THE UNIVERSITY OF SYDNEY

Advanced analyses of physiological signals and their role in Neonatal Intensive Care

Principal Supervisor
Prof. Alistair McEwan

Author
Jacqueline Huvanandana
B. Eng

Auxiliary Supervisors
Dr. Mark Tracy
Dr. Cindy Thamrin
Dr. Peter Jones

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

School of Electrical & Information Engineering

24 MAY 2018

Statement of authentication

This thesis is submitted to the University of Sydney in fulfilment of the requirement for the degree of Doctor of Philosophy. The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Jacqueline Huvanandana   24 May 2018

Acknowledgements

Thank you to the University of Sydney and School of Electrical & Information Engineering, for the opportunity to pursue my PhD and the scholarships supporting my studies.

I would like to express my sincere gratitude to the many people who have supported me throughout my PhD:

To my supervisor, Prof. Alistair McEwan, for his guidance, insightful discussion and the opportunities to develop and diversify my work. Thank you for his timely advice and understanding when it came to managing time, work and the many challenges along the way.

To my auxiliary supervisors and the balance of their perspectives: Dr. Mark Tracy, for his clinical insight and enthusiasm for all things NICU. To Dr. Cindy Thamrin, for her clarity and the opportunity to grow in research. To Dr. Peter Jones, for his pragmatic advice and patience.

To those who offered encouragement and guidance along the way: Prof. Heather Jeffery, Dr. Angela Carberry, Murray Hinder, Joseph Prinable, Dr. Chinh Nguyen, Dr. Alun Pope, Dr. Robin Turner, A/Prof. Camille Raynes-Greenow, Emily Bek and Prof. Philip de Chazal.

This PhD would not have become a reality without my family: to my Mum, for her guidance and belief in me, to Nina, for her encouragement, and to Coco & Ollie, for their company. To my Father, for the opportunities his support has allowed me. To Liam, for his confidence in me. To my Grandfather, for his inspiration and enduring enthusiasm for engineering.

Abstract

Preterm infants admitted to the neonatal intensive care unit (NICU) face an array of life-threatening diseases requiring procedures such as resuscitation and invasive monitoring, as well as other risks related to exposure to the hospital environment, all of which may have lifelong implications. For several decades, intensive care monitoring systems have displayed physiological data such as electrocardiogram (ECG), blood oxygen saturation and arterial blood pressure signals. Datapoints were subsequently transposed crudely to paper records in hourly intervals, rendering long-term pattern recognition impossible. In recent years, there has been considerable work utilising high-definition intensive care physiological data. The applications of advanced analyses in the NICU range from predictive monitoring to characterising patterns of disease. Time series analysis of ECG-derived heart rate data has demonstrated predictive value in early identification of neonatal sepsis (Griffin et al. 2001, Moorman et al. 2011) and intraventricular haemorrhage (IVH) (Tuzcu et al. 2009).

This thesis examined a range of applications for advanced signal analyses in the NICU, from identifying physiological patterns associated with neonatal outcomes, to evaluating the impact of certain treatments on physiological variability. Firstly, the thesis evaluated the potential of detrended fluctuation analysis (DFA), pre-processing and multivariable models to identify infants at risk of developing IVH, often interrelated with factors leading to preterm birth and use of mechanical ventilation.

Hypoxia and prolonged apnoeas may also predispose infants to white matter brain injury and/or IVH. A known respiratory stimulant, caffeine is regularly administered to infants to prevent and treat apnoea of prematurity, and to facilitate weaning off mechanical ventilation. The thesis applied Poincare analysis and DFA to heart rate and arterial blood pressure data to characterise the cardiovascular impact of caffeine in preterm infants, finding greater pulse pressure variability and elevated parasympathetic modulation. These observations suggested an enhanced responsiveness of the autonomic nervous system.

Cerebral autoregulation maintains cerebral blood flow despite fluctuations in arterial blood pressure and is an important consideration for preterm infants

who are especially vulnerable to brain injury. Using various time and frequency domain correlation techniques, the thesis found acute changes in cerebral autoregulation of preterm infants following caffeine therapy.

Nutrition in early life may also affect neurodevelopmental outcomes and morbidity in later life. This thesis developed a range of models for identifying malnutrition risk and demonstrated the predictive value of anthropometry and near-infrared interactance features in body composition assessment.

This thesis has presented and validated a range of ways in which advanced analyses including time series analysis, feature selection and model development can be applied to neonatal intensive care. There is a clear role for these analyses in early detection of clinical outcomes, characterising the effects of relevant treatments or pathologies and identifying infants at risk of later morbidity.

Contents

List of Appendices

The following publications and draft manuscript do not form a core part of the thesis, though were undertaken over the course of the PhD candidature. These appendices are additional to the body of work presented in the thesis.

I	First author conference paper on Logistic regression models for predicting intraventricular haemorrhage in preterm infants using respiratory and blood pressure signals
II	Co-author paper on Reducing false arrhythmia alarms in the ICU using multimodal signals and robust QRS detection
III	Co-author paper on Length-free near infrared measurement of newborn malnutrition
IV	Co-author paper on How do different brands of size 1 Laryngeal Mask Airway compare to facemask ventilation in a dedicated Larygeal Mask Airway teaching manikin?
V	Co-author paper on Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features
VI	Co-author abstract from 2017 PSANZ conference on Laryngeal Mask Airway comparison (Canberra, Australia)
VII	First author draft manuscript on Model development for body composition assessment using near-infrared interactance in infants aged 0-2 years

List of Abbreviations

ADP	Air displacement plethysmography
AUC	Area under curve
BF%	Body fat percentage
BMI	Body mass index
CAP	Caffeine for Apnoea of Prematurity
Coh	Coherence
COx	Cerebral Oximetry Index
CPAP	Continuous positive airway pressure
CRIB	Clinical Risk Index for Babies
DBP	Diastolic blood pressure
DCCA	Detrended cross-correlation analysis
DFA	Detrended fluctuation analysis
DXA	Dual energy X-ray absorptiometry
ECG	Electrocardiogram
EEG	Electroencephalogram
FFM	Fat-free mass
FM	Fat mass
HRC	Heart rate characteristic
HRV	Heart rate variability
IVH	Intraventricular haemorrhage
LF	Low frequency
LOOCV	Leave-one-out cross-validation
MAP	Mean arterial pressure
NEC	Necrotising enterocolitis
NICU	Neonatal Intensive Care Unit
NIR	Near-infrared
NIRS	Near infrared spectroscopy
PDA	Patent ductus arteriosus
ROC	Receiver-Operator Characteristic
SBP	Systolic blood pressure
SD	Standard deviation
TOI	Tissue Oxygenation Index
VLF	Very low frequency

Chapter 1

Introduction

This thesis focuses on the application of advanced analysis techniques in the neonatal intensive care unit (NICU), from the early identification of infants at risk of intraventricular haemorrhage (IVH) and malnutrition, to characterising physiological variability associated with caffeine therapy and cerebral autoregulation. This chapter discusses the motivation, objectives and research questions, and provides an overview of the thesis structure and related publications.

