.231--0.122]</td>
<td>0.146 [0.134–0.158]</td>
<td>39.485 [37.742–41.227]</td>
<td>0.091 [0.031–0.151]</td>
<td>0.218 [0.153–0.284]</td>
<td>0.428 [0.404–0.453]</td>
</tr>
<tr>
<td>S3</td>
<td>-0.500 [-0.589--0.411]</td>
<td>0.251 [0.231–0.271]</td>
<td>60.292 [57.430–63.155]</td>
<td>-0.728 [-0.826–-0.630]</td>
<td>-0.628 [-0.734–-0.521]</td>
<td>0.112 [0.071–0.152]</td>
</tr>
<tr>
<td>S4</td>
<td>-0.219 [-0.276--0.162]</td>
<td>0.144 [0.131–0.157]</td>
<td>41.749 [39.903–43.594]</td>
<td>0.041 [-0.022–0.105]</td>
<td>0.103 [0.034–0.172]</td>
<td>0.396 [0.370–0.422]</td>
</tr>
</tbody>
</table>

**Table 3.**

Fixed effects regression coefficients of ventilation parameters vs sensors. Values indicate fixed effect regression coefficient with the 95% confidence interval between brackets.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Minute Volume</th>
<th>etCO2</th>
<th>Tidal Volume</th>
<th>Inspiratory pressure</th>
<th>Peak pressure</th>
<th>PEEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.000 [-0.016--0.015]</td>
<td>0.000 [-0.003–0.004]</td>
<td>0.014 [-0.494–0.522]</td>
<td>-0.001 [0.018–0.016]</td>
<td>0.000 [-0.019–0.018]</td>
<td>0.000 [-0.007–0.007]</td>
</tr>
<tr>
<td>S2</td>
<td>-0.001 [-0.016--0.015]</td>
<td>0.000 [-0.003–0.004]</td>
<td>0.023 [-0.485–0.532]</td>
<td>-0.001 [0.018–0.016]</td>
<td>0.000 [-0.019–0.018]</td>
<td>0.000 [-0.007–0.007]</td>
</tr>
<tr>
<td>S3</td>
<td>-0.001 [-0.016--0.015]</td>
<td>0.000 [-0.003–0.004]</td>
<td>0.014 [-0.494–0.523]</td>
<td>-0.001 [0.018–0.017]</td>
<td>0.000 [-0.019–0.018]</td>
<td>0.000 [-0.007–0.007]</td>
</tr>
<tr>
<td>S4</td>
<td>-0.001 [-0.016--0.015]</td>
<td>0.000 [-0.003–0.004]</td>
<td>0.023 [-0.485–0.532]</td>
<td>-0.001 [0.018–0.016]</td>
<td>0.000 [-0.019–0.018]</td>
<td>0.000 [-0.007–0.007]</td>
</tr>
</tbody>
</table>

**Table 4.**

Fixed effects regression coefficients of ventilation parameters vs sensors. Values indicate fixed effect regression coefficient with the 95% confidence interval between brackets.
Non-invasive breath monitoring with eNose does not improve glucose diagnostics in critically ill patients in comparison to Continuous Glucose Monitoring in blood —

Non-invasive breath monitoring with eNose does not improve glucose diagnostics in critically ill patients in comparison to Continuous Glucose Monitoring in blood —


A shorter version of this manuscript appeared in Journal of Breath Research 2017, 2
ABSTRACT

Introduction: Continuous glucose monitoring (CGM) can be beneficiary in critically ill patients. Current CGM devices rely on subcutaneous or blood plasma glucose measurements and consequently there is an increased risk of infections and the possibility of loss of blood with each measurement. A potential method to continuously and non-invasively measure blood glucose levels is using exhaled breath. A correlation between blood glucose levels and volatile organic compounds (VOCs) in the exhaled breath was already reported. VOCs can be analyzed continuously using a so-called electronic nose (eNose). We hypothesize that continuous exhaled breath analysis using an eNose can be used to accurately predict blood glucose levels in intubated, mechanically ventilated ICU-patients.

Methods: Mechanically ventilated patients whose blood glucose concentration was monitored with a CGM device were eligible. An eNose with 4 metal oxide sensors was used to continuously measure changes in exhaled breath. After pre-processing the data, several regression models were trained, consisting of: (1) only eNose sensor values. (2) only the 1st and 2nd principal components (PC) of eNose values. (3) eNose sensor values and last known blood glucose value as random effect. (4) 1st and 2nd PC of eNose sensor values and CGM value of one minute ago as fixed effect. (5) CGM value of one minute ago as fixed effect. Model performance was measured using the R2 value, the Akaike Information Criterion (AIC) and the Clarke error grid.

Results: 23 patients were included in the study and 1165 hours of measurements were collected. Performance was low in model 1, 2 and 3 with a mean R2 of 0.07 [95%-CI: 0.00 – 0.28], 0.10 [95%-CI: 0.00 – 0.40] and 0.30 [0.02 – 0.79], respectively. Performance in models 4 and 5 was better with a mean R2 of 0.77 [0.02 – 1.00]. Subsequently, eNose data in model 4 had no added value over using CGM only in model 5.

Conclusion: In our study, continuous exhaled breath analysis using an eNose cannot be used to accurately predict blood glucose levels in intubated, mechanically ventilated ICU-patients.

Keywords: eNose, breath, analysis, glucose, critically ill, continuous

INTRODUCTION

Dysglycemia, and glycemic variability are associated with an increased risk of morbidity and mortality in critically ill patients [1-3]. While there is no consensus whether tight glucose control is beneficiary for patients on the intensive care unit (ICU) [4-6], it has been suggested that continuous glucose monitoring (CGM) will improve glucose control. The CGM devices that have been developed so far rely on subcutaneous or blood plasma glucose measurements and consequently there is an increased risk of infections, the possi-
bility of loss of blood with each measurement and discomfort for the patient. Additionally the accuracy for the prediction of blood glucose greatly varies between devices [7]. A potential method to continuously and non-invasively measure blood glucose levels is using exhaled breath [8]. Since the majority of ICU patient is ventilated, collection of exhaled breath would not be invasive for these patients.

**Exhaled breath analysis**
The scent of acetone in the breath of patients with diabetic keto-acidosis is a famous example of breath analysis and supports the theory that blood glucose levels may be measured via analysis of exhaled volatile organic compounds (VOCs) [9]. Recently, exhaled breath had been investigated as a diagnostic tool in various medical domains [10] using several technologies [11]. Several researchers have found a correlation between blood glucose levels and concentrations of different VOCs in the exhaled breath of healthy volunteers and diabetic patients [8]. However, there are currently no reports on continuous exhaled breath analysis for the monitoring of the plasma glucose concentration. Many studies use gas chromatography and mass spectroscopy (GC/MS) for breath analysis [12-14]. This technique takes a detailed snapshot of VOCs in breath but is not suitable for use in clinical practice as it is relatively expensive, not available at the bedside and time-consuming. VOCs can also be analyzed continuously and at the bedside, using a so-called electronic nose (eNose) [15]. eNoses rely on a cross-reactive sensor array to detect changes in VOC-profiles. These breath fingerprints can subsequently be used to train an algorithm that aims at predicting the plasma glucose level. We hypothesized that continuous exhaled breath analysis using an eNose could be used to accurately predict blood glucose levels in intubated, mechanically ventilated ICU-patients.

**METHODS**
Mechanically ventilated patients included in one of two other studies conducted in the ICU of the Academic Medical Center, Amsterdam, the Netherlands were eligible [16,17]. These studies investigated CGM devices and therefore plasma glucose concentration was monitored continuously. Patient age had to be > 18 years and anticipated life expectancy had to be > 96 hours. When patients were not expected to survive for a considerable amount of time, they were excluded. The Institutional Review Board of the Academic Medical Center approved the two aforementioned studies but concluded that the legislation on human participation in research was not applicable on this study because of its non-invasive nature and the need for informed consent was waived. The study was conducted in the 34 bed mixed medical/surgical ICU of the Academic Medical Center in Amsterdam, the Netherlands.
Standard of care for ventilated patients
Intubated and mechanically ventilated patients on the ICU of the Academic Medical Center in Amsterdam were ventilated according to a standardized protocol. This protocol included, but was not limited to: tidal volume between 6 and 8 mg/kg predicted body weight and Positive End Expiratory Pressure (PEEP) ≥ 5 cm H2O. In addition, a pulmonary toilet and nebulization of salbutamol and acetylcysteine were performed every 6 hours. Patients were routinely ventilated with passive humidification with a heat-moist exchanger. Blood gas analysis with the measurement of glucose were performed three times per day at minimum, or more frequently when deemed necessary.

Study design
Exhaled breath measurements initiated when CGM measurements were started in a mechanically ventilated patient. Breath was monitored continuously for the entire duration of CGM measurements and mechanical ventilation. Continuous breath measurements were halted when either CGM measurements or mechanical ventilation was stopped.

Continuous glucose measurements
Glucose was measured using one of two different CGM devices and was stored every minute. Sentrino® (Medtronic MiniMed, Northridge, CA) is a CGM that measures interstitial glucose levels using a subcutaneous sensor. EIRUS® (Maquet Critical Care AB, Solna, Sweden) used a specialized central venous catheter to measure blood glucose using microdialysis. CGM measurements were saved every minute on the measurements device and exported when measurements were done. Both devices were used on the ICU as part of clinical research into their accuracy.

Exhaled breath data collection
Breath was monitored using an eNose that was specifically adapted for continuous clinical use in the ICU (Comon Invent, Delft, the Netherlands). A similar device has been described in earlier studies [18]. The eNose consisted of one sensor array with 4 metal oxide sensors (Figaro, Osaka, Japan), chosen for their stability and safety, a roller pump for a continuous flow of exhaled breath, a plastic body that can easily be cleaned and an offline mode to prevent interference with other medical devices at bedside. Using a T-piece, the eNose was connected to the expiratory tube of the ventilator. Metal oxide and humidity sensor data were stored every minute.
**Pre-processing**
Data were pre-processed to diminish noise introduced by normal patient care. These sources of noise include time delay of sensors to reach steady state, changes in humidity in the side-stream connector, disconnections and nebulization of medication and changing ventilator settings. We have discussed our approach and considerations in a previous study [19].

**Model development**
Using the pre-processed data, several linear mixed effects models were trained, these models are used for the analysis of continuous dependent and independent variables with a grouping variable that serves as random intercept. In this way a linear regression coefficient is obtained for all independent variables, while correcting for correlated data within a patient. We developed the following models:

1. A linear mixed model with patient ID as random intercept and the four different sensors as fixed effects. This model only used pre-processed eNose data and is the simplest: Predicted Blood Glucose Level ~ eNose sensors + Patient ID as intercept.
2. A model similar to model (1), but using principal components obtained from the four metal oxide sensors: Predicted Blood Glucose Level ~ First two principle components of eNose sensors + Patient ID as intercept.
3. A model similar to model (2) with the last measured blood glucose value as random intercept. As blood glucose was measured at least three times a day, the last measurement was taken between 0 and 8 hours before the predicted value. By using the last known blood glucose value as random effect, this model essentially recalibrated when a new blood glucose value was available. : Predicted Blood Glucose Level ~ First two principle components of eNose sensors + Patient ID and last known blood glucose value as intercept.
4. A model similar to model (2) with the CGM value of one minute ago as fixed effect. This method could possibly be used to predict the trend of the signal using the eNose: Predicted Blood Glucose Level ~ First two principle components of eNose sensors + CGM value of 1 minute ago + Patient ID as intercept.
5. A model without any eNose sensor data, with the CGM value of 1 minute ago as fixed effect. This model can be used to compare the added value of eNose measurements Predicted Blood Glucose Level ~ CGM value of 1 minute ago + Patient ID as intercept.

**Model performance**
Performance of the developed models was measured using the coefficient of
determination R2 and the Akaike Information Criterion (AIC). The R2 ranges between 0 (model explains no variance) and 1 (model explains all variance). The AIC penalizes the likelihood of the model based on model complexity. Therefore, the higher the number of variables and thus complexity, the higher the penalty. In this way, model complexity and likelihood are balanced and the model with the lowest AIC will tend to resist overfitting. The AIC value is useful for comparing models but in itself is not informative. Additionally, a Clarke Error Grid, often used in studies evaluating glucose measurement devices, was used to measure point accuracy [20]. There are 5 zones in this grid and >95% of the samples should be in zone A, with a maximum of 5% in zone B to classify the method as accurate [21].

RESULTS

eNose data and CGM data were collected in 23 patients between October of 2012 and July of 2015. A total of 1165 hours of paired CGM eNose measurements were collected with a median duration of 51 (Interquartile range [41 – 65]) hours per patient. Included patients had a median APACHE II–score of 22, a median SAPS II–score of 52 and a median length of stay on the ICU of 12 days. Twenty-one patients were medical admissions, while 2 were planned surgery admissions. Patient characteristics are described in Table 1.

Performance of the eNose as stand-alone test

In Models 1 and 2, in which only use eNose data as predictors, model performance (table 2) had a mean R2 of 0.07 [95%-CI: 0.00 – 0.28] and 0.10 [95%-CI: 0.00 – 0.40] and an AIC of 317503 and 317529, respectively. 73.2% of values in zone A of the Clarke error grid in both models. Model predictions were plotted in Figure 1 for 1 patient. Model coefficients can be found in table 3.

Performance of the eNose in conjunction with previous glucose measurements

Performance of model 3, in which the last known blood glucose value was added as random effect, had a mean R2 of 0.30 [0.02 – 0.79] and an AIC of 288561. 78.9% of values were in zone A of the Clarke error grid.

Model 4, with the CGM value from one minute before as fixed effect, had a mean R2 of 0.77 [0.02 – 1.00] and an AIC of 277301. 85.7% of values in zone A of the Clarke error grid. Similar to model 4, but without any eNose data, model 5 performed similar with a mean R2 of 0.77 [0.02 – 1.00] and an AIC of 277320 and 85.7% of values in zone A of the Clarke error grid. While the addition of eNose variables in model 4 improved the model significantly (p = 1.386e-08) compared to model 5, the quantity of model performance improvement was negligible.