1.1   Research motivation

With technological advancement and the continuous availability of physiological signals monitored in the NICU, there has been a growing interest in advanced analyses of these signals for a range applications. Multiple opportunities are now available to characterise the behaviour of a complex system such as the human body [1], to identify potentially life-threatening changes (such as the onset of disease or injury), and to better understand the mechanisms behind treatments or the conditions under which they may be beneficial or harmful. These analyses contribute to the improvement of patient outcomes in both the short and long term.

Monitoring and management of patients in the NICU represents a unique balance of necessary life-saving procedures with other risks and potentially life-long impacts. Infants born preterm face the risk of serious brain injury, cerebral palsy and other neurodevelopmental outcomes [2–4]. Despite advancements in perinatal care which include using antenatal corticosteroids and magnesium sulphate, and the postnatal use of surfactant, their underdeveloped respiratory systems may still require mechanical ventilation. Cardiorespiratory management within the first 24 hours of birth is a key contributor to the development and timing of brain injury such as IVH [5, 6], with the well-recognised association

that patient-ventilator asynchrony may be a potential causal pathway [7, 8]. Apnoea of prematurity is similarly linked to immature respiratory control systems. While the long-term consequences of apnoea are unclear, the short-term effects include reduced systemic blood pressure and cerebral hypoperfusion, which may in turn contribute to injury of the immature, pressure-passive brain [9]. Caffeine is commonly prescribed as a respiratory stimulant to prevent and treat apnoea of prematurity [10], and to facilitate weaning off mechanical ventilation, being associated with improved neurodevelopmental outcomes [11, 12]. There is also evidence for the contribution of nutrition and body fat in early life to neurocognitive development [13, 14] and morbidity in later life [15, 16].

This thesis examines the role of advanced analyses in neonatal intensive care, from identifying infants at risk of IVH and malnutrition, to what physiological insights could be gained using these techniques with respect to caffeine therapy and cerebral autoregulation. Most studies to date have quantified physiological changes using linear metrics such as mean and standard deviation (SD). These metrics fail to take into account the pattern of changes over time or their codependency on one another [1, 17]. In contrast, advanced analysis techniques such as detrended fluctuation analysis (DFA) and Poincare analysis may better characterise the non-linear behaviour of systems such as heart rate, respiratory and blood pressure dynamics. Feature selection and cross-validation techniques may also offer value in developing and evaluating models to identify infants at risk of morbidity.

Firstly, this thesis explores the early identification of preterm infants at risk of developing IVH. Cardiorespiratory management in the early hours of life is an important risk factor in its development [5, 6], although no specific treatments have reached consensus. Earlier studies have reported altered heart rate variability following the development of IVH [18, 19]. However, there has been limited work exploring the predictive value of such physiological variability for identifying risk of IVH in the early hours of life [20]. There may also be value in evaluating the variability and dynamics of heart rate and blood pressure following caffeine therapy, the mechanisms of which remain unknown. Previous work in this area has focused on short- and long-term clinical outcomes [12, 21, 22], and used linear analyses to characterise acute physiological changes [23–25]. Non-linear analysis approaches such as DFA or Poincare analysis may be better placed to handle the non-stationarities often seen in physiological signals, where statistical properties are not constant over time. This work also includes the novel application of detrended cross-correlation analysis, a generalisation of the DFA method, to describe the dynamic relationship between blood pressure and cerebral oxygenation. Finally, this thesis develops models to identify neonatal and infant malnutrition.

Nutrition in early life may have lifelong implications and plays an important role in neurocognitive development [13, 14], although its widespread measurement is limited by a lack of simple, cost-effective and portable methods.

1.2 Objectives

The core objectives of this thesis are as follows:

1. To critically appraise the literature on variability analysis as it has been applied to neonatal intensive care

2. To explore the potential of time series analysis of physiological signals in the early identification of preterm infants at risk of intraventricular haemorrhage

3. To characterise the effect of caffeine, a commonly used treatment for apnoea of prematurity, on variability and complexity of heart rate and blood pressure

4. To characterise the cerebral autoregulation in preterm infants following caffeine therapy

5. To identify infants at risk of malnutrition and later morbidity using near-infrared (NIR) interactance signals for body composition assessment

1.3 Research Questions

The first objective relates to the following research questions:

1: What advanced analysis techniques have been applied or could be potentially relevant in the neonatal intensive care unit?

The second objective pertains to the following research questions:

2: Can fluctuation analysis and multivariable models be applied to improve the early detection of IVH in the NICU setting?

3: Can pre-processing techniques improve the performance of DFA?

The third objective relates to:

4: What are the acute affects of a loading dose of intravenous caffeine on the dynamics of heart rate and arterial blood pressure?

The fourth objective relates to:

5: Can detrended cross-correlation analysis, a relatively novel method to jointly describe the time patterns of two signals, be used to characterise cerebral autoregulation?

6: How does caffeine affect variability of cerebral autoregulation?

The fifth objective focuses on:

7: What anthropometric features are associated with neonatal malnutrition?

8: What NIR features can be used to model body composition measures of nutritional status?

1.4   Thesis Contributions

The core contribution of this thesis is demonstrating the value of advanced analyses in neonatal intensive care, from predictive monitoring applications to better characterising of physiological variability associated with specific conditions and treatments:

1. The thesis presents to the best of our knowledge, the first review on variability analysis techniques and the potential for improving outcomes specific to neonatal intensive care (Chapter 2)

2. The thesis applies DFA to arterial blood pressure and respiratory data, showing the potential of these techniques in the identification of neonates at high risk of intraventricular haemorrhage (Chapter 3)

3. The thesis also shows the value of pre-processing data and correcting, if necessary, for long-term signal drift and other non-linear trends on the impact of DFA (Chapter 3)

4. Using time series analysis of both heart rate and arterial blood pressure signals, the thesis shows altered cardiovascular dynamics in preterm infants following a standard loading dose of caffeine (Chapter 4)

5. The thesis quantifies increased pulse pressure variability following caffeine which may have implications for caffeine therapy in infants with possible impaired cerebral autoregulation (Chapter 4)

6. The thesis demonstrates the predictive value of detrended cross-correlation analysis for characterising cerebral autoregulation in potentially non-stationary time series (Chapter 5)

7. The thesis demonstrates reduced coherence corresponding and potential improved cerebral autoregulation following a loading dose of caffeine (Chapter 5)

8. The thesis develops and presents linear regression models for neonatal fat estimation as a gauge for undernutrition (Chapter 6)

9. The thesis demonstrates the potential of near-infrared interactance signals for detection of neonatal and infant malnutrition (Chapter 6)

1.5 Presentation of thesis

This thesis is presented as a combination of 1 review and 6 research papers and consists of seven chapters:

- Chapter 1 provides an overview of the aims and research questions addressed in this dissertation, followed by the list of publications.

- Chapter 2 (a published journal paper) reviews a number of techniques across both time and frequency domains for characterising variability of complex physiological systems. We present these techniques with particular emphasis on neonatal intensive care applications, outlining potential areas where these techniques have offered or may offer predictive value.

- Chapter 3 (a published journal paper) explores the utility of DFA of arterial blood pressure in the early identification of preterm infants at risk of intraventricular haemorrhage. It highlights the importance of pre-processing and multivariable models in predictive monitoring.

- Chapter 4 (a published journal paper) presents the use of non-linear variability analysis techniques including DFA and Poincare anlaysis for characterising the cardiovascular impact of caffeine in preterm infants. Caffeine is often administered in neonatal intensive care as treatment for apnoea of prematurity or to facilitate weaning off mechanical ventilation.

- Chapter 5 (a manuscript submitted for consideration) compares a range of techniques for describing cerebral autoregulation in preterm infants. The analysis focuses on their discriminative value in characterising the acute effects of caffeine therapy.

- Chapter 6 (a published journal paper and conference paper) focuses on the identification of neonates and infants at risk of later morbidity and mortality using anthropometric and NIR interactance features for body composition assessment.

- Chapter 7 summarises the research findings and discusses directions of future work.

1.6 List of Publications

The publications in the following section form the core part of this thesis.

1.6.1 Journal publications

Huvanandana, J., Thamrin, C., Tracy, M., Hinder, M., Nguyen, C. and McEwan, A., 2017. Advanced analyses of physiological signals in the neonatal intensive care unit. *Physiological Measurement*, 38:253-279

Huvanandana, J., Nguyen, C., Thamrin, C., Tracy, M., Hinder, M. and McEwan, A.L., 2017. Prediction of intraventricular haemorrhage in preterm infants using time series analysis of blood pressure and respiratory signals. *Scientific reports*, 7, p.46538.

Huvanandana, J., Carberry, A.E., Turner, R.M., Bek, E.J., Raynes-Greenow, C.H., McEwan, A.L. and Jeffery, H.E. 2018. An anthropometric approach to characterising neonatal morbidity and body composition, using air displacement plethysmography as a criterion method. *PLOS ONE*, 13, e0195193.

Huvanandana, J., Thamrin, C., McEwan, A., Hinder, M., Nguyen, C. and Tracy, M., 2018. Cardiovascular impact of intravenous caffeine in preterm infants. *Acta Paediatrica*

Huvanandana, J., Thamrin, C., McEwan, A., Hinder, M., Nguyen, C. and Tracy, M., 2018. The effect of caffeine loading on cerebral autoregulation in preterm infants **under review with Acta Paediatrica**

1.6.2 Conference publications

Huvanandana, J., Jones, P., Jeffery, H.E., Carberry, A.E., Norris, S. and McEwan, A.L., 2017, December. A near-infrared interactance model for the estimation of infant body composition. In *Life Sciences Conference (LSC), 2017 IEEE* (pp. 149-152). IEEE.

Huvanandana, J., Thamrin, C., Nguyen, C., Tracy, M., Hinder, M. and McEwan, A., 2016, October. Logistic regression models for predicting intraventricular haemorrhage in preterm infants using respiratory and blood pressure signals. In *Biomedical Engineering (BME-HUST), International Conference on* (pp. 18-22). IEEE.

Sadr, N., **Huvanandana, J.**, Nguyen, D.T., Kalra, C., McEwan, A. and de Chazal, P., 2015, September. Reducing false arrhythmia alarms in the ICU by Hilbert QRS detection. In *Computing in Cardiology Conference (CinC), 2015* (pp. 1173-1176). IEEE.

Mustafa, F.H., Jones, P.W., **Huvanandana, J.** and McEwan, A.L., 2016, October. Improvement of near infrared body fat sensing at 45-degree source-detector position angle. In *Biomedical Engineering (BME-HUST), International Conference on* (pp. 70-74). IEEE.

Tracy, M. B., Priyadarshi, A., Lowe, K., Goel, D., **Huvanandana, J.**, and Hinder, M., 2017. How do different LMA brands compare to facemask ventilation in a LMA teaching manikin?. In *Journal of Paediatrics and Child Health*, 53, pp. 43-44. (Abstract)

1.6.3   Other published work

Research work carried out during the duration of the degree but not forming a part of the thesis has been published in:

Sadr, N., **Huvanandana, J.**, Nguyen, D. T., Kalra, C., McEwan, A., and de Chazal, P., 2016. Reducing false arrhythmia alarms in the ICU using multimodal signals and robust QRS detection. *Physiological measurement*, 37(8), 1340.

Mustafa, F. H., Bek, E. J., **Huvanandana, J.**, Jones, P. W., Carberry, A. E., Jeffery, H. E., Jin, C. T. and McEwan, A. L., 2016. Length-free near infrared measurement of newborn malnutrition. *Scientific reports*, 6.