DISCUSSION

Performance of model 1 through 3 is low and as seen in model 5, good perfor-
mance of model 4 is completely due to inclusion of the last known blood glucose level. Therefore, our findings suggest that continuous exhaled breath analysis using an eNose cannot be used to accurately predict blood glucose levels in intubated, mechanically ventilated ICU-patients.

In contrast to our findings, several other studies were able to find a correlation between exhaled breath and blood glucose levels [8,22-27]. In a systematic review by our group, we found that 7 studies reported a correlation with a mean linear regression coefficient of 0.82 [range: 0.08–0.98] between one or more VOCs in exhaled breath, and blood glucose levels [8]. In these studies, an oral glucose tolerance test or clamp study design was often used, which resulted in a predictable trajectory of blood glucose levels. Results were positive in healthy and type 1 diabetes mellitus subjects, but no correlation could be found in type 2 diabetes mellitus patients [28,29]. Studies under less controlled circumstances, such as the present study, have not been reported on yet. In addition, none of these studies used an eNose for exhaled breath measurements. A study that used similar sensors as our eNose was able to distinguish healthy subjects from diabetic patients [30]. However, subject-specific prediction models were used to compensate for inter-subject variance, which makes the results not generalizable to clinical practice.

There are several possible explanations for the low performance of this eNose for continuous measurement of glucose levels in critically ill patients. Large inter-patient variation in critically ill ICU patients makes it harder to use a single predictive model. Metabolites linked to glucose levels, including acetone, ethanol, methanol and propane vary in quantity between healthy individuals with similar glucose levels [8]. In critically ill patients, differences in gut microbiome, also due to varying antimicrobial therapy, further differentiate levels of exhaled VOCs linked to glucose levels [31]. In addition, markers of oxidative stress that are found to be related to high blood glucose levels will be non-specific, as they will also increase with conditions common in the ICU population such as sepsis and acute respiratory distress syndrome [32].

In addition to inter-patient variation, the sensors in the eNose used in our study were possibly not specific enough. While the non-specific sensors in the eNose might be able to pick up differences in breath composition between healthy subjects and patients with a single medical condition such as asthma or lung cancer [18], it may not be possible to use the device in critically ill patients with multiple illnesses [33,34]. Each specific medical condition supposedly has a unique breathprint. With comorbidity, these breathprints may be overlaid within the breath of a single person, which makes it difficult, if not impossible, to infer the state of the disease of interest. This is a major challenge for the clinical application of any electronic nose type breath test, which has to be addressed more thoroughly in the future.
This study has several limitations. First, in order to get a reference standard with a high number of subsequent measurements, we opted to only include patients monitored by CGM. As our hospital only has access to CGM devices used in a study setting, and study patients were limited, we had little control over patient selection. Therefore, patients in our study were admitted for a wide variety of conditions, which may have increased inter-subject variation and therefore decreased the performance of the model. Second, we used an eNose that did not contain sensors that were specifically selected for sensitivity to VOCs of interest in glucose monitoring. Therefore, it is likely that the reported accuracy is an underestimation of the potential of a perfectly adapted eNose. However, performance is so low at the moment, that even with considerable improvements, the eNose would not be likely to accurately predict blood glucose levels in critically ill ICU patients. There were also several strengths to this study. It was the first study to continuously measure exhaled breath in critically ill patients using an electronic nose. In addition, a wide variety of data analysis methods were used for eNose model development, and several metrics were used to evaluate model performance. Finally, this is the first study to evaluate breath analysis for glucose monitoring in a clinical setting, and not in a clamp or glucose infusion study with predictable glucose dynamics. The results of this study have several implications. Further exhaled breath analysis research in critically ill patients should include concurrent measurements with GC–MS to better understand differences in breath profiles between patients. While the results of these measurements can in turn be used to develop an eNose with more specific sensors suitable for continuous glucose measurements in mechanically ventilated critically ill patients, it is unlikely that cross-reactive sensors used in eNoses are sensitive enough for continuous measurement of blood glucose in exhaled breath.

CONCLUSION

In our study, continuous exhaled breath analysis using an eNose cannot be used to accurately predict blood glucose levels in intubated, mechanically ventilated ICU-patients.

References


**FIGURES**

Figure 1. Example of the 5 different models in 1 patient. Predictions of the 5 different models. CGM device measurements are indicated by the black dashed line.

**TABLES**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, median [IQR]</td>
<td>67 [62 – 75]</td>
</tr>
<tr>
<td>Male gender, number (%)</td>
<td>9 (39%)</td>
</tr>
<tr>
<td>Admission type, number (%)</td>
<td></td>
</tr>
<tr>
<td>– Medical</td>
<td>21 (91%)</td>
</tr>
<tr>
<td>– Planned Surgery</td>
<td>2 (9%)</td>
</tr>
<tr>
<td>APACHE II, median [IQR]</td>
<td>22 [19 – 28]</td>
</tr>
<tr>
<td>SAPS II, median [IQR]</td>
<td>52 [42 – 65]</td>
</tr>
<tr>
<td>ICU LOS, days, median [IQR]</td>
<td>12 [9 – 15]</td>
</tr>
<tr>
<td>ICU mortality, number (%)</td>
<td>10 (43%)</td>
</tr>
<tr>
<td>Measurement duration, median [IQR]</td>
<td>51 [41 – 65]</td>
</tr>
</tbody>
</table>

Table 1. Patient characteristics. IQR, Interquartile range; BMI, body mass index; APACHE II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Sepsis-related Organ Failure Assessment score II; ICU, Intensive Care Unit; LOS, Length of stay.
<table>
<thead>
<tr>
<th>Model</th>
<th>R² (mean [Range])</th>
<th>AIC</th>
<th>CEG</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Pre-processed eNose values.</td>
<td>0.07 [0.00 – 0.28]</td>
<td>317503</td>
<td>73.2%</td>
<td>25.1%</td>
<td>0.0%</td>
<td>1.7%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>2: 1st and 2nd PC of eNose values.</td>
<td>0.10 [0.00 – 0.40]</td>
<td>317529</td>
<td>73.2%</td>
<td>25.1%</td>
<td>0.0%</td>
<td>1.7%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>3: Pre-processed eNose values and last known blood glucose value as random effect.</td>
<td>0.30 [0.02 – 0.79]</td>
<td>288561</td>
<td>78.9%</td>
<td>19.7%</td>
<td>0.0%</td>
<td>1.7%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>4: 1st and 2nd PC of eNose values and CGM value of one minute ago as fixed effect.</td>
<td>0.77 [0.02 – 1.00]</td>
<td>277301</td>
<td>85.7%</td>
<td>13.0%</td>
<td>0.0%</td>
<td>1.3%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>5: CGM value of one minute ago as fixed effect.</td>
<td>0.77 [0.02 – 1.00]</td>
<td>277320</td>
<td>85.7%</td>
<td>13.0%</td>
<td>0.0%</td>
<td>1.3%</td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. AIC, Akaike Information Criterium, CEG, Clarke error grid, PC, Principal Component, CGM, Continuous glucose measurement.
Part 3
Chapter 8

Comparison of classification methods in breath analysis by electronic nose


Journal of Breath Research 2015, 4
ABSTRACT
Currently, many different methods are being used for pre-processing, statistical analysis and validation of data obtained by electronic nose technology from exhaled air. These various methods, however, have never been thoroughly compared. We aimed to empirically evaluate and compare the influence of different dimension reduction, classification and validation methods found in published studies on the diagnostic performance in several datasets. Our objective was to facilitate the selection of appropriate statistical methods and to support reviewers in this research area.

We reviewed the literature by searching Pubmed up to the end of 2014 for all human studies using an electronic nose and methodological quality was assessed using the QUADAS-2 tool tailored to our review. Forty-six studies were evaluated regarding the range of different approaches to dimension reduction, classification and validation. From forty-six reviewed articles only seven applied external validation in an independent dataset, mostly with a case-control design. We asked their authors to share the original datasets with us. Four of the seven datasets were available for re-analysis. Published statistical methods for eNose signal analysis found in the literature review were applied to the training set of each dataset. The performance (area under the receiver operating characteristics curve (ROC-AUC)) was calculated for the training cohort (in-set) and after internal validation (leave-one-out cross validation). The methods were also applied to the external validation set to assess the external validity of the performance. Risk of bias was high in most studies due to non-random selection of patients. Internal validation resulted in a decrease in ROC-AUCs compared to in-set performance: -0.15, -0.14, -0.1, -0.11 in dataset 1 through 4, respectively. External validation resulted in lower ROC-AUC compared to internal validation in dataset 1 (-0.23) and 3 (-0.09). ROC-AUCs did not decrease in dataset 2 (+0.07) and 4 (+0.04). No single combination of dimension reduction and classification methods gave consistent results between internal and external validation sets in this sample of four datasets. This empirical evaluation showed that it is not meaningful to estimate the diagnostic performance on a training set alone, even after internal validation. Therefore, we recommend the inclusion of an external validation set in all future eNose projects in medicine.

INTRODUCTION
An electronic nose (eNose) is a chemical vapor analyzer, containing an array of cross-reactive sensors [1,2]. Metal-oxides, polymers, cantilevers and optical arrays are among the techniques that have been utilized for breath analysis. When exposed to a gas mixture, the physical properties of the sensors change. Those changes are measured and quantified resulting in a unique, composite
breath signal. Importantly, sensors in an eNose are typically semi-selective and promiscuous; they do not react with just one (group of) volatile organic compound(s) (VOCs). Therefore, breath data analysis largely relies on differential signaling of multiple sensors capturing spectra of partially overlapping VOCs. The sensor signals are subsequently analyzed by pattern recognition, after which the eNose signals form what is referred to as a “breathprint”. These breathprints are representing patterns of VOC mixtures and can be used for real-time probabilistic diagnostic and monitoring purposes[3]. In principle, such probabilistic outcome does not require identification of individual molecular constituents. Therefore, an eNose potentially represents a non-invasive, relatively easy to use diagnostic instrument, which makes it very attractive from a clinician’s as well patient’s point of view [4-8].

Several steps have to be taken to get from “raw” sensor data to validated classification models. This process includes: data pre-processing, dimensionality reduction, classification, and validation of classification models [9]. Research pertaining to the eNose has not been consolidated yet on these aspects, which is reflected by its explorative nature. Choosing the preprocessing, dimension reduction, classification, and validation techniques may have large impact on model performance. This is, however, not paralleled with extensive expertise of the merits of the techniques used, thereby resulting in arbitrary technique choices and suboptimal results. If one method were to be superior, standardization could be pursued [10].

Overfitting and validation represent two major issues concerning eNose analysis. Overfitting refers to the phenomenon when the model becomes so complex as to describe the idiosyncrasies of training data, reflecting noise, instead of the true underlying relationships in the data [11,12]. The breath research community has also identified the problems with false discoveries due to overfitting in recent years [13-15] but thus far there has been no objective evaluation of the influence of data analysis methods on diagnostic performance. This is particularly relevant as eNose signals may impose additional difficulties because sensor variables are highly correlated, necessitating controlling model complexity in order to avoid overfitting. Second, the majority of studies has not performed validation due to a case control design, and if so, this is mostly confined to internal validation. External validation, in which a model is evaluated on new independent samples from a setting different from the one used for model generation, should be applied to assess whether overfitting is manifest.

The aim of this paper was two-fold: (1) to strengthen eNose research by metho-
dologically evaluating eNose literature and (2) to examine how methodological choices affect external model validity. Therefore, we empirically evaluated and compared the influence of different dimension reduction, classification and validation methods found in the literature on diagnostic performance of eNose analysis approaches in published studies encompassing an external validation cohort. To this end we obtained datasets from authors of papers that included an external validation of their model. Our data may assist researchers in selecting appropriate methods and support reviewers to critically assess papers in this field.

**METHODS**

**Systematic review**

To get an overview of dimension reduction, classification and validation methods currently used, we searched Pubmed for available literature up to the end of 2014. Search terms used were “((((Electronic and Nose)) OR Sensor OR nano* OR nanosens* OR nanotub* OR nanopart* OR smell) AND breath) AND (diagnosis OR discrimin* OR classifi* OR diagnosing)”. No limits were imposed on publication year or language. Only studies in humans were included. Two independent reviewers (JHL, LDB) selected articles for full text assessment when the abstract suggested the work investigated exhaled breath using an electronic nose of nano-sensing technology to diagnose a certain medical condition. Disagreement between the two reviewers on inclusion of studies was resolved by consensus among them.

To assess the methodological quality of the included papers, the two reviewers (JHL, LDB) used an adapted version of the QUADAS-2 tool [16]. The adapted version of this tool can be found in eSupplement 1. Extracted data include: characteristics of the study (first author, year, country, condition studied, number of included patients), methods used (pre-processing, dimension reduction, feature selection, classification algorithm), performance measure(s), and validation method.

**Empirical evaluation**

We selected studies that attempted external validation and asked the authors to share the original datasets with us. We used the received datasets to evaluate the range of different approaches to dimension reduction, classification and validation.

Data processing was performed on training and validation data separately, that is, the training cohort was used exclusively to train the diagnostic algorithms. This separation ensures the inaccessibility of the external validation
set before all analyses on training data were performed. The dimensions of the datasets were reduced using one of three methods: principal component analysis (PCA), selection of sensors based on Akaike Information Criterion (AIC) [17] and PCA, and partial least square analysis – discriminant analysis (PLS-DA) in which information on the class variable is used. Alternatively, the raw values were maintained (figure 1, line 2). Classification was performed by one of five methods: neural networks (NN), logistic regression, linear discriminant analysis (LDA), support vector machine (SVM) or random forest (RF) (figure 1, line 3). Performance was quantified as the area under the receiver operating characteristics curve (ROC-AUC), without validation (in-set), with internal validation (leave-one-out cross-validation (LOOCV)) or after external validation (figure 1, line 4). Higher AUCs indicate better performance. For external validation, the models obtained in the training phase for both dimension reduction and classification were applied to the external validation set. Finally, we calculated the theoretical performance in the scenario that the external validation set would have been used as if it were the training cohort.

A recently introduced framework [18] was used to interpret differences in performance measures between training and validation sets by assessment of the reproducibility and transferability of the models. If similar model performance is observed in the internal and external validation set, it can be concluded that the model is reproducible provided that the training and validation samples are derived from the same population, whilst the model is transportable if they originate from a different population [19]. The difference in case-mix between the external validation and training sample was quantified by calculating the C-statistic (ROC-AUC) of a logistic regression model attempting to discriminate between these two cohorts (thus not between diseased and non-diseased subjects). A difference between the two sets can be the result of case mix heterogeneity or severity. Heterogeneity is reflected in the variation (SD) in predicted probability of group membership between the two samples. Case mix severity was measured by the difference in mean predicted group membership between the training and validation cohort [18].

RESULTS

Systematic review
The search identified 408 studies (figure 2) of which 344 were excluded due to being on different topics after reading title and abstract. Full text was assessed in 64 articles, 18 of which were excluded due to being on different topics. The 46 remaining articles were included in our analysis. Characteristics of these studies can be found in table 1.
Risk of bias
The QUADAS-2 tool measures bias in four domains: Patient selection, index test, reference test, and flow and timing. The risk of bias was high in most studies (table 1). This is primarily due to the non-random selection of patients in most studies. A correct index test (e.g. eNose) and reference standard (the diagnostic criteria for disease) was often used.

Pre-processing
Sensor signals have to be pre-processed before dimension reduction and classification can be performed. Despite the importance of this step, details on how data were pre-processed were not reported in 26 of the 46 (57%) studies included in our systematic review.

Dimension reduction
Five out of 46 (11%) studies used no dimension reduction technique (table 2). PCA was used in 30 (65%) and PLS in 1 (2%) studies. The remainder used other techniques (3/46; 7%) or it was not reported (7/46; 15%).

Classification methods
Discriminant analysis was the most frequently used classification method (22/46; 48%) (table 2). Four studies used logistic regression (9%), 3 (7%) used a neural network, 4 used SVM (9%) and 1 (2%) used a random forest. Eight out of 46 (17%) studies used other techniques and 4 (9%) did not report the technique that was used.

Validation
Ten out of 46 (22%) studies did not report on any validation of their results (that is, they only reported the in-set performance) (table 2). Internal validation was performed in 29 (63%) studies. External validation was performed in 7 (15%) studies.

Empirical evaluation
6 out of 7 contacted authors agreed to share their data sets with us [20-26]. Two out of six datasets were not included in our analysis due to issues preventing the reliable recreation of the original data (data was missing and/or the disease group which observations belonged to was not documented). The four remaining datasets were formatted in a similar way with cases as rows and sensors as columns [20,21,24,26]. These datasets encompassed pregnant women, acute respiratory distress syndrome patients, and asthma and chronic obstructive pulmonary disease patients. Datasets of three included studies used a Cyranose 320 as a measurement device yielding 32 sensor values per patient. One study
used an eNose developed by the University of Rome (Tor Vergata) with 8 sensor values per measurement.

**In-set performance**

In-set performance varied greatly between the four datasets and between tested combinations of dimension reduction and classification method (Figure 3). Dataset 1 [24] had a mean ROC-AUC of 0.98 [range 0.95 – 1]. Datasets 2[26], 3 [20] and 4 [21] had a lower ROC-AUCs of 0.86 [range 0.58 – 1], 0.83 [range 0.61 – 1] and 0.69 [range 0.5 – 0.94], respectively.

**Performance with internal validation**

Internal validation generally resulted in a lower ROC-AUC: -0.15 [range -0.10 – +0.00], -0.14 [range -0.35 – 0.00], -0.1 [range -0.30 – +0.11], -0.11 [range -0.34 – 0.00] in dataset 1 through 4, respectively (Figure 3). Mean ROC-AUCs after internal validation were 0.97 [range 0.91 – 1], 0.72 [range 0.48 – 0.87], 0.73 [range 0.43 – 0.91] and 0.59 [range 0.5 – 0.69] in datasets 1-4, respectively.

**Performance with external validation**

External validation resulted in a lower ROC-AUC in dataset 1 (-0.23 [range -0.46 – +0.13]) and 3 (-0.09 [range -0.34 – +0.22]) compared to the internal validation set. In contrast, the AUC-ROC did not decrease in dataset 2 (+0.07 [range -0.11 – +0.26] and 4 (+0.04 [range -0.13 – +0.15]) after external validation. Mean ROC-AUCs were 0.74 [range 0.5 – 0.86], 0.80 [range 0.5 – 0.99], 0.63 [range 0.46 – 0.84] and 0.63 [range 0.51 – 0.72] for dataset 1 through 4 respectively.

**Performance when the external validation set was used as training set**

External validation also resulted in a lower ROC-AUC compared to the theoretical maximum performance in dataset 1 (-0.18 [range -0.37 – +0.05]) and 3 (-0.16 [range -0.41 – +0.04]). Performance differed less between external validation and theoretical maximum performance in datasets 2 (-0.03 [range -0.41 – +0.24]) and 4 (-0.08 [range -0.41 – +0.08]). Mean ROC-AUCs were 0.86 [range 0.71 – 0.93], 0.83 [range 0.63 – 0.96], 0.80 [range 0.63 – 0.99] and 0.71 [range 0.57 – 0.92] for dataset 1 through 4, respectively.

**Best combination of dimension reduction and classification**

There was not a single combination of dimension reduction and classification that had the best performance in all four datasets (table 4). However, in datasets 1 through 3, the combination of PCA and LDA yielded the best results. The combination of methods that ranked at places 2 to 5 are mostly different between datasets. The use of a dimension reduction technique (either PCA, variable selection + PCA, or PLS-DA) provided better results than using raw values...
in almost every tested situation.

**Interpretation of external validation results**
Dataset 1 and 3 showed a decrease in ROC-AUC in the external validation cohort. The mean C-statistic between the training and validation set was 0.80 and 0.82, respectively (figure 4). Additionally, case mix heterogeneity decreased in these validation sets (figure 4A). With decreased variation, it is increasingly difficult to discriminate between cases and controls. Dataset 1 also showed an increase in case mix severity (figure 4B).

Dataset 2 and 4 showed no decrease in ROC-AUC between the external and internal validation cohort. The C-statistic between the training and validation set was 0.63 and 0.55, for dataset 2 and 4, respectively. The heterogeneity was similar in the training and validation cohorts in these datasets. In dataset 2 the case-mix severity was decreased in the external validation set.

**DISCUSSION**
This study evaluated and compared pre-processing, statistics and validation amongst published eNose studies of exhaled breath in medicine. Seven out of forty-six reviewed eNose studies performed external validation, and none attempted to interpret this validation. Risk of bias in the included studies was high in the majority of studies. Empirical evaluation of the data analysis methods showed that internal validation generally resulted in a lower performance, compared to inset performance. In two out of four datasets, performance further decreased after external validation. No single combination of dimension reduction and classification gave consistent results between internal and external validation sets in this sample of four datasets.

A systematic search in the literature provided evidence that a wide range of statistical methods is used in breath research with an eNose. The frequencies of these findings have been summarized in table 3. Furthermore, the risk of bias was high in these studies, mainly due to the non-random selection of patients. This is a direct consequence of the case-control design that is frequently used in pilot studies. In conjunction with the generally small sample size and wide range of diseases that was studied, the present study demonstrates that thus far breath research has been in an explorative phase. This may very well explain the limited number of studies that performed external validation, as this is frequently considered a second phase in diagnostic studies [27].

No single combination of dimension reduction and classification methods was superior in breath analysis with an electronic nose. In comparison, Gromski et al. previously reported that SVM and LDA were superior to Random Forests and PLS-DA in the statistical analysis of eNose signals [28]. One impor-
tant difference between our and the latter study is that we used patient data from four different studies that were published in peer-reviewed journals instead of data derived in the lab by exposing an eNose to selected volatile organic compounds (acetone, DMMP, methanol and 1-propanol). Therefore, the present results more closely represent the situation that will be encountered in future clinical settings or trials.

Although a “one size fits all” method cannot be recommended, we made the following observations about combinations of dimension reduction and classification. In three out of four tested sets, the combination of PCA and LDA gave the best results in the external validation set. More in depth investigation of these studies showed that all three studies used a case-control design and included patients with a limited number of comorbidities. As PCA is an unsupervised method for dimension reduction it performs well in situations where the largest variance is explained by the differences between cases and controls. However, this combination was not in the top 5 best performing combinations in dataset 4. Interestingly, the latter study had a prospective cohort design and included many patients with multiple comorbidities. This may have limited the possibility for PCA to capture the relevant differences between cases and controls. Indeed, the influence of comorbidities was previously suggested to limit the use of PCA and LDA in patients with pulmonary embolisms [29]. We postulate that in such cases supervised dimension reduction techniques such as PLS-DA might be preferable. Our data suggest that data analysis techniques should be tailored to the clinical question and included population.

The performance during internal validation was similar to the externally validated performance in dataset 2 and 4. Thus, model performance was highly reproducible in these datasets and the datasets were indistinguishable (C-statistic close to 0.5). Because the datasets are indistinguishable, one does not know how the model performance transports to a different dataset. Indeed, the article from which dataset 2 was derived, described that training and validation cohorts were drawn from one population. For dataset 4, the external validation cohort came from the same intensive care unit of the same hospital as the training set. In contrast, the C-statistic was approximately 0.8 between the training and validation cohorts in dataset 1 and 3, which allowed us to test transportability. However, the transportability was limited in these datasets as indicated by a decreased performance after external validation. This is not unexpected, since the samples in the validation cohort in dataset 1 were taken in other hospitals as the training cohort. Additionally, healthy controls served as the control group in the training cohort, whilst COPD patients with fixed airway obstruction were included as control patients in the validation study. In dataset 4, pregnant women in their third trimester served as cases in the training cohort but in the validation cohort also women in their second tri-
mester of pregnancy were included. Taken together, the framework proposed by Debray et al. can be used to explain the differences in model performance between the internal and external validation set.

Several limitations of our study should be noted. First, only four datasets were included in the empirical evaluation of the statistical methods. Although most authors we contacted were willing to share their data with us, not all datasets were usable due to issues that prevented the reliable recreation of the original data. Therefore, we were not able to apply all methods on more datasets, which would have strengthened our findings. This limits the generalizability of the results. Second, not all possible combinations of dimension reduction and classification techniques were investigated. However, since there are numerous possibilities we focused on the plausible ones based on a systematic review on all methods used for eNose analysis in the medical field. This assured that we investigated combinations of dimension reduction and classification techniques that were actually used previously for this purpose. It is however interesting to investigate other methods, such as non-linear variable selection techniques, in future studies. It must also be noted how the hyphenation of a dimension reduction technique like PLS-DA with a classification method like RF work: while the PLS-DA focuses on the variance in a linear fashion it might exclude more complex forms of information that methods such as RF would have captured otherwise. Our study also had several strengths. We systematically reviewed all the available literature on eNose analysis for the diagnosis of medical conditions. Furthermore, this is the first study to empirically evaluate the influence of different combinations of statistical methods in this field. Therefore, otherwise arbitrary methodological decisions can now be motivated, allowing explicit recommendations at the end of this manuscript. Finally, we were able to explain our findings using a novel framework, in conjunction with clinical data described in the original articles.

The presented data have several implications. Internal validation provided more realistic results than in-set analysis. Therefore, especially with a limited number of patients, internal validation has to be performed. However, it must be noted that cross-validation, as do other re-sampling validation methods, does not consider omission of sampling variance and in this sense comprises a sub-optimal form of test set validation[30]. Also, when looking at figure 3, it can also be observed that in any dataset a good internally valid result can be obtained if sufficient analytical methods are explored. Since statistical fishing expeditions by authors cannot be traced during peer-review, unless an a priori data analysis plan had been published beforehand, the results of any paper without an external validation set should be treated as potentially over-optimistic. Of note, this over-optimism is only one of possible other independent
biases such as publication bias, bias due to study design, and other biases that could positively influence diagnostic accuracy [31]. All of these items can only be assessed if the TRIPOD and STARD guidelines are used for reporting [32,33]. On the other hand, the use of a stringent data-analysis plan could result in false-negative results. For example, one could conclude from study 1, 2 and 3 that a combination of PCA with LDA provides reasonable results and this is in fact used frequently in the literature. However, this approach would result in very low diagnostic accuracy in study 4 (figure 3). Therefore, such an inflexible approach cannot be recommended for all studies. A possible solution may be the use of more flexible analytical plans that are part of the protocol that is preferably published in an online registry before the study is completed. This flexibility implies that external validation is even more important.

CONCLUSIONS
Published diagnostic studies with eNoses have used many types of dimension reduction and classification techniques. Only 7 out of 46 studies performed external validation. This empirical evaluation demonstrated that no single combination of techniques is superior to the others. Our data showed that it is not meaningful to estimate the diagnostic performance on a training cohort alone.

RECOMMENDATIONS
- No single statistical analytical method is superior in breath analysis by an electronic nose. Therefore, several analytical methods should be tested on the training set to obtain an optimal result.
- Internal validation is essential to estimate the performance of a statistical method in the training cohort. It should therefore always be performed.
- An external validation cohort extracted from a related population that is separated from the training cohort by either time or place, should always be included to estimate the true diagnostic performance. This is especially true if multiple analytical methods are tested.
- Differences between internal and external validation should be interpreted for example by using the framework described in Debray et al. following a similar approach as in the present paper [18].

ACKNOWLEDGEMENTS
We would like to thank Serpil Ezurum for sharing her dataset [25] with us.
References


52. Hanson, C. W. J.; Thaler, E. R. Electronic nose prediction of a clinical pneumonia score: biosensors and microbes. Anesthesiology 2005, 102, 63–68.


57. Lazan, Z.; Fens, N.; van der Maten, J.; van der Schee, M. P.; Wagener, A. H.; de Nijs, S. B.; Dijkers, E.; Sterk, P. J. Electronic nose breathprints are independent of acute changes in airway caliber in asthma. Sensors (Basel) 2010, 10, 9127–9138.