Tracy, M. B., Priyadarshi, A., Goel, D., Lowe, K., **Huvanandana, J.**, and Hinder, M., 2017. How do different brands of size 1 laryngeal mask airway compare with face mask ventilation in a dedicated laryngeal mask airway teaching manikin?. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, fetalneonatal-2017.

Befrui, N., Elsner, J., Flesser, A., **Huvanandana, J.**, Jarrousse, O., Le, T.N., Mller, M., Schulze, W.H., Taing, S. and Weidert, S., 2018. Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features. *Medical & Biological Engineering & Computing*, pp. 1-16.

Chapter 2

Literature Review

This chapter addresses the first research question:

> 1: What advanced analysis techniques have been applied or could be potentially relevant in the neonatal intensive care unit?

The content presented in this chapter is currently published as:

2.1 Advanced analyses of physiological signals in the neonatal intensive care unit

Statement of Contributions of Joint Authorship

- Jacqueline Huvanandana (Candidate): corresponding author, provided the main idea, writing, reviewing and editing of the manuscript

- Cindy Thamrin: provided the main idea, reviewing and editing the manuscript

- Mark Tracy: provided the main idea, reviewing and editing the manuscript

- Murray Hinder: reviewing and editing the manuscript

- Chinh Nguyen: reviewing and editing the manuscript

- Alistair McEwan (Principal Supervisor): provided the main idea, reviewing and editing the manuscript

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Alistair McEwan   **Date:** 24 May 2018

Topical Review

Advanced analyses of physiological signals in the neonatal intensive care unit

J Huvanandana[1] , C Thamrin[2] , M B Tracy[3,4], M Hinder[1,3], C D Nguyen[2] and A L McEwan[1,5]

[1] School of Electrical and Information Engineering, University of Sydney, Sydney, Australia
[2] Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia
[3] Westmead Hospital, Sydney, Australia
[4] School of Paediatrics and Child Health, University of Sydney, Sydney, Australia
[5] Cerebal Palsy Alliance, Australia

E-mail: j.huvanandana@gmail.com

CrossMark

Abstract

Management and monitoring of infants within the neonatal intensive care unit represents a unique challenge. It involves an array of life-threatening diseases, procedures with potentially lifelong impacts, co-morbidities associated with preterm birth and risk of infection from prolonged exposure to the hospital environment. With the integration of monitoring systems and increasing accessibility of high-resolution data, there is a growing interest in the utility of advanced data analyses in predictive monitoring and characterising patterns of disease. Such analyses may offer an opportunity to identify infants at high risk of certain conditions and to detect the onset of disease prior to manifestation of clinical signs. This allows caregivers more time to respond and mitigate any abnormal or potentially fatal changes. We review techniques for variability analysis as they have been or have the potential to be applied to neonatal intensive care, the disease conditions in which they have been tested, and technical as well as clinical challenges relevant to their application.

Keywords: signal processing, variability analysis, neonatal, intensive care, time series analysis

(Some figures may appear in colour only in the online journal)

1. Introduction

The management and monitoring of neonates within the neonatal intensive care unit (NICU) presents many unique challenges, given the prematurity of the patients admitted as well as the criticality and range of conditions observed. The human body is a complex system, comprising numerous interacting components. True to such systems, it is more than the sum of its parts (Kane and Higham 2015), i.e. the interactions between the components results in behaviour that is far more complex than expected by simply adding the behaviour of the individual components. The state of complex systems can be characterised by monitoring changes in patterns over time (Volpe 1989b, Seely and Macklem 2004). This becomes particularly relevant in monitoring critical care environments where changes are often abrupt and life threatening.

In recent years, there has been growing interest in modelling and characterising deviations from normal physiological patterns, made possible by improved capabilities to process large amounts of data. In medicine, particularly in the critical care setting, the analyses of variability and physiological patterns over time have shown promising results in predictive monitoring for critical illness and the opportunity to adjust healthcare and potentially improve outcomes and mortality (Moorman *et al* 2011 and Sullivan and Fairchild 2015). The work by Griffin and Moorman (2001) has demonstrated the potential for early detection of neonatal sepsis with the HeRO monitoring system. It uses a proprietary heart rate characteristic (HRC) index to characterise both variability and transient decelerations in heart rate which occur during systemic inflammation (Griffin *et al* 2005). The HRC index becomes increasingly abnormal prior to abrupt clinical deterioration or other recognisable clinical indicators suggesting sepsis (Griffin and Moorman 2001). In a randomised trial across 9 neonatal intensive care units, Moorman *et al* reported a reduction in mortality by more than 20% compared to settings where this information was not displayed (Moorman *et al* 2011). Such an approach to monitoring offers the opportunity to precede abrupt or catastrophic deterioration which may be life-threatening, with conventional diagnosis of sepsis often not confirmed until after significant haemodynamic compromise (Sullivan and Fairchild 2015).

There have been a number of recent reviews in this area. These range from those with a focus on the variability analysis techniques available and their underlying assumptions (Seely and Macklem 2004, Bravi *et al* 2011) to the range of applications in clinical settings, including but not limited to predictive monitoring, early detection and improving clinical outcomes (Ahmad *et al* 2009b, ChuDuc *et al* 2013, Billman *et al* 2015, Sullivan and Fairchild 2015). The majority of applications reviewed pertain to the adult population, following successful prediction results in mortality after myocardial infarction (Voss *et al* 1996, Schmidt *et al* 1999) and the onset of sepsis in adult bone marrow transplant patients (Ahmad *et al* 2009a, 2009b). This paper offers an overview of variability analysis as applied to the neonatal population, or more specifically neonatal intensive care, taking into account its unique challenges and conditions. This follows recent promising results in mortality reduction (Moorman *et al* 2011), the clinical interest in identifying patterns of critical illness and the potential to respond in a more timely manner to critical conditions that have long-term, even lifelong impacts.