**FIGURES**

![Diagram](image-url)

**Figure 1.**

Methods subtracted from studies in our systematic review. Each of these steps was applied to the datasets supplied by the authors we contacted.
Figure 2. Flow diagram of study selection

Figure 3. Area under the receiver operator characteristic curve of the four included datasets for each combination of dimension reduction and classification. The first column shows the in-set performance, the second column shows internal-validation results (LOOCV), and the third column shows external validation results. The last column shows theoretical performance if the external validation set were to be used as if it were the training cohort and the accuracy was validated with LOOCV. Red indicates low AUC values, green indicates high AUC values.
Figure 4.
Framework by Debray et al applied to the four included datasets. The top graph a combination of the C-statistic (ROC-AUC) between the training and validation set, and the case mix heterogeneity. The lower graph is the combination of the C-statistic (ROC-AUC) between the training and validation set, and the case mix severity.
### TABLES

#### RISK OF BIAS

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Flow and timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benedek[34]</td>
<td>2013</td>
<td>Yes</td>
<td>Unknown</td>
<td>No</td>
<td>Unknown</td>
</tr>
<tr>
<td>Bikov[20]</td>
<td>2011</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bikov[35]</td>
<td>2011</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Unknown</td>
</tr>
<tr>
<td>Bos[21]</td>
<td>2014</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Broza[36]</td>
<td>2013</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bruins[22]</td>
<td>2013</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bruno[37]</td>
<td>2008</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Chapman[23]</td>
<td>2012</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cohen-Kaminsky[38]</td>
<td>2013</td>
<td>Unknown</td>
<td>Unknown</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>D'amico[39]</td>
<td>2010</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>de Heer[40]</td>
<td>2013</td>
<td>Unknown</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Di Natale[41]</td>
<td>2003</td>
<td>Unknown</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Dragoneiri[42]</td>
<td>2007</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Dragoneiri[43]</td>
<td>2009</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Dragoneiri[44]</td>
<td>2012</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Dragoneiri[45]</td>
<td>2013</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Fens[46]</td>
<td>2009</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Fens[24]</td>
<td>2011</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Fens[47]</td>
<td>2011</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Fens[48]</td>
<td>2013</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Greulich[49]</td>
<td>2013</td>
<td>Unknown</td>
<td>No</td>
<td>Unknown</td>
<td>No</td>
</tr>
<tr>
<td>Guo[50]</td>
<td>2010</td>
<td>Unknown</td>
<td>No</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Hakim[51]</td>
<td>2011</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hanson 3rd[52]</td>
<td>2005</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hattesoh[53]</td>
<td>2011</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hockstein[54]</td>
<td>2005</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hockstein[55]</td>
<td>2004</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kovacs[56]</td>
<td>2013</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Lazar[57]</td>
<td>2010</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Leunis[58]</td>
<td>2013</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Machado[25]</td>
<td>2005</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Marom[59]</td>
<td>2012</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mazzone[60]</td>
<td>2007</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mazzzone[61]</td>
<td>2012</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mohamed[62]</td>
<td>2003</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Montuschi[26]</td>
<td>2010</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Nonaka[63]</td>
<td>2005</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Pat</td>
<td>Diseases studied</td>
<td>Dimension reduction</td>
<td>Classification algorithm</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>-----</td>
<td>--------------------------------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Benedek[34]</td>
<td>2013</td>
<td>28</td>
<td>Other pulmonary disease</td>
<td>PCA</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Bikov[35]</td>
<td>2011</td>
<td>10</td>
<td>Exercise</td>
<td>PCA</td>
<td>None</td>
</tr>
<tr>
<td>Bos[21]</td>
<td>2014</td>
<td>48</td>
<td>Other pulmonary disease</td>
<td>Raw sensor values</td>
<td>(S)PLS</td>
</tr>
<tr>
<td>Broza[36]</td>
<td>2013</td>
<td>12</td>
<td>Pulmonary malignancy</td>
<td>Raw sensor values</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Bruins[22]</td>
<td>2013</td>
<td>30</td>
<td>Pulmonary infection</td>
<td>PARAFAC and TUCKER3 kernels</td>
<td>Neural network</td>
</tr>
<tr>
<td>Bruno[37]</td>
<td>2008</td>
<td>28</td>
<td>Rhinosinusitis</td>
<td>Raw sensor values</td>
<td>None</td>
</tr>
<tr>
<td>Chapman[23]</td>
<td>2012</td>
<td>80</td>
<td>Other malignancy</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Cohen-Kaminsky[38]</td>
<td>2013</td>
<td>45</td>
<td>Other pulmonary disease</td>
<td>Not reported</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>D’amico[39]</td>
<td>2010</td>
<td>92</td>
<td>Pulmonary malignancy</td>
<td>Not reported</td>
<td>(S)PLS</td>
</tr>
<tr>
<td>de Heer[40]</td>
<td>2013</td>
<td>11</td>
<td>Other pulmonary disease</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Di Natale[41]</td>
<td>2003</td>
<td>60</td>
<td>Pulmonary malignancy</td>
<td>Raw sensor values</td>
<td>(S)PLS</td>
</tr>
<tr>
<td>Dragoneiri[42]</td>
<td>2007</td>
<td>40</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Dragoneiri[43]</td>
<td>2009</td>
<td>30</td>
<td>Pulmonary malignancy, Asthma or COPD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Reference</td>
<td>Year</td>
<td>Sample Size</td>
<td>Disease/Condition</td>
<td>Methodology</td>
<td>Analysis Type</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Dragoneiri[44]</td>
<td>2012</td>
<td>39</td>
<td>Other malignancy</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Dragoneiri[45]</td>
<td>2013</td>
<td>56</td>
<td>Other pulmonary disease</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Fens[46]</td>
<td>2009</td>
<td>90</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Fens[24]</td>
<td>2011</td>
<td>100</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Fens[47]</td>
<td>2011</td>
<td>32</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Not reported</td>
</tr>
<tr>
<td>Fens[48]</td>
<td>2013</td>
<td>157</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Greulich[49]</td>
<td>2013</td>
<td>60</td>
<td>Sleep apnoea</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Guo[50]</td>
<td>2010</td>
<td>442</td>
<td>Other malignancy, Diabetes</td>
<td>PCA</td>
<td>K-Nearest neighbour</td>
</tr>
<tr>
<td>Hakim[51]</td>
<td>2011</td>
<td>87</td>
<td>Other malignancy</td>
<td>PCA</td>
<td>SVM</td>
</tr>
<tr>
<td>Hanson 3rd[52]</td>
<td>2005</td>
<td>38</td>
<td>Pulmonary infection</td>
<td>PLS</td>
<td>(S)PLS</td>
</tr>
<tr>
<td>Hattesohl[53]</td>
<td>2011</td>
<td>33</td>
<td>Asthma or COPD</td>
<td>Not reported</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Hockstein[54]</td>
<td>2005</td>
<td>44</td>
<td>Pulmonary infection</td>
<td>Not reported</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Hockstein[55]</td>
<td>2004</td>
<td>25</td>
<td>Pulmonary infection</td>
<td>PCA</td>
<td>SVM</td>
</tr>
<tr>
<td>Kovacs[56]</td>
<td>2013</td>
<td>49</td>
<td>Lung transplant</td>
<td>PCA</td>
<td>Mahalanobis distance</td>
</tr>
<tr>
<td>Lazar[57]</td>
<td>2010</td>
<td>10</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Not reported</td>
</tr>
<tr>
<td>Leunis[58]</td>
<td>2013</td>
<td>59</td>
<td>Other malignancy</td>
<td>Not reported</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Marom[59]</td>
<td>2012</td>
<td>62</td>
<td>Kidney failure</td>
<td>PCA</td>
<td>SVM</td>
</tr>
<tr>
<td>Mazzone[60]</td>
<td>2007</td>
<td>143</td>
<td>Pulmonary malignancy</td>
<td>Not reported</td>
<td>Random Forest</td>
</tr>
<tr>
<td>Mazzone[61]</td>
<td>2012</td>
<td>229</td>
<td>Pulmonary malignancy</td>
<td>Raw sensor values</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Mohamed[62]</td>
<td>2003</td>
<td>10</td>
<td>Rhinosinusitis</td>
<td>PCA</td>
<td>Neural network</td>
</tr>
<tr>
<td>Montuschi[26]</td>
<td>2010</td>
<td>51</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Neural network</td>
</tr>
<tr>
<td>Nonaka[63]</td>
<td>2005</td>
<td>66</td>
<td>Malodor</td>
<td>Not reported</td>
<td>Logistic regression, Multiple linear regression</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Sample Size</td>
<td>Disease/Condition</td>
<td>Analysis Method</td>
<td>Reporting Location</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>-------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Paff[64]</td>
<td>2013</td>
<td>73</td>
<td>Other pulmonary disease</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Peng[65]</td>
<td>2009</td>
<td>62</td>
<td>Pulmonary malignancy</td>
<td>PCA</td>
<td>Not reported</td>
</tr>
<tr>
<td>Peng[66]</td>
<td>2010</td>
<td>177</td>
<td>Pulmonary malignancy, Other malignancy</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Timms[67]</td>
<td>2012</td>
<td>81</td>
<td>GORD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Tisch[68]</td>
<td>2012</td>
<td>19</td>
<td>Asymptomatic Nigrostriatal Dopaminergic Lesion</td>
<td>DFA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>van der Scheel[69]</td>
<td>2013</td>
<td>45</td>
<td>Asthma or COPD</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Witt[70]</td>
<td>2011</td>
<td>22</td>
<td>Discrimination smokers/ non smokers</td>
<td>PCA</td>
<td>Discriminant analysis</td>
</tr>
<tr>
<td>Wlodzimirow[71]</td>
<td>2013</td>
<td>14</td>
<td>Hepatic failure</td>
<td>PCA</td>
<td>K-Nearest Neighbour</td>
</tr>
<tr>
<td>Xu[72]</td>
<td>2013</td>
<td>130</td>
<td>Other malignancy</td>
<td>DFA</td>
<td>Discriminant analysis</td>
</tr>
</tbody>
</table>

Table 2.
Study characteristics extracted from the studies included in the systematic review part of this paper. PCA: Principal Component Analysis, (S)PLS: (Sparse) Partial Least Squares, COPD: Chronic Obstructive Pulmonary Disease, SVM: Support Vector Machine, GORD: Gastroesophageal Reflux Disease, DFA: Discriminant Factor Analysis.
<table>
<thead>
<tr>
<th>Risk of bias: Patient selection</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Yes</td>
<td>24 (52%)</td>
</tr>
<tr>
<td>- No</td>
<td>13 (28%)</td>
</tr>
<tr>
<td>- Unknown</td>
<td>9 (20%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risk of bias: Index test</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Yes</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>- No</td>
<td>40 (87%)</td>
</tr>
<tr>
<td>- Unknown</td>
<td>3 (7%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risk of Bias: Reference Standard</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Yes</td>
<td>8 (17%)</td>
</tr>
<tr>
<td>- No</td>
<td>34 (74%)</td>
</tr>
<tr>
<td>- Unknown</td>
<td>4 (9%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risk of Bias: Flow and Timing</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Yes</td>
<td>8 (17%)</td>
</tr>
<tr>
<td>- No</td>
<td>34 (74%)</td>
</tr>
<tr>
<td>- Unknown</td>
<td>4 (9%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension Reduction method:</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- PCA</td>
<td>30 (65%)</td>
</tr>
<tr>
<td>- RAW</td>
<td>5 (11%)</td>
</tr>
<tr>
<td>- PLS</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>- DFA</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>- Other</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>- Not reported</td>
<td>7 (15%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification Method:</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- LR</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>- Mahalanobis distance</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>- (S)PLS</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>- DA</td>
<td>22 (48%)</td>
</tr>
<tr>
<td>- NN</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>- RF</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>- KNN</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>- SVM</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>- none</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>- not reported</td>
<td>3 (7%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation Method:</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>- In-set</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>- Internal</td>
<td>29 (63%)</td>
</tr>
<tr>
<td>- External</td>
<td>7 (15%)</td>
</tr>
<tr>
<td>- Not reported</td>
<td>8 (17%)</td>
</tr>
</tbody>
</table>

*Table 3.* Summary of frequencies of table 1 and 2
<table>
<thead>
<tr>
<th>Set</th>
<th>Dimension Reduction</th>
<th>Classification</th>
<th>Combined (Top 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1. Variable selection + PCA</td>
<td>1. RF</td>
<td>1. PCA + LDA</td>
</tr>
<tr>
<td></td>
<td>2. PCA</td>
<td>2. LDA</td>
<td>2. Variable selection + PCA + LDA</td>
</tr>
<tr>
<td></td>
<td>3. Raw</td>
<td>3. SVM</td>
<td>3. Raw values + RF</td>
</tr>
<tr>
<td></td>
<td>5. Neural Network</td>
<td></td>
<td>5. Variable selection + PCA + RF</td>
</tr>
<tr>
<td>2</td>
<td>1. PCA</td>
<td>1. LDA</td>
<td>1. PCA + LDA</td>
</tr>
<tr>
<td></td>
<td>2. Variable selection + PCA</td>
<td>2. RF</td>
<td>2. Variable selection + PCA + LDA</td>
</tr>
<tr>
<td></td>
<td>3. Raw values</td>
<td>3. SVM</td>
<td>3. PCA + RF</td>
</tr>
<tr>
<td></td>
<td>4. PLS-DA</td>
<td>4. NN</td>
<td>4. Variable selection + PCA + RF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Logistic Regression</td>
<td>5. Raw values + RF</td>
</tr>
<tr>
<td>3</td>
<td>1. PCA</td>
<td>1. LDA</td>
<td>1. PCA + LDA</td>
</tr>
<tr>
<td></td>
<td>2. Raw values</td>
<td>2. NN</td>
<td>2. PCA + Logistic Regression</td>
</tr>
<tr>
<td></td>
<td>3. PLS-DA</td>
<td>3. Logistic Regression</td>
<td>3. PCA + RF</td>
</tr>
<tr>
<td></td>
<td>4. Variable selection + PCA</td>
<td>4. SVM</td>
<td>4. PCA + NN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. RF</td>
<td>5. PCA + SVM</td>
</tr>
<tr>
<td>4</td>
<td>1. PLS-DA</td>
<td>1. SVM</td>
<td>1. PLS-DA + Logistic Regression</td>
</tr>
<tr>
<td></td>
<td>2. PCA</td>
<td>2. LDA</td>
<td>2. PLS-DA + LDA</td>
</tr>
<tr>
<td></td>
<td>3. Variable selection + PCA</td>
<td>3. RF</td>
<td>3. PLS-DA + SVM</td>
</tr>
<tr>
<td></td>
<td>4. Raw values</td>
<td>4. NN</td>
<td>4. PLS-DA + NN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Logistic Regression</td>
<td>5. PCA + SVM</td>
</tr>
</tbody>
</table>

Table 4. Ranking data-reduction and classification methods and their combinations
Volatile organic compound profiles in outlet air from extracorporeal life-support devices differ from breath profiles in critically ill patients

ABSTRACT

**Introduction:** It is highly uncertain where volatile organic compounds (VOCs) in exhaled breath originate. Indeed, VOCs could be produced locally, in the air compartment or lung tissue, but they may also originate from the bloodstream. We compared VOC mixtures in exhaled breath, and in air coming from extracorporeal support devices in critically ill patients.

**Methods:** First, we investigated whether it was safe to connect an eNose or a gas sampling pump to extracorporeal support membranes. Then, breath and air from extracorporeal support devices were collected simultaneously for continuous monitoring of VOC mixtures using an electronic nose (eNose). In addition, samples for gas–chromatography and mass spectrometric detection (GC-MS) analysis were taken daily on the two sites.

**Results:** Ten critically ill patients were monitored for 73 [IQR –113] hours; in total, we had 887 hours of air sampling. The eNose signals of breath correlated moderately with signals of air from the extracorporeal support devices (R² between 0.25 and 0.44). After GC-MS analysis, ninety-six VOCs were found both in breath and air from the extracorporeal support devices; of these, 29 (30%) significantly correlated (P < 0.05) between the two sites, of which 17 were identified. VOCs that did not correlate were found in a higher concentration in breath than in air from the extracorporeal support devices.

**Conclusion:** This study suggests VOC analysis in the extracorporeal circulation is safe and that VOCs of non–pulmonary origin can be measured in the breath and in the extracorporeal circulation of critically ill patients. For VOCs that did not correlate between the two collection sites, the breath concentration was higher, suggesting pulmonary production of these molecules in a highly selected population of patients received extracorporeal support.

INTRODUCTION

Intensive Care medicine requires intensive monitoring of numerous biological markers indicative for organ function, infection status or metabolic needs. Often, these markers are blood based and can only be analyzed intermittently. Preferably, biological data are monitored continuously and as noninvasively as possible in critically ill patients [1].

Exhaled breath contains hundreds of volatile organic compounds (VOCs) [2]. Analysis of VOCs has been suggested to provide information about numerous pathophysiological processes and might improve the diagnosis and monitoring of several diseases [3-6]. VOCs can be separated, quantified and identified by gas chromatography and mass spectrometry (GC-MS). Alternatively, patterns of VOCs can be analyzed by sensor arrays called electronic noses (eNoses). VOCs result from oxidative stress [3], inflammation [7] or come from bacterial metabolism [7-9] and can be of systemic origin or can be produced in the lung itself.
Studies in healthy volunteers have established that certain VOCs, such as acetone, are transported to the lung via the bloodstream [10]. Others, such as propofol, are infused by the physician and we therefore know the biological origin of propofol in exhaled breath [11]. However, the origin of many VOCs that are encountered in severe illness is still unknown.

Patients with severe respiratory or circulatory problems who are admitted to the intensive care unit may receive extracorporeal life support to support with oxygenation (extracorporeal membrane oxygenation [ECMO]) or CO2 removal (extracorporeal CO2 removal [ECCO2R]). The application of these extracorporeal support devices allows for the comparison of (profiles of) VOCs coming from the external air outlet of the membrane with those in exhaled air.

We hypothesize that there is a correlation between VOCs in exhaled breath and VOCs coming from the membrane of an extracorporeal support device for VOCs that are known to be of systemic origin in critically ill patients requiring extracorporeal support. If so, we will use this technique to identify those compounds that are likely to have a non-pulmonary origin in critical illness. The aim of this study is: (1) assess the feasibility and safety of connecting an electronic nose to an extracorporeal circulation device and (2) to evaluate the correlation between VOCs in exhaled breath and VOCs coming from the membrane of an extracorporeal support device.

METHODS

Design

Feasibility and safety, and clinical measurements were all performed at the university hospital of Regensburg, Germany. The Institutional Review Board of the university hospital of Regensburg waived the need for informed consent due to the non-invasive nature of the study.

VOC analysis

VOCs were measured in exhaled breath and in the extracorporeal circulation using two different techniques. First, air was monitored continuously using an eNose adapted for clinical use (Comon Invent, Delft, the Netherlands). This device with a sensor array with 4 metal oxide sensors (Figaro, Osaka, Japan) has been described in a previous study [12]. The eNose was connected to the membrane via an intermediate piece of tubing to the external outlet of the membrane. Samples for GC-MS were gathered with a gas sampling pump (Acti-VOC; Markes, Llantrisant, UK) from the same disposable side-stream connection for at 200 mL/min for 10 min and trapped on a sorbent tube (TD100, Markes, Llantrisant, UK) [13,14]. These tubes were analyzed by means of GC-MS as described previously [15]. In short, sorbent tubes were heated to 280°C for 15 min with a flow of 30 ml/min. VOCs were captured on a cold trap at 10°C and then
reinjected by rapidly heating the trap to 300 °C for 1 minute. Subsequently the molecules were splitless injected through a transfer line at 180 °C onto an Inertcap 5MS/Sil GC column [30 m, ID 0.25 mm, film thickness 1 µm, 1,4-bis(dimethylsiloxy)phenylenedimethyl polysiloxane. (Restek, Breda, The Netherlands)] with a flow of 1.2 ml/min. Oven temperature was isothermal at 40°C for 5 minutes, increased to 280 ° at 10°C/min, and kept isothermal at 280 °C for 5 minutes. Molecules were ionized using electron ionization (70 eV), and the fragment ions were detected using a quadrupole mass spectrometer (GCMS-GP2010, Shimadzu, Den Bosch, The Netherlands) with a scan range of 37–300 Da. GC-MS analysis, de-noising, peak detection, and alignment were performed using the R Xcms package (Scripps Center for Metabolomics, La Jolla, CA) and resulted in an ion fragment peak table as input for statistical analysis. Collinearity of ion fragment peaks were checked for every detected peak in the total ion chromatogram to exclude co-elution that would significantly increase the measured VOC concentration. The predictive VOCs were manually checked in the raw chromatograms and the corresponding metabolites were tentatively identified based on National Institute of Standards and Technology library matching. Metabolites were considered identified if the first five hits in the library were the same compound and all matching factors were 90%. In the case of multiple likely library hits, a chemical standard (Sigma-Aldrich, Zwijndrecht, The Netherlands) was injected for identification. When these two procedures did not result in identification, the compound was deemed unidentified.

**Extracorporeal circulation: Feasibility and safety**

Since the operation of extracorporeal support devices should in no way be disturbed, it had to be investigated whether it was safe to connect an eNose or a gas sampling pump to the device. Therefore, it was tested if the connector, a T-piece with the same diameter as the gas outlet, added airflow resistance that could possibly disturb the extracorporeal support device. Additionally, the connection had to be tested to make sure that no condensed water droplets entered into the eNose together with the airflow. The gas inlet of the membrane was connected to the hospital oxygen supply. Positioned in front of the inlet port with a three-way valve, a cuff blood pressure meter was placed. The adapter to which the eNose or gas sampling pump could be connected was placed on the gas outlet of the membrane. An eNose or gas sampling pump was connected to this adapter. Pre-oxygenator pressure data was recorded with and without the connector in threefold, at 3 different in-membrane pressures (60, 120 and 150 mmHg) using 15 different levels of sweep gas flow (1-15 L/min) in three-fold. These measurements were performed in the different membranes used in the hospital: Maquet Quadrox-iR and HLS Advanced 7.0, Novalung iLA...
Activve, Sorin ECC.O 5 and Medos HILITE AF7000. Mean difference in in-membrane pressure and the range of differences was calculated afterwards.

**Extracorporeal circulation: Measurements in patients**

After safety was guaranteed, measurements in patients were started. Inclusion criteria were: (1) Patients admitted to an Intensive Care Unit (ICU) of the university hospital in Regensburg, (2) patients received extracorporeal support and (3) were expected to receive extracorporeal support for at least 72 hours. There were no exclusion criteria. Due to the exploratory nature of the study, every available patient between September and December of 2015 was included and no power calculation was made. eNose measurements were taken during the entire duration of extracorporeal support. Every 24 hours an air sample was taken for GC-MS analysis. When a patient was monitored for 7 days and another eligible is available, the eNose was transferred to the new patient. Patient characteristics and oxygenator variables were collected.

**Breath measurements**

Breath was monitored through a side-stream connector placed distal of the heat-moist exchanger, as described before [13,16]. The electronic nose was connected to the side-stream connector with Teflon tubing. As with the extracorporeal circulation measurements, eNose measurements were taken during the entire duration of extracorporeal support. Breath samples for GC-MS analyses were taken simultaneously with those taken from the extracorporeal circulation.

**Pre-processing of the eNose signal**

Using available ventilator and oxygenator variables, eNose data were pre-processed to diminish the effect of changes in humidity in exhaled breath and air coming from the membrane. This approach has been discussed in detail before [16]. In short, the pre-processing included normalization for changes in humidity and smoothing of the signal.

**Data analysis**

After eNose data were pre-processed, each sensor was correlated to the corresponding sensor in both measurement sites for each patient using the lm() method in R 3.3.1 for linear regression modelling [17]. VOCs, measured by GC-MS, were correlated between the two measurement sites in a similar fashion. The abundance of the VOCs was also compared between the breath and the extracorporeal circulation (ECC) and a ratio was calculated (Breath/ECC ratio). The association between the correlation coefficient and the Breath/ECC ratio was visualized and was analyzed by linear regression.
RESULTS
Between September and December of 2015, 10 patients were included in the study. Median patient age was 56 [IQR 51–65] and all patients were critically ill. 6 out of 10 patients were diagnosed with ARDS. Patients were admitted to the ICU for a median duration of 36 days [IQR 31–57] and monitored by the eNose for a median duration of 73 hours [IQR 72–113]. A total of 887 hours of eNose data was collected. Patient characteristics are described in Table 1.

Feasibility and Safety
The eNose had no effect on pre-oxygenator pressure in the 4 different membranes that were tested. Table 2 shows the mean additional pre-oxygenator pressure for the 4 different membranes, with several in-membrane pressures. This indicated that we could safely connect the eNose to the membrane without interfering with extracorporeal life support.

eNose measurements
The eNose signal from the individual sensors was compared between the extracorporeal circulation and breath. A mean R² [95% CI] of 0.37 [0.15 – 0.59], 0.44 [0.21 – 0.67], 0.25 [0.02 – 0.49] and 0.44 [0.21 – 0.68] was found for sensors 1-4 respectively. Correlation plots for the four sensors can be found in Figure 1, with individual plots in the online supplement.

GC-MS measurements
At 69 sample moments, a total of 96 different VOCs were found by GC-MS in the breath and in the outlet gas from the extracorporeal circulation. Of these 96 VOCs, 29 (30%) were significantly correlated (P< 0.05) between the two sites. Seventeen VOCs that correlated significantly were identified (Table 3, Figure 2). These included VOCs from well described endogenous sources (e.g. 2-methyl-3-hexanone, acetone, ethyl acetate), VOCs from possible exogenous sources (e.g. D-Limonene, Chlorobenzene) and VOCs due to anesthetics (profofol and 1,1,1,3,3,3-hexafluoro-2-propanol [metabolite of sevoflurane]). Figure 3 shows the fold change between the breath concentration and extracorporeal air concentration, and the correlation coefficient between the two sampling sites. A lower correlation between the two sample sites was associated with a higher concentration of the molecule in breath compared to the extracorporeal air (Figure 3).

DISCUSSION
VOC analysis from the gas outlet of the extracorporeal support membrane was considered safe, as it did not increase pre-oxygenator pressure. With eNose measurements, a moderate correlation was found between the two measure-
ment sites, after correction for ventilator settings, suggesting that at least part of the exhaled VOCs are of non-pulmonary origin in critically ill mechanically ventilated ICU patients with severe pulmonary problems. Several VOCs that are known to originate from the circulation (e.g. acetone and propofol) were found to strongly correlate between exhaled breath and the gas outlet of the extracorporeal circulation. However, two-thirds of the VOCs that were found in breath and in the gas outlet did not significantly correlate and were found in a higher concentration in breath, which might suggest that these VOCs do not exclusively originate from the systemic circulation.

The concentration of two important VOCs, acetone and isoprene, that could be informative about the metabolic status of a patient [10] were found in higher concentrations in the outlet gas from the extracorporeal circulation than in the breath of these patients. These two concentrations can differ due to several factors. First, local production in the lung, for example by bacteria [18] or lipid peroxidation [19] might lead to a higher exhaled concentration. Second, the diffusion capacity may differ selectively between the lung and the extracorporeal circulation for different compounds, which might increase or decrease the breath concentration. Third, in mechanically ventilated patients there is frequently a profound ventilation/perfusion mismatch in the lung, which might result in lower exhaled concentration of non-pulmonary produced VOCs due to non-circulated alveoli (e.g. dead space ventilation). The ventilation/perfusion mismatch seems most plausible for acetone and isoprene based on the presented results in this study as VOCs known to originate from blood are consistently found in lower concentrations in breath in this study (table 3). This is also in line with previous literature; we were not able to predict changes in glucose concentrations based on exhaled VOC in critically ill mechanically ventilated patients [12]. In a very selected group of critically ill patients the higher amounts of acetone measured in air coming from external outlet of an extracorporeal life support membrane could provide possibilities for online monitoring.