1.1. The neonate

The neonatal period refers to the first 4 weeks of life following term birth delivery at 40 weeks gestation (Robertson and Rennie 1992). Survival is possible with extreme prematurity as early as 23 weeks gestation, and time between neonatal admission and discharge can be as long as 17 to 20 weeks in total. For preterm infants, this period is particularly critical; many are admitted to the NICU with congenital abnormalities and other conditions such as patent

ductus arteriosus (PDA) or may develop an array of conditions ranging from sepsis, necrotising enterocolitis (NEC) to intraventricular haemorrhage (IVH), where treatment, interventions and management can have significant and long-term impacts. Table 1 provides an overview of certain NICU-relevant conditions and their current diagnosis. To further complicate their management, these conditions may occur concurrently and exhibit similar clinical signs, with diagnosis often taking place after significant deterioration or compromise. They are often interrelated with factors leading to preterm birth (Villar *et al* 2006) such as infection and poor placental function (acute = asphyxia, chronic = intrauterine growth restriction). Importantly, lifelong injury of tissues such as the retina, brain, kidneys and gastro-intestinal can be related in part or wholly to recurrent pathological deviations in the physiological state in preterm newborns. Recurrent hypocarbia can be associated in a dose dependent manner with severity of brain damage (Collins *et al* 2001, Erickson *et al* 2002). There is thus difficulty in characterising the precise and long-term impact of a patient's stay in the NICU, where assessment of developmental outcomes, from physical to neurological, require longer term studies.

1.2. The neonatal intensive care environment

The NICU provides life-saving support for newborns born prematurely. The length of stay in the NICU varies with survival, gestational age, birthweight, intrauterine growth restriction and use of antenatal steroids (Lee *et al* 2013). Over this time, management and monitoring of the neonate in this environment presents a combination of challenges, from the need for intubation and mechanical ventilation to intravenous catheterisation—all the while, risk of infection increases with every day of exposure to the hospital environment (Taeusch *et al* 2005). Administration of care in and the introduction of new systems to this environment requires consideration of noise and light exposure levels as well as the existing framework for monitoring and life-saving support. It also presents the opportunity to initiate individualised treatment. A recent report from Moody *et al* demonstrated the potential for reducing length of stay in the NICU through early initiation of developmental care (Moody *et al* 2017).

1.3. Current monitoring approaches and limitations

Physiological parameters monitored in these units range from oxygen saturation (SpO_2) using pulse oximetry, heart rate derived from temporal distances between consecutive R peaks in the electrocardiogram (ECG) signal, and arterial blood pressure which is measured invasively using an umbilical or peripheral arterial catheter. Other parameters include cerebral oximetry which is monitored using near-infrared spectroscopy and respiration rate which can be derived from chest electrical impedance monitoring.

Conventional monitoring of these physiological parameters involves a limit-based approach, where deviations from the set thresholds may trigger alarms to which healthcare providers can respond. This approach requires patient-customisation for effective use, such as gestational age and birthweight-adjusted limits for arterial blood pressure. Even with appropriate limits set, false alarms persist. An excess of alarms contributes to desensitisation or alarm fatigue, especially when a high proportion of these may be false or not clinically important. Examples include those triggered by instances of excessive tidal volume from deep spontaneous breaths during weaning from mechanical ventilation or by motion and other artefacts (Imhoff and Kuhls 2006). Addressing alarm overload may also come with a trade-off, with McClure *et al* showing that while longer time windows for averaging SpO_2 signals may reduce the number of alarms, it may underrepresent the number and severity of events (McClure *et al* 2016).

Table 1. Overview of prevalent conditions, their incidence rates and the potential for early prediction in the NICU. Signals of use include electrocardiogram (ECG), electroencephalography (EEG), arterial blood pressure (ABP) and tissue oxygenation index (TOI).

Clinical conditions	Definition	Current diagnosis	Clinical signs	Incidence rates	Signals of use	Interventions initiated at diagnosis
Late-onset sepsis	Systemic infection occurring after 3 d of life (Boghossian et al 2013)	Positive blood cultures	Frequent apnoea episodes, increase in desaturations and apnoea episodes, temperature instability, lethargy, feeding intolerance as well as abdominal distension. Hypotension and/or shock, acidosis, multi-system organ failure.	Incidence ranges from 0–30% (Boghossian et al 2013, Lahra et al 2009, Tröger et al 2014) and is inversely proportionate to birthweight (Stoll et al 2002)	ECG, ABP and pulse oximetry as measures for shock (Bohanon et al 2015, Funk et al 2009)	Antibiotics, haemodynamic support
Intraventricular haemorrhange (IVH)	Bleeding, originating from the germinal matrix and possibly extending to the ventricular area	Cranial ultrasound	Falling haemoglobin, shock, bleeding disorders, seizures, apnoea bradycardia, pupillary and cranial nerve abnormalities, changes in levels of consciousness, movement or tone.	Over 20% in infants <30 weeks gestation Heuchan et al (2002)	ECG, ABP, TOI, respiratory signals	Haemodynamic and respiratory support, monitoring for hydrocephalus, supportive care

(*Continued*)

Table 1. (*Continued*)

Clinical conditions	Definition	Current diagnosis	Clinical signs	Incidence rates	Signals of use	Interventions initiated at diagnosis
Necrotising enterocolitis (NEC)	Disease of the gastrointestinal tract, possible mucosal injury	Abdominal radiography and ultrasound	Frequent apnoea episodes, increase in desaturations and apnoea episodes, temperature instability, lethargy, feeding intolerance as well as abdominal distension (Bell *et al* 1978). Blood-stained or bilious gastric aspirates, blood in stools, pneumatosis intestinalis.	1–5% of all newborns admitted to the NICU (Lin and Stoll 2006), 7–14% of very low birthweight (<1500 g) infants (Luig and Lui 2005, Neu *et al* 2008)	ECG, ABP and pulse oximetry	Antibiotics, bowel rest and decompression
Seizures	Abnormal central nervous system discharge	EEG	May involve random eye movements, thrashing, apnoea, stiffening of the muscles	1–3.5 in 1000 births for term neonates (Lanska *et al* 1995, Saliba *et al* 1999, Ronen *et al* 1999)	EEG, ECG, respiratory signals	Neuroimaging, administration of anticonvulsant agents (Glass *et al* 2012)
Patent ductus arteriosus (PDA)	Persistence of the ductus arteriosus connecting systemic and pulmonary circulation	Doppler echocardiography	Widened pulse pressure, failure to wean ventilator pressures, long systolic murmurs and distinct peripheral pulses (Evans *et al* 2004)	30% of very low birthweight infants (Lemons *et al* 2001)	ABP	Intravenous ibuprofen or indomethacin (Ohlsson *et al* 2013)

Furthermore, abnormal signals may fail to trigger an alarm and/or a response, as demonstrated by an analysis of nurse-recorded as well as monitor and algorithm detected apnoea events. Of the algorithm-detected prolonged apnoea events, only 26% had nurse documentation within 1 hr and only 23% activated the monitor's apnoea alarm (Vergales *et al* 2014). There is also an under-utilised potential to display trends in the data over various timescales and make this available for review, which is highlighted by the more recent promising work in predictive monitoring (Fairchild *et al* 2013, Sullivan and Fairchild 2015).