Only approximately one third of the VOCs that were found persistently in breath and air from the outlet of the extracorporeal circulation significantly correlated. As discussed above, this can have three reasons. 1-propanol is one of the VOCs that is found in higher concentration in breath. This molecule has recently been linked to bacterial metabolism and an inflammatory response in the lung [20]. We speculate that many of the molecules that were not found to correlate are produced locally in the alveoli and airways and are therefore not detectable in the extracorporeal circulation. This is further supported by the observation that VOCs that did not correlate between the breath and the air from the extracorporeal circulation were more likely to be present in higher concentrations in the breath (figure 3). It should be noted that VOCs from the extracorporeal
circulation and from breath can only be compared in a highly selected group of patients and it is unknown how these results translate to other patient groups. However, it seems plausible that for patients with severely damaged lungs, requiring extracorporeal support, the airways and alveoli contribute to a large proportion of the VOCs in the exhaled breath.

This is the first study to analyze VOCs in the air coming from external outlet of an extracorporeal life support membrane using an eNose and GC-MS. The results of this study indicate that this can be done safely, and, based on the molecular identification of the VOCs found, we speculate that there is a potential for monitoring of VOCs that are likely to have a non-pulmonary origin in the extracorporeal circulation. Nonetheless, several weaknesses of the study must be acknowledged. Because of the pragmatic observational nature of this study, we did not have control over patient selection, nor did we distinguish between different membranes used in these patients. In addition, adapters fitted to the external gas outlet of the membranes differed between the different membrane types as manufacturers used their own proprietary designs. Furthermore, we did not tightly control the ventilation parameters and the flow in the extracorporeal circulation, all of which might influence the concentration of the identified VOCs.

CONCLUSION

The results from this study support safety and feasibility of extracorporeal VOC analysis. This study indicates that VOCs of non-pulmonary origin can be measured in breath and in the extracorporeal circulation. For VOCs that did not correlate between the two collection sites, the breath concentration was higher, suggesting pulmonary production of these molecules in a highly selected population of critically ill patients requiring extracorporeal support.

References


Figure 1. eNose sensor values.
Breath sensor values plotted against ECMO sensor value for each sensor. Each color represents a different patient. The black line represents the correlation between the two sampling sites.
Figure 2.
Correlation between the concentration of VOCs in breath and in air coming from the extracorporeal circulation. The y-axis is the correlation coefficient between the concentration of VOCs in breath and in air coming from the extracorporeal circulation and the y-axis is the retention time of the VOCs. Numbers correspond to numbers in the list on the right.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acetone</td>
</tr>
<tr>
<td>2</td>
<td>Isoprene</td>
</tr>
<tr>
<td>3</td>
<td>2-Propanol, 1,1,1,3,3,3-hexafluoro-</td>
</tr>
<tr>
<td>4</td>
<td>1-Propanol</td>
</tr>
<tr>
<td>5</td>
<td>Pentane, 3-methyl-</td>
</tr>
<tr>
<td>6</td>
<td>Pentane, 2,4-dimethyl-</td>
</tr>
<tr>
<td>7</td>
<td>Furan, 2-methyl</td>
</tr>
<tr>
<td>8</td>
<td>Ethyl Acetate</td>
</tr>
<tr>
<td>9</td>
<td>2,4-Hexadiene</td>
</tr>
<tr>
<td>10</td>
<td>1-Pentene, 2,4,4-trimethyl-</td>
</tr>
<tr>
<td>11</td>
<td>Unknown</td>
</tr>
<tr>
<td>12</td>
<td>Unknown</td>
</tr>
<tr>
<td>13</td>
<td>Unknown</td>
</tr>
<tr>
<td>14</td>
<td>Unknown</td>
</tr>
<tr>
<td>15</td>
<td>Benzene, chloro-</td>
</tr>
<tr>
<td>16</td>
<td>3-Hexanone, 2-methyl-</td>
</tr>
<tr>
<td>17</td>
<td>2-Propanol, 1,1,1-trichloro-2-methyl-</td>
</tr>
<tr>
<td>18</td>
<td>Unknown</td>
</tr>
<tr>
<td>19</td>
<td>Pentane, 2,2,3-trimethyl-</td>
</tr>
<tr>
<td>20</td>
<td>Unknown</td>
</tr>
<tr>
<td>21</td>
<td>D-Limonene</td>
</tr>
<tr>
<td>22</td>
<td>Unknown</td>
</tr>
<tr>
<td>23</td>
<td>Undecane</td>
</tr>
<tr>
<td>24</td>
<td>Unknown</td>
</tr>
<tr>
<td>25</td>
<td>Unknown</td>
</tr>
<tr>
<td>26</td>
<td>Unknown</td>
</tr>
<tr>
<td>27</td>
<td>Unknown</td>
</tr>
<tr>
<td>28</td>
<td>Unknown</td>
</tr>
<tr>
<td>29</td>
<td>Propofol</td>
</tr>
</tbody>
</table>
Figure 3.
Fold change between the breath concentration and extracorporeal air concentration and the correlation coefficient between the two sampling sites. The x-axis corresponds to the correlation between VOCs in breath and in air coming from the extracorporeal circulation and the Y-axes is the fold change in VOC abundance between the two sites.

Table 1.
Patient characteristics. IQR, Interquartile range; ARDS, Acute Respiratory Distress Syndrome; ICU, Intensive Care Unit; LOS, Length of stay
<table>
<thead>
<tr>
<th>Device</th>
<th>In-membrane pressure (mmHg)</th>
<th>Additional pressure, Mean [Range, cmH2O]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maquet 60</td>
<td>-0.56 [-2.00 – 0.33]</td>
<td></td>
</tr>
<tr>
<td>Maquet 120</td>
<td>-0.18 [-3.00 – 1.33]</td>
<td></td>
</tr>
<tr>
<td>Maquet 150</td>
<td>0.07 [-1.33 – 1.00]</td>
<td></td>
</tr>
<tr>
<td>Medos 60</td>
<td>-0.04 [-0.67 – 0.33]</td>
<td></td>
</tr>
<tr>
<td>Medos 120</td>
<td>-0.02 [-1.00 – 0.67]</td>
<td></td>
</tr>
<tr>
<td>Medos 150</td>
<td>0.04 [-0.33 – 0.33]</td>
<td></td>
</tr>
<tr>
<td>Novalung 60</td>
<td>-0.11 [-1.33 – 0.00]</td>
<td></td>
</tr>
<tr>
<td>Novalung 120</td>
<td>-0.31 [-1.33 – 0.00]</td>
<td></td>
</tr>
<tr>
<td>Novalung 150</td>
<td>-0.04 [-0.67 – 0.33]</td>
<td></td>
</tr>
<tr>
<td>Sorin 60</td>
<td>-0.04 [-0.67 – 0.00]</td>
<td></td>
</tr>
<tr>
<td>Sorin 120</td>
<td>-0.16 [-1.00 – 0.33]</td>
<td></td>
</tr>
<tr>
<td>Sorin 150</td>
<td>0.00 [0.00 – 0.00]</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Mean additional pressure in front of the membrane when eNose is connected to membrane gas outlet

<table>
<thead>
<tr>
<th>Retention Time (m)</th>
<th>Correlation r</th>
<th>P-value</th>
<th>Proportion Breath/ECMO</th>
<th>Identity</th>
<th>Presumed Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.32</td>
<td>0.56</td>
<td>0.0018</td>
<td>0.61</td>
<td>Acetone</td>
<td>Endogenous</td>
</tr>
<tr>
<td>2.47</td>
<td>0.37</td>
<td>0.0472</td>
<td>0.55</td>
<td>Isoprene</td>
<td>Endogenous</td>
</tr>
<tr>
<td>2.78</td>
<td>0.84</td>
<td>0.0000</td>
<td>0.99</td>
<td>2-Propanol, 1,1,1,3,3,3-hexafluoro-</td>
<td>Anesthetic</td>
</tr>
<tr>
<td>3.22</td>
<td>0.38</td>
<td>0.0395</td>
<td>1.26</td>
<td>1-Propanol, 3-methyl-Pentane, 2,4-dimethyl-</td>
<td>Endogenous</td>
</tr>
<tr>
<td>3.50</td>
<td>0.45</td>
<td>0.0139</td>
<td>0.4</td>
<td>Pentane, 3-methyl-</td>
<td>Endogenous</td>
</tr>
<tr>
<td>3.80</td>
<td>0.41</td>
<td>0.0255</td>
<td>1</td>
<td>2,4-dimethyl-</td>
<td>Endogenous</td>
</tr>
<tr>
<td>4.20</td>
<td>0.48</td>
<td>0.0090</td>
<td>0.80</td>
<td>Furan, 2-methyl</td>
<td>Exogenous</td>
</tr>
<tr>
<td>4.22</td>
<td>0.53</td>
<td>0.0029</td>
<td>0.54</td>
<td>Ethyl Acetate</td>
<td>Endogenous</td>
</tr>
<tr>
<td>5.78</td>
<td>0.43</td>
<td>0.0197</td>
<td>0.99</td>
<td>2,4-Hexadiyne</td>
<td>Exogenous</td>
</tr>
<tr>
<td>7.33</td>
<td>0.89</td>
<td>0.0000</td>
<td>1</td>
<td>1-Pentene, 2,4,4-trimethyl-</td>
<td>Exogenous</td>
</tr>
<tr>
<td>8.85</td>
<td>0.82</td>
<td>0.0000</td>
<td>1.18</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>9.72</td>
<td>0.38</td>
<td>0.0431</td>
<td>1.27</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>9.83</td>
<td>0.79</td>
<td>0.0000</td>
<td>1.19</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>11.40</td>
<td>0.62</td>
<td>0.0004</td>
<td>0.89</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>PM</td>
<td>Vol</td>
<td>Ppm</td>
<td>Rel</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>11.47</td>
<td>0.67</td>
<td>0.0001</td>
<td>0.83</td>
<td>Exogenous</td>
<td></td>
</tr>
<tr>
<td>11.95</td>
<td>0.68</td>
<td>0.0000</td>
<td>0.41</td>
<td>Endogenous</td>
<td></td>
</tr>
<tr>
<td>14.47</td>
<td>0.56</td>
<td>0.0015</td>
<td>0.83</td>
<td>Exogenous</td>
<td></td>
</tr>
<tr>
<td>14.47</td>
<td>0.70</td>
<td>0.0000</td>
<td>0.99</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>14.88</td>
<td>0.41</td>
<td>0.0286</td>
<td>0.62</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>15.23</td>
<td>-0.44</td>
<td>0.0161</td>
<td>6.35</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>15.55</td>
<td>0.73</td>
<td>0.0000</td>
<td>1.22</td>
<td>D-Limonene</td>
<td></td>
</tr>
<tr>
<td>17.07</td>
<td>0.60</td>
<td>0.0006</td>
<td>0.90</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>18.55</td>
<td>0.67</td>
<td>0.0001</td>
<td>0.66</td>
<td>Undecane</td>
<td></td>
</tr>
<tr>
<td>19.03</td>
<td>0.62</td>
<td>0.0003</td>
<td>0.84</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>19.53</td>
<td>0.76</td>
<td>0.0000</td>
<td>0.82</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>20.65</td>
<td>0.86</td>
<td>0.0000</td>
<td>0.77</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>21.10</td>
<td>0.75</td>
<td>0.0000</td>
<td>0.88</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>21.22</td>
<td>0.60</td>
<td>0.0005</td>
<td>0.94</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>21.32</td>
<td>0.40</td>
<td>0.0299</td>
<td>0.18</td>
<td>Propofol</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.**

VOCs significantly correlated between the external outlet of the membrane and the outlet of the tube of the ventilator of the patient.
Chapter 10

Summary and General Discussion

Jan Hendrik Leopold
The objectives of this thesis outlined in the introduction were divided into three parts. First, we aimed to compare and test different (continuous) blood glucose measurement methods (part I). Second, we aimed to predict blood glucose levels by analysis of exhaled breath in intubated and ventilated ICU patients (part II). Finally, we aimed to investigate further development of exhaled breath analysis techniques and data analysis methods (part III). In this chapter, we will summarize the three parts of this thesis and will draw final conclusions, followed by future perspectives.

**SUMMARY**

**Part I**

Chapters 2-4 of this thesis show that many different techniques are currently being investigated and used for the continuous measurement of blood glucose levels in patients in the ICU [1-3]. Most methods of the studied blood glucose measurement systems require the insertion of a catheter in the blood stream or measure subcutaneously and are thus invasive in one way or another. Also, while measurement accuracy is often good, continuous glucose measurement methods tend not to be as accurate as point of care-based methods, which are used in daily practice. The interstitial continuous glucose monitoring (CGM) device tested in chapter 3 performed poorly with only 77% of 929 comparative measurements in zone A of the Clarke error grid [2]. Therefore, it did not meet the set accuracy standards set by consensus implying that 95% of paired values need to be in zones A of the Clarke error grid to qualify a device as point accurate [4]. In addition, many reliability problems concerning the subcutaneous sensors were encountered. With 93.6% of values in zone A of the Clarke error grid in the study using a micro-dialysis-based device in chapter 4, accuracy was moderate to good [3]. However, these results indicate that this device also does not meet the standards. As in the study in chapter 3, there were reliability problems: both glucose sensors and specialized central venous catheters had to be replaced prematurely, or where malfunctioning unexpectedly. Therefore, it cannot be recommended that any of the tested devices is used in clinical practice before these practical problems are overcome.

**Part II**

We set out to review literature available on the correlation between volatile organic compounds (VOCs) in exhaled breath and the plasma concentration of glucose to evaluate the state of the art evidence. In chapter 5, we showed that seven out of nine studies included in our systematic review indeed found a strong correlation between VOCs in exhaled breath and blood glucose levels [5]. Glucose levels were associated with VOCs such as ketone bodies, VOCs produced by gut flora, exogenous compounds and markers of oxidative stress.
While many of these studies used an oral glucose tolerance test or clamp study design, which resulted in a predictable trajectory of blood glucose levels, it does indicate that using exhaled breath for prediction of blood glucose is feasible. Because of the chosen study design in the literature and the non-continuous fashion of the measurements, we had no indication on how breath analysis for glucose prediction would perform in an ICU setting. In addition, there had not been any continuous eNose measurements in critically ill ICU patients. Therefore, we had to investigate how to filter possible noise from the derived signal. In chapter 6 this was investigated by continuously monitoring exhaled breath in 23 ICU patients for a total of 1251 hours [6]. Ventilator variables were also collected prospectively. We aimed to investigate whether changes in humidity in the side-stream, patient-ventilator disconnections and changes in ventilator settings introduced noise into the signal, and to discuss several approaches to reduce this noise. We found that humidity in the side-stream had an effect on eNose sensor values and were able to correct for this. Also, sensor values were associated with end-tidal CO2, tidal volume and the pressure variables and were corrected accordingly. This study paved the way to investigating the correlation between the eNose signal, and continuous glucose measurements in Chapter 7. The same dataset collected in the previous chapter was combined with CGM values collected in chapter 2 and 3 [7]. A total of 1165 hours of paired measurements in 23 patients were investigated. We hypothesized that continuous exhaled breath analysis using an eNose could be used to predict blood glucose levels in intubated and mechanically ventilated ICU-patients. We tested several regression models to correlate the two signals but we were not able to find a meaningful correlation between the two. This was in contrast to previous studies under controlled conditions that did find a strong correlation between exhaled breath and blood glucose [5]. A possible explanation for why we were unable to find a correlation was large inter-patient variation in ICU patients. Due to their critical illness, ICU-patients may have great variance in VOCs in exhaled breath. Additionally, the sensors in the eNose used in this study were possibly not specific enough in detecting VOCs that are associated with changes in glucose levels.

Part III
Chapter 8 of this manuscript dives into the different methods used throughout the field for the analysis of eNose data [8]. Our objective for this manuscript was to investigate which statistical methods were best for eNose data analysis. To get an insight into the current state of the art on pre-processing, classification and validation, we systemically reviewed the literature and listed all the statistical analyses and validation techniques that were used. A total of 64 studies were included in our analysis. Authors of studies that included an
external validation cohort were contacted and asked to share their original datasets with us to compare the analysis methods we found. Four datasets were used for the comparison. No single combination of dimension reduction and classification methods gave consistently better results between internal and external validation sets in this sample of four datasets. Also, performance of analysis methods was compared between the training cohort, after internal validation, and after external validation to assess the external validity of the performance. As expected, performance of the classification methods decreased after internal validation, and decreased further after external validation. We concluded that it is not meaningful to estimate diagnostic performance without external validation and that several classification methods should be assessed when performing studies on eNose data. Authors should set out to define the statistical techniques they will use before finalizing the data and should report adhering strictly to their analysis plan. In Chapter 9, we studied whether or not it is safe to investigate air coming from the membrane of extracorporeal support devices. Also, we compared signals to those in exhaled breath. We hypothesized that there is a correlation between VOCs in exhaled breath and VOCs collected in extracorporeal support devices for VOCs that are known to be of non-pulmonary origin. After assessment of the safety of connecting a breath-sampling device to four different extracorporeal support membranes, ten critically ill patients were monitored in this study. This was done by using an eNose to continuously monitor breath and air from extracorporeal support devices at the same time. In addition, a daily sample for analysis by gas chromatography and mass spectroscopy (GC-MS) was simultaneously taken at both sites. We found that eNose signals showed a moderate correlation (R² between 0.25 and 0.44) between the two measurement sites. Using GC-MS, we found that 96 VOCs were present both in breath and air from the extracorporeal support devices. Of these 96, 29 correlated significantly between the two sites, of which 17 were identified. VOCs that did not correlate were found in a higher concentration in breath than in air from the extracorporeal support devices, suggesting pulmonary production of these molecules. The results of this study suggest that in critically ill patients, VOCs that are most likely of non-pulmonary origin could be measured reliably in air coming from the membrane of extracorporeal support devices.

GENERAL DISCUSSION
A wide number of different techniques for glucose measurement are currently studied on the ICU. All of these techniques are invasive to a varying extent and most are not as accurate as point of care methods that are routinely available. In addition, the tested CGM devices have varying accuracy and low reliability due to problems with sensors and other disposables. While there is hope that
CGM can be beneficial for ICU patients, both in quality and safety of blood glucose control, this is still up for discussion.

In a systematic review, we found that studies suggest that there is a possible association between VOCs in exhaled breath and blood glucose levels. However, designs of the studies reviewed in chapter 5 all lead to a very predictable trajectory of blood glucose levels. In addition, patients in these studies were fairly similar and apart from having DM type 1 & 2, lacked many co-morbidities. Therefore, it is likely that an actual breath-based glucose measurement device can be developed for this particular group of patients. Trying to get similar results in critically ill ICU patients using a continuous eNose however, was not successful. This means that currently, continuous exhaled breath analysis using an eNose cannot be used to accurately predict blood glucose levels in intubated, mechanically ventilated patients. Large inter-patient variation and a limited specificity of the used sensor array are the most likely cause for this. Nonetheless, we now have a clear picture on the sources of noise that disturb the eNose signal in this group of patients and know how to filter this noise out. A breath based technique for glucose measurement in this group though, is still far from reality.

A wide variety of statistical techniques are used when analyzing eNose datasets. When comparing these techniques, there does not seem to be a single best method for analysis of eNose data. Therefore, several (combinations of) analysis techniques should be tested. The data clearly show that it is not meaningful to interpret study performance without use of an external validation set to check the obtained results. Finally, it seems safe to monitor air coming from the membrane of extracorporeal support devices. We found that there is a moderate overall correlation between exhaled breath and air coming from the membrane of extracorporeal support devices. This correlation was better for VOCs thought to be of non-pulmonary origin. VOCs that did not correlate between the two sites were found in a higher concentration in the exhaled breath, suggestive of an alveolar or interstitial source of production.

The results of the various chapters in this thesis call for investing in further research in several topics. While most clinicians agree that some form of glycemic control is necessary, the optimal range of blood glucose is still unknown [9]. This should therefore be the main focus of studies on glycemic control in the ICU. To further investigate whether glucose can be monitored through exhaled breath several steps could be taken. First, the continuous eNose used in our study should be adapted to contain sensors which are more sensitive to the VOCs we found in our systematic review. Second, a more homogenous group of intubated and ventilated patients with a smaller number of co-morbidities should be investigated to determine the potential of the technique. A possibility would be to use patients who undergo non-critical surgery.
This would minimize inter-patient variation and would possibly yield better results. As for the analysis of air coming from the membrane of extracorporeal support devices, more research is necessary. The first experiments as described in chapter 9 show promise but were performed under sub-optimal conditions. Future studies should have a pre-defined group of patients, preferably using one type of extracorporeal support membrane. In addition, more clinical parameters should be registered continuously.

**FUTURE PERSPECTIVES**

The data analysis methods used in several chapters of this thesis can be considered machine learning techniques. These techniques enable researchers and physicians to increasingly rely on computers to make decisions for them based on previous collected data. While computerized decision support systems (CDDS) previously relied on hard-coded rules programmed by their developers, it is now possible that these systems make their own conclusions based on the data we feed in. This enables the systems to keep improving while they gather new data, and most importantly, learn from their mistakes. While a simple CDDS may only need little knowledge to function, more complex problems can only be solved with a large amount of data. It is therefore important that investments are made in platforms that enable researchers to share their (anonymized) patient data for others to use. Ideally, this is combined with a world-wide network of computers to analyze this immense amount of data. This method of extreme-parallel computing, already successfully used in the search for extra-terrestrial life [10] and gene folding [11], will help to vastly speed up the process of analyzing large datasets. Continuous eNose data combined with the many clinical parameters gathered in intensive care units today is the perfect candidate for such a platform. We must however keep in mind that while these approaches may make the use of highly intelligent CDDS a reality, they will most likely solely act as a tool for physicians for the years to come because personal contact between patients and their doctor is hard to replace.

**References**

diagnostics in critically ill patients in comparison to continuous glucose monit... - PubMed - NCBI. J Breath Res 2017, 11, 026002.


CONTINUOUS GLUCOSE AND EXHALED BREATH ANALYSIS IN THE INTENSIVE CARE UNIT
**ACHTERGROND**

Als patiënten worden opgenomen op een Intensive Care (IC) zijn ze ernstig ziek. Een groot gedeelte van deze ernstig zieke patiënten wordt geïntubeerd en mechanisch geventileerd, en worden hun vitale functies gemonitord. Idealiter gebeurt dit monitoren continu, maar meestal is dit technisch niet mogelijk waardoor de waarde van de metingen beperkt blijft. Patiënten op een IC zouden dan ook groot belang hebben bij technologische vooruitgang in monitoring. Idealiter zouden dit continue en non-invasieve metingen zijn (dus niet door middel van frequent bloedprikken, zoals nu meestal het geval is).

Eén van de gebieden waarin continue monitoring voordeel zou kunnen brengen in IC-patiënten is de glucosehuishouding. Er bestaan verschillende continue glucosemonitoren (CGM) maar deze zijn vaak invasief en soms ook niet accuraat. In het eerste deel van dit proefschrift vergelijken we manieren om continue glucose te meten en testen we een tweetal apparaten. Een mogelijke manier om glucose non-invasief en continue te meten is in uitademingslucht van een beademde patiënt. Dit is wellicht mogelijk doordat er in uitademingslucht naast koolstofdioxide (CO₂), zuurstof, stikstof en waterdamp ook duizenden vluchtige organische componenten (VOC’s) zitten. Mogelijk kunnen deze VOC’s gerelateerd worden aan fysiologische processen in het lichaam. Wij behandelen twee methoden voor uitademingsluchtanalyse in dit proefschrift. Ten eerste de continue elektronische neus (eNose), waarmee combinaties van VOC’s herkend kunnen worden doordat ze een “vingerafdruk” achterlaten. Gebruikmakend van patroonherkenning proberen we deze “vingerafdruk” aan bepaalde fysiologische veranderingen te koppelen. Daarnaast behandelen we gaschromatografie en massaspectroscopie (GC–MS), waarmee VOC’s in een gasmengsel gescheiden, geïdentificeerd en gekwantificeerd kunnen worden. Hierdoor kan er met deze techniek een zeer accurate momentopname gegeven worden van de samenstelling van uitademingslucht. In het tweede deel van dit proefschrift testen we de hypothese dat het mogelijk is om met een continue eNose bloedglucose te meten in de uitademingslucht van, beademde patiënten. In het derde en laatste deel van dit proefschrift onderzoeken we welke dataanalyse methoden er gebruikt worden voor eNose data, en welke combinatie van methoden het beste presteert. Ook doen we onderzoek in een groep ernstig zieke patiënten die door slechte pulmonaire of cardiale functie extracorporale ondersteuning nodig hebben. In deze groep testen we de hypothese dat er een correlatie is tussen VOC’s in uitademingslucht en VOC’s in de gas die langs het membraan van extracorporale ondersteuningsapparaten passeert.

**SAMENVATTING**

**Deel I**

In het eerste deel van deze thesis laten we zien dat er veel verschillende tech-
nieken worden gebruikt om continue bloedglucose te meten op de IC [1-3]. In hoofdstuk 2 werd duidelijk dat de meeste continue bloedglucosemettechnieken invasief zijn, ze vereisen dat er een katheter in de bloedstroom wordt ingebracht, of subcutaan meten. Daarnaast wordt duidelijk dat, ondanks de redelijke diagnostische accuratesse van de onderzochte apparaten, laboratorium- of bloedgas-gebaseerde technieken nog altijd accurater zijn. In hoofdstuk 3 onderzochten we een CGM-apparaat dat glucose met kleine naaldjes onder huid meet. Dit GCM-apparaat presteerde slecht, aangezien maar 77% van de 929 vergelijkbare metingen accuraat genoeg waren [2]. Daarmee voldeed het niet aan de vereiste accuraatheid aangezien minimaal 95% van de waarden voldoende accuraat moeten zijn [4]. Daarnaast waren er ook betrouwbaarheidsproblemen met de subcutane sensoren. Het op microdialyse gebaseerde CGM apparaat dat in hoofdstuk 4 onderzocht werd presteerde matig tot goed met 93.6% van de meetwaarden genoeg accuraat [3]. Ook dit GCM-apparaat had last van betrouwbaarheidsproblemen. Zowel de glucosesensoren als de speciale centrale lijnen moesten vroegtijdig vervangen worden, of stopten er onverwachts mee. Hierdoor is het niet aan te raden om deze apparaten in de klinische praktijk te gebruiken tot deze problemen overkomen zijn.