2. Variability analysis techniques

The prevailing notion in variability analysis of biological signals is that illness itself usually causes a reduction in system variability Goldberger (1997), though the opposite can also occur. The available methods for this can be broadly classified into time domain, frequency domain or a combination of the two.

2.1. Time domain analysis

These are measures which describe the signals of interest in terms of their variation over time. One example are moments, which are statistical quantities that characterise the variability of a series: $x_1, x_2, x_3...x_n$. The *rth* sample moment is denoted by $\mu_n = \frac{1}{n} \sum_{i=1}^{n} x_i^r$. For a mean-centred series, the moments from $n = 2$ onwards are referred to as central moments, with the second central moment denoting variance, the third skewness, and the fourth kurtosis. Standard deviation (SD), that is the square root of the variance, is often used to quantify variability. Like the mean, these features are most informative for Gaussian-distributed data, where cases which deviate from this may not be discerned or characterised based on time features alone (Stanley *et al* 1999).

In relation to HRV, these parameters are the basis for calculating SDANN, the standard deviation of average NN (normal sinus to normal sinus) interbeat intervals of an ECG calculated for 5 min segments across a 24 h recording, and RMSSD which represents the root mean square of differences between adjacent NN intervals (Shaffer *et al* 2014).

The Poincaré plot is derived from the time domain signal, whereupon a signal is plotted against a version of itself shifted by a time log. It offers a visual representation of self-similarity, where reduced variability manifests as increased elongation of the ellipse fitted to the data points. This can also be characterised using SD1 and SD2, which represent the standard deviation perpendicular and parallel to the line of identity. These two parameters have been shown to be highly correlated with other statistics such as SDANN (Brennan *et al* 2001).

Using HRV as an example, with ECG derived RR_n on the *x*-axis and RR_{n+1} on the *y*-axis, figure 1 displays examples of regular (a) and irregular (b) neonatal HRV, respectively. Note that figures 1 and 2 were generated from data collected as part of a study approved by the Sydney West Area Health Service Human Research and Ethics and conducted according to the World Medical Association Declaration of Helsinki.

2.2. Frequency/time-frequency domain analysis

Transformations, that is the reproducible and reversible mapping of one set of values to another using a function, is at the core of frequency or time-frequency domain analysis. The Fourier transform represents one such function, simplifying a given time series as sinusoidal waves of varying periods superimposed on one another. This is often applied as the short-time
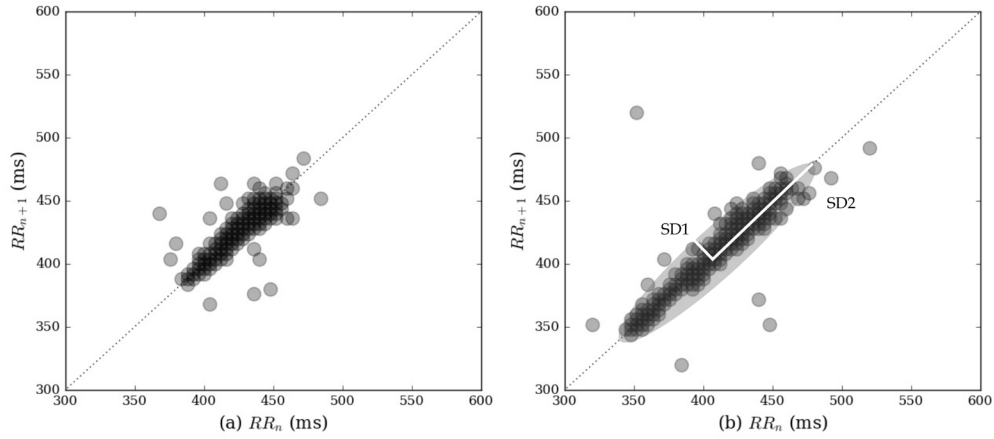
(a) RR_n (ms)   (b) RR_n (ms)

Figure 1. Poincaré plots showing heart rate variability from two excerpts of neonatal heart rate with low (a) and higher (b) levels of regularity. SD1 and SD2 are defined as the standard deviation perpendicular to and along the line of identity. For an ellipse fitted to approximately 95% of the points in along both of these axes, SD1 and SD2 are one half the minor and major axes, respectively. The ratio of SD1/SD2 is thus equivalent to the ratio of minor and major elliptical axes.

(a) 4 minute segment (μ, σ)

(b) 2 minute segment (μ, σ)

(c) 1 minute segment (μ, σ)
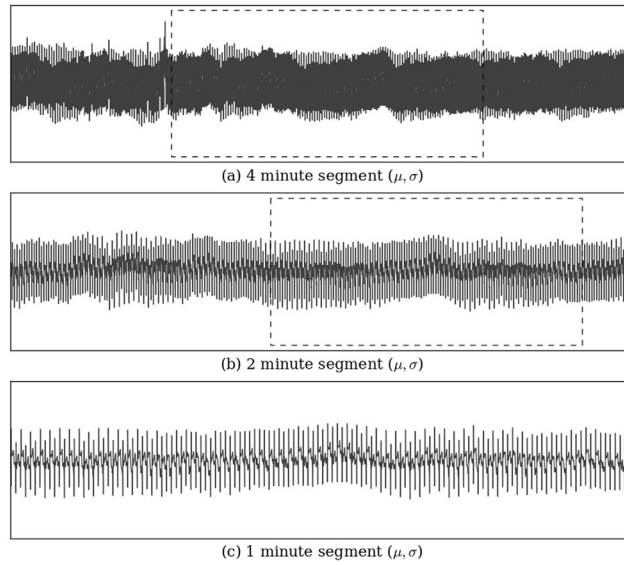
Figure 2. Statistical self-similarity over various time scales. The subplots (a) to (c) show a neonatal electrocardiogram signal over 4, 2 and 1 minute segments, respectively. The means (μ), standard deviations (σ) and relative intervals between R peaks appear similar regardless of the observed time scale.

Fourier transform where this transform is applied to the window of interest. Most frequency-domain features involve the total power over a given frequency band (a–b Hz) and denoted by

$$\int_a^b [F(t)]^2 \mathrm{d}t \tag{1}$$

R259

17

Figure 3. Application of detrended fluctuation analysis, showing the divisions of the integrated mean-centred signal into var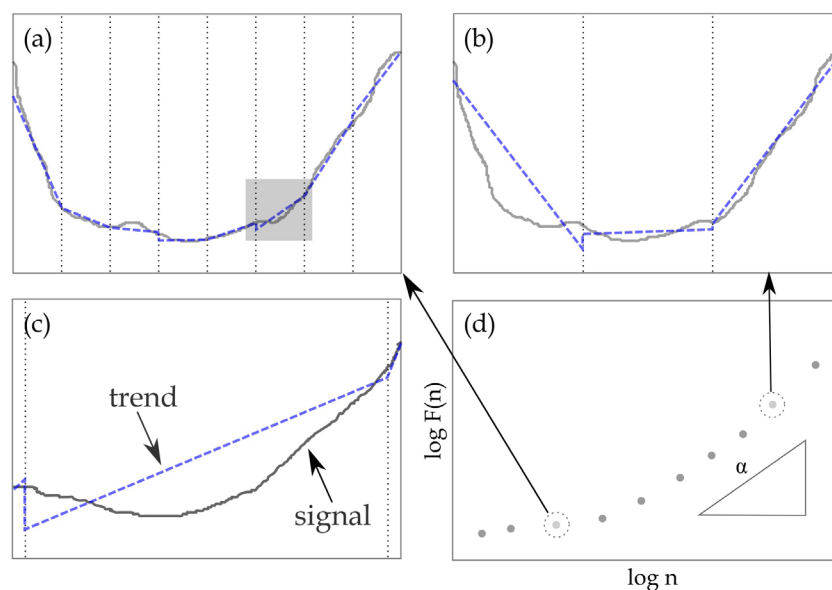ious box sizes (a) and (b). (c) shows the linear trend determined for a single box, which is subsequenty removed to determine the root mean-squared fluctuation. (d) shows the logarithmic plot of this fluctuation F(n) against the box size n, the slope of which is defined as the scaling exponent α.

where F(t) represents the fast Fourier transform (FFT). The validity of this approach rests on assumptions of periodicity and stationarity (Seely and Macklem 2004), that is, where statistical properties such as mean and SD remain constant over time. The latter description is not often accurate for physiological data such as heart rate and blood pressure signals, which may exhibit local or intermittent fluctuations.

The wavelet transform is another such function, where the time series is expressed in terms of correlation at varying scales and times to a reference signal or mother wavelet. The process is described in further detail elsewhere (Ivanov *et al* 1996, Torrence and Compo 1998), though can also be used to define similar components over a range of frequencies.

These techniques assume stationarity and periodicity of the signal (Mansier *et al* 1996) and are sensitive to artefacts, introducing bias in the calculated low frequency or high frequency components. The ranges are also not definitive and may vary depending on the individuals and other factors (Furlan *et al* 1990).

2.3. Fractal analysis and power law behaviour

Fractal analysis deals specifically with self-similarity across various scales. This can manifest geometrically, as observed in many biological structures such as in the human airways (ER 1962, Horsfield and Cumming 1968), though can also manifest as statistically identical properties over time (Thamrin and Stern 2010). Any given section of a self-similar signal, when scaled to the original signal, would exhibit an identical mean and standard deviation, as demonstrated in figure 2.

Developed by Peng and co-workers, detrended fluctuation analysis (DFA) is a method used to quantify long range power-law correlations (Peng *et al* 1995). It has since been applied to
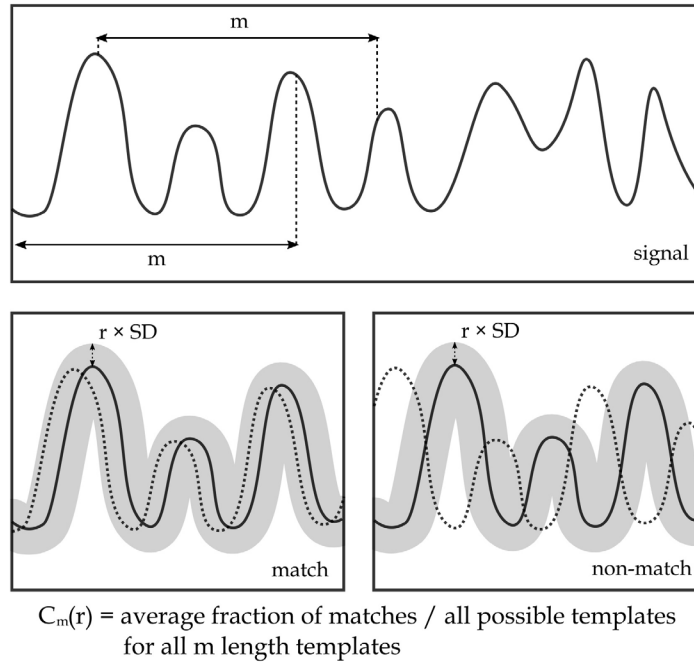
18

$C_m(r)$ = average fraction of matches / all possible templates
for all m length templates

Figure 4. Illustration of approximate entropy matches and non-matches. A template pattern is defined according to pattern length *m* which is defined by continually shifting along the window of analysis. For each of these templates, matches between the template and signal are counted when differences between the two do not exceed the defined tolerance *r*. This parameter *r* is often expressed as a function of the standard deviation. Once the fraction of matches from all possible windows of analyses is determined for each *m*-length template, C_m represents the mean of these fractions. This process is then repeated for a template of pattern length $m + 1$, and the approximate entropy determined from the natural logarithm of the ratio between C_m and C_{m+1}.

various clinical contexts, given its capacity to mitigate the effects of non-stationarities. This view has developed since the introduction of this technique, with studies to caution the risk of introducing bias in the presence of non-linear trends as well as the need for appropriate pre-processing (Bryce and Sprague 2012). Its application also requires a substantial number of data points (Seely and Macklem 2004).

The application of DFA involves several steps for a given time series *x(k)*:

(i) Integration of the mean-centred signal

$$y(k) = \sum_{j=1}^{k} x(j) - \bar{x} \tag{2}$$

(ii) Division of the integrated signal into boxes of equal length *n* and each box is detrended by subtracting the local linear trend

(iii) The root mean square (RMS) fluctuation of the time series is determined using the following expression

$$F(n) = \left(\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2 \right)^{1/2} \tag{3}$$

where $y_n(k)$ is the local linear trend per box of length n

(iv) The process detailed in (2) to (3) is repeated for varying box sizes n

 (v) A linear trend is fit to the log-log relationship between box size n and the RMS fluctuation $F(n)$ to determine scaling exponent α

$$F(n) = An^{\alpha}. \tag{4}$$

This process is illustrated in figure 3. The scaling exponent α is thought to represent stability in the system and is thus deranged in disease. There is some suggestion that either a low or high α value may represent pathology (Rojo-Álvarez *et al* 2007), though clinical interpretation may be difficult, especially in attributing it to specific pathophysiological causes (Glass 1999). Power laws governing the distribution of inter-breath intervals, which are related to the scaling exponent α, have been associated with maturation in preterm and term infants (Frey *et al* 1998).

2.4. Approximate and sample entropy

We can also approach variability characterisation using entropy or the rate of information generated. The work of Grassberger and Procaccia (Grassberger and Procaccia 1983) on the extraction of correlation exponent v from time series data served as a precursor for subsequent entropy measures (Eckmann and Ruelle 1985). With respect to time series analysis of physiological signals, it is commonly used in the form of approximate entropy (ApEn) or sample entropy (SampEn).

ApEn was originally presented by Pincus and is a measure of logarithmic likelihood that similar patterns will be further followed by similar patterns (Pincus 1991). A time series with a low ApEn is often interpreted as having predictable or deterministic fluctuations. This statistic denotes the conditional probability that a given template pattern of length m, will not be followed by a similar pattern in the proceeding time epoch, within a tolerance of r. This is expressed as a fraction of the SD (Pincus 1991), as illustrated in figure 4. Discussions surrounding its use and reliability have targeted the inherent bias towards regularity as it takes into account self-matches (to avoid the possibility of $\ln(0)$) (Richman and Moorman 2000).

ApEn is defined according to the following equation:

$$\text{ApEn} = \ln \frac{C_m(r)}{C_{m+1}(r)} \tag{5}$$

where C_m represents the mean fraction of matches for all possible templates of length m.

In addressing the limitations of ApEn, SampEn was developed, denoting the likelihood that two similar sequences for a given window of length m, remain similar at the proceeding time epoch, within a tolerance of r (Richman and Moorman 2000). The definition of SampEn also displays relative consistency. Put simply, where one record has a lower SampEn than another, this observation will be consistent for any given definition of m and r (Lake *et al* 2002). The application of these statistics assume stationarity (Pincus and Singer 1996) and thus requires appropriate interpretation. Reduced entropy estimates were largely due to the presence of non-stationarities (such as spikes in the data, inflating the SD and thus the tolerance for identifying matches) and not necessarily an indication of underlying complexity changes (Moorman *et al* 2006, 2011). Analysis is also affected by parameter selection and data length.

2.5. Cross-correlation and multi-signal approaches

The above techniques can also be extended to incorporate multiple sources of information in a single feature. Linear dependency between two simultaneously measured time series can be quantified with the linear correlation coefficient, r with bounds of -1 and 1, representing perfect negative and positive correlation, respectively. Running correlation coefficients can also be calculated for short window definitions, forming the basis for studies on synchronisation and functional connectivity. Cardiorespiratory synchronisation or coupling represents a more signal-specific approach, where a parameter λ indicates the strength of synchronisation between heart rate and respiratory signals. This was found to characterise early changes in breathing control in infants (Nguyen *et al* 2012).

Existing variability statistics such as ApEn and SampEn can also be extended to account for multiple signals, as in their cross-entropy counterparts, cross-ApEn and cross-SampEn. These were introduced to quantify the degree of asynchrony of two time series, where the more asynchronous the series, the higher the statistics (Pincus and Singer 1996, Pincus *et al* 1996, Richman and Moorman 2000). More complex measures include multiscale entropy (Costa *et al* 2002, 2005) and multiscale cross-sample entropy (Xia and Shang 2012). Interpretation of these measures needs to account for underlying dynamics and the potential influence of sampling time and selected time scales (Thuraisingham and Gottwald 2006).

The detrended cross-correlation coefficient, ρDCCA (Zebende 2011), is the fractal analysis counterpart to the linear correlation coefficient that can quantify correlations in non-stationary signals (Kristoufek 2014). Extensions of this work, such as the generalised detrended cross-correlation coefficient, have been described elsewhere (Oświęcimka *et al* 2014, Kwapień *et al* 2015, Shen *et al* 2015).

2.6. Considerations in technique selection and data collection

Technique selection ought to take into consideration the underlying assumptions of each technique, data requirements (how many data points are required, how large a dataset is required) and the impact of deviations from these assumptions. Most of the techniques described above depend strongly on assumptions of stationarity, being susceptible to bias when this does not hold. An exception to this is DFA, which was originally presented as a means of addressing non-stationarities (Peng *et al* 1995), mainly through the linear detrending step. While this may be true of certain non-stationarities, there is also the potential to introduce bias in the presence of non-linear trends (Bryce and Sprague 2012). Other assumptions include those of periodicity, permitting valid transformation of a signal to the frequency domain.

Signal availability and the nature of their collection may also impose restrictions on the techniques applied. This is certainly the case for EEG, which may not be routinely monitored in the NICU environment, as well as for ECG, where sufficient separation of the electrodes may not be possible in a small-sized preterm infant, given other monitoring devices and potential irritation (Baird *et al* 1992). Blood pressure can be particularly helpful, though its continuous measurement requires an invasive (umbilical or peripheral) arterial catheter. Temperature measurements, though continuously available, are not necessarily useful as deviations from the normal range may be masked by temperature-controlled incubator environments. The routine collection of oxygen saturation data from pulse oximetry and its signal availability renders it a good candidate for analysis, though acute changes