Deel II
Om onze hypothese te toetsen dat uitademingslucht gebruikt kan worden bloedglucosewaarden te meten begonnen we met een review van de beschikbare literatuur. In hoofdstuk 5 vonden we dat zeven van de negen geïncludeerde studies een sterke correlatie vinden tussen VOC’s en bloedglucosewaarden [5]. Veranderingen in glucose werden onder anderen geassocieerd met ketonlichamen, VOC’s die werden geproduceerd door darmflora, exogene componenten die werden ingeademd en markers van oxidatieve stress. De meeste studies waren intraveneuze of orale glucosetolerantietesten waardoor patiënten te maken hadden met voorspelbare veranderingen in bloedglucosewaarden. Desondanks lijkt het mogelijk om bloedglucosewaarden te voorspellen met behulp van uitademingslucht. Door het studieontwerp en de niet-continue metingen in de onderzochte studies was het niet duidelijk of de resultaten direct te vertalen waren naar IC-patiënten. Daarnaast waren er tot dan toe geen studies met een continu metende eNose bij IC-patiënten uitgevoerd. Hierdoor was het niet duidelijk hoe dergelijke signalen voorbehandeld moesten worden met betrekking tot mogelijke ruis in het signaal. In hoofdstuk 6 van dit proefschrift onderzochten we dit door continue eNose uitademingslucht te onderzoeken in 23 patiënten [6]. Dit leverde in totaal 1251 uur aan eNose-data op. Ventilatorwaarden en beademingsinstellingen werden eveneens prospectief verzameld. Het doel van het onderzoek was nagaan of veranderingen in luchtvochtigheid in de slang naar de eNose, ontkoppelingen van de patiënt van de
ventilator, en veranderingen in ventilatorinstellingen invloed hadden op het signaal van de eNose. Daarnaast werd onderzocht wat de beste manieren waren om deze ruimte te verminderen. Verandering van de luchtvochtigheid in de slang bleek, zoals verwacht, invloed te hebben op eNose-sensorwaarden en het was mogelijk om hiervoor te corrigeren. Ook de koolstofdioxide concentratie in de adem, teugvolume en verschillende drukparameters bleken invloed te hebben op eNose-sensorwaarden. Ook hiervoor kon gecorrigeerd worden. Met de in deze studie ontwikkelde methoden konden we onderzoeken of er een correlatie was tussen het eNose-signalen en continue glucosemetingen. In hoofdstuk 7 combineerden we de data en methodes van hoofdstuk 6, met de continue glucosemetingen van hoofdstukken 3 en 4 [7]. In totaal onderzochten we 1165 uur aan gepaarde metingen van 23 verschillende patiënten. Onze hypothese was dat continue uitademingsanalyse met een eNose gebruikt kon worden om bloedglucosewaarden te voorspellen in geïntubeerde, beademde IC-patiënten. Met verschillende regressiemodels konden we echter geen noemenswaardige correlatie vinden tussen de twee signalen. Een mogelijke verklaring hiervoor was de hoge mate van variabiliteit tussen IC-patiënten die ontstaat door ernstige ziekte. Hierdoor ontstaan er naar alle waarschijnlijkheid grote verschillen in de samenstelling van VOC’s in uitademingslucht. Een andere mogelijke verklaring voor dit resultaat was dat de sensoren in de gebruikte eNose niet specifiek genoeg waren voor de VOC’s die geassocieerd zijn met glucosewaarden.

Deel III
In hoofdstuk 8 onderzochten we welke verschillende methodes er worden gebruikt om eNose data te analyseren [8]. Om een inzicht te krijgen in welke verwerkings-, analyse- en validatiemethodes er worden gebruikt bij eNose-onderzoek, zochten we in alle beschikbare literatuur. In totaal selecteerden we 64 studies en de in deze studies gebruikte technieken werden bestudeerd. Aan auteurs van studies met een validatiecohort werd gevraagd of we die dataset konden gebruiken om de gevonden technieken op te testen. In totaal ontvingen we vier datasets. Er bleek geen enkele combinatie van een verwerkings- en analysetechniek te zijn die in elke dataset het beste presteerde. Daarnaast werd externe validiteit van de gevonden resultaten onderzocht door prestaties van het trainingscohort na interne validatie en vervolgens na externe validatie te vergelijken. Zoals verwacht werd duidelijk dat prestaties minder werden na interne validatie, en nog verder afnamen na externe validatie. Hierdoor concludeerden we dat het niet zinvol was om prestaties van een eNose weer te geven zonder externe validatie. Ook moeten er verschillende data-analyse technieken vergeleken worden wanneer er onderzoek wordt gedaan met eNose data. In hoofdstuk 9 onderzochten we of het veilig is om lucht te onderzoeken dat van het membraan van extracorporale ondersteuningsapparaten komt. Onze hypo-
these was dat er een correlatie is tussen VOC's in uitademingslucht en VOC's in lucht die van het membraan van extracorporale ondersteuningsapparaten komt, voor VOC's van niet-pulmonale oorsprong. Nadat we gevonden hadden dat het veilig is om luchтанalyse te doen op extracorporale ondersteuningsmembranen, begonnen we met analyses in tien ernstig zieke patiënten. We deden dit door met een eNose continu te meten in zowel de uitademingslucht als de lucht die van het membraan van extracorporale ondersteuningsapparaten komt. Ook namen we elke dag een luchtmonster voor GC-MS-analyse. Met de eNose vonden we een gematigde correlatie (R² tussen 0,25 en 0,44) tussen de twee meetlocaties. Met GC-MS vonden we 96 VOC's die in beide luchtronnen voorkwamen waarvan er 29 significant gecorreleerd waren. Zeventien hiervan konden worden geïdentificeerd. VOC's die niet significant correleerden, kwamen in hogere concentraties voor in de uitademingslucht. Dit suggereert dat in VOC's van niet-pulmonale oorsprong met enige vorm van betrouwbaarheid konden worden gemeten in lucht dat van het membraan van extracorporale ondersteuningsapparaten komt.

References

Appendices
Continuous Glucose Monitoring–devices for Use in Intensive Care Units
R.T.M. van Hooijdonk: Literature review, writing manuscript. J.H. Leopold: Literature review, writing manuscript. M.J. Schultz: Literature review, writing manuscript.

Point Accuracy and Reliability of an Interstitial Continuous Glucose Monitoring Device in Critically Ill Patients: A Prospective study
R.T.M. van Hooijdonk: study design, data collection, data analysis, drafting manuscript. J.H. Leopold: study design, data collection, data analysis and reviewing manuscript. T. Winters: study design, data collection, reviewing manuscript. J.M. Binnekade: study design, data analysis, reviewing manuscript. N.P. Juffermans: study design, data collection, reviewing manuscript, J. Horn: study design, data collection, reviewing manuscript, J.C. Fischer: study design, reviewing manuscript. E.C. van Dongen–Lases: study design, reviewing manuscript. M.J. Schultz: study design, data collection, drafting manuscript.

Point and Trend Accuracy of a Continuous Intravenous Microdialysis–based Glucose–monitoring Device in Critically Ill Patients – a Prospective study
J.H. Leopold: study design, data collection, data analysis, drafting manuscript. R.T.M. van Hooijdonk: study design, data collection, data analysis and reviewing manuscript. M. Boshuizen: data collection and reviewing manuscript. T. Winters: data collection and reviewing manuscript. L.D. Bos: study design, data analysis and reviewing manuscript. A. Abu–Hanna: study design and reviewing manuscript. A.M.T. Hoek: study design, reviewing manuscript. J.C. Fischer: study design, reviewing manuscript. E.C van Dongen–Lases: study design, reviewing manuscript. M.J. Schultz: study design, reviewing manuscript.

Glucose Prediction by Analysis of Exhaled Metabolites – a Systematic Review
J.H. Leopold: study design, data extraction, data review, drafting manuscript. R.T.M. van Hooijdonk: study design, reviewing manuscript. P.J. Sterk: study design, reviewing manuscript. A. Abu-Hanna: study design, reviewing manuscript. M.J. Schultz: study design, reviewing manuscript. L.D.J. Bos: study design, data extraction, data review, drafting manuscript.

Factors influencing continuous breath signal in intubated and mechanically–ventilated intensive care unit patients measured by an electronic nose
J.H. Leopold: study design, data collection, data analysis, drafting manuscript.
A. Abu-Hanna: study design, data analysis, reviewing manuscript. C. Colombo: Data analysis, reviewing manuscript. P.J. Sterk: study design, reviewing manuscript, M.J. Schultz: study design, reviewing manuscript. L.D.J. Bos: study design, data analysis, drafting manuscript.

**Non-invasive breath monitoring with eNose does not improve glucose diagnostics in critically ill patients in comparison to Continuous Glucose Monitoring in blood**
J.H. Leopold: study design, data collection, data analysis, drafting manuscript, L.D.J. Bos: study design, data analysis, drafting manuscript, C. Colombo: data analysis, reviewing manuscript, P.J. Sterk: Study design, reviewing manuscript, M.J. Schultz: Study design, reviewing manuscript, A. Abu-Hanna: Study design, data analysis, reviewing manuscript.

**Comparison of classification methods in breath analysis by electronic nose**
J.H. Leopold: study design, data collection, data analysis, drafting manuscript. L.D.J. Bos: study design, data collection, data analysis, drafting manuscript. P.J. Sterk: study design, reviewing manuscript. M.J. Schultz: study design, reviewing manuscript, N.Fens: study design, reviewing manuscript. I. Horvath: study design, reviewing manuscript. A. Bikov: study design, reviewing manuscript. P. Montuschi: study design, reviewing manuscript. C. Di Natale: study design, reviewing manuscript. D.H. Yates: study design, reviewing manuscript. A. Abu-Hanna: study design, data analysis, reviewing manuscript.

**Volatile organic compound profiles in outlet air from extracorporeal life-support devices differ from breath profiles in critically ill patients**
J.H. Leopold: study design, data collection, data analysis, drafting manuscript. A. Philipp: study design, reviewing manuscript. T. Bein: study design, reviewing manuscript. A. Redel: study design, reviewing manuscript. M. Gruber: study design, reviewing manuscript. M.J. Schultz: study design, reviewing manuscript. A. Abu-Hanna: study design, reviewing manuscript. P. Brinkman: data analysis, reviewing manuscript. H.G. Janssen: data analysis, reviewing manuscript. L.D.J. Bos: study design, data analysis, drafting manuscript.
**PhD period:** August 2013 – July 2017

### 1. PhD training

<table>
<thead>
<tr>
<th>Courses</th>
<th>Year</th>
<th>ECTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Crash Course</td>
<td>2013</td>
<td>0,4</td>
</tr>
<tr>
<td>- AMC World of Science</td>
<td>2013</td>
<td>0,7</td>
</tr>
<tr>
<td>- BROK</td>
<td>2014</td>
<td>0,9</td>
</tr>
<tr>
<td>- Project Management</td>
<td>2014</td>
<td>0,6</td>
</tr>
<tr>
<td>- E-Science</td>
<td>2014</td>
<td>0,7</td>
</tr>
<tr>
<td>- Informed Consent</td>
<td>2016</td>
<td>0,3</td>
</tr>
</tbody>
</table>

#### Seminars, workshops and master classes

- Doctoral Consortium AIME 2016 2016 0,5

#### Presentations

- Poster ISICEM 2014 0,5
- Poster ISICEM 2015 0,5
- Doctoral Consortium AIME 2016 0,5

(Inter)national conferences

- Medinfo 2013 1
- ISICEM 2014 1
- ISICEM 2015 1
- WATCH-conference 2015 0,5
- AIME 2016 1

**Other**

- Intensive Care Research meeting (weekly) 2013-2016 12
- LEICA Research meeting (weekly) 2013-2016 12
- Journal club LEICA 2013-2016 4
- Bite seminar AMC 2015 0,1
- Promovendiday Kik 2015 0,5
- Promovendiday Kik 2016 0,5

### 2. Teaching

**Lecturing**

- R-Course ICU PhD students 2013 1

**Supervising**

- Research internship (Medicine) 2014 1
- Research internship (Engineering) 2014 1

**Other**

- Aprove Board 2014 1
- Aprove Board 2014 1
- Aprove Board 2014 1

### 3. Parameters of Esteem

**Grants**

- Maquet cardiopulmonary 2014
4. Publications


Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord
—
Dankwoord

Ik ben veel dank verschuldigd aan

Alle patiënten, en hun families, die meegedaan hebben aan mijn onderzoek. Zonder jullie is medische vooruitgang onmogelijk. De IC-verpleegkundigen. Jullie waren altijd behulpzaam, ook al zat ik jullie vaak in de weg. Marcus, voor het vertrouwen en de kans die je mij hebt gegeven. Ik heb enorm veel van je geleerd. Ameen, voor je enthousiasme, betrokkenheid en warmte. Ik ben blij dat ik nu nog veel met je samen kan werken. Lieuwe, voor de onmisbaarheid in het opzetten van de studies, de analyses en het schrijven van de stukken. Zonder jou had ik er veel langer over gedaan. De leden van de promotiecommissie, voor de tijd die jullie hebben geïnvesteerd in het kritisch beoordelen van mijn proefschrift. Paranimfen Frank en Diederik. Frank voor het altijd behulpzaam zijn, altijd willen luisteren maar vooral ook voor je (slechte) humor. Je was een geweldige kamergenoot. Diederik voor het zijn van een fijne studiegroot, maar vooral voor het zijn van een fantastische vriend. Bij ons thuis gedraag je je als onderdeel van het gezin en dat is me zeer dierbaar. Margreeth, Nicole, Janneke, Frederique en Jan. Ik heb veel van jullie geleerd op de IC! Tineke voor je enorme dosis humor, maar vooral voor al je hulp. Je was onmisbaar voor me! Annelou, voor je vele hulp en voor het zijn van een fijne kamergenoot. Roos, voor het helpen van mij als student, en collega. Ik heb altijd enorm veel lol met je gehad. Emma, voor je humor en voor het helpen pesten van Frank. Ik heb enorm veel bewondering voor hoe je met tegenslagen omgaat. Marjolein, ik heb veel van je geleerd en was toen je kwam blij dat ik niet meer de enige nerd was | Gerie, voor het mij leren hoe je kinderen en promoveren kunt combineren. Marleen, doordat je Betondorp voor mij op de kaart hebt gezet en ik altijd enorm met je heb gelachen! Matt, Anna-Linda, David, Fabienne, Laura, Luuk, Margit, Matthijs, Poulte, Robert, Jenny, Sanne, Laween, Esther, Hendrik, Maike, Sophia, Marjon, Rianne, Sabrine, Friso, Charlotte, Lonneke en alle andere collega’s van de IC, omdat het zonder jullie veel minder leuk was geweest! Al mijn (nieuwe) collega’s op de KIK. De laatste maanden van mijn PhD heb ik bij jullie doorgebracht. Ik voel mij hier al helemaal thuis! Mijn APROVE collega’s Juliette, Ruth, Anne, Sanne, Louise, Marlot, Kilian, Arnoud, Wouter, Mitra, Simone, Olvert en Margriet. Door jullie weet ik dat er naast de IC ook nog ander leven is in het AMC! Simon, Fred en alle andere bij Comon Invent. Mede dankzij jullie techniek kon dit onderzoek gedaan worden. Excuses voor het meermaals verzuipen van de eNose. Thomas, Micheal, Andreas, Alois and all my other colleagues in Regensburg. Thanks for having us over and making our time there unforgettable. Peter Sterk, voor de goede adviezen en waardevolle bijdragen aan de artikelen in mijn proefschrift die de kwaliteit enorm hebben doen toenemen. Paul Brinkman en anderen bij de afdeling longziekten voor jullie hulp, en het vele goede werk in het veld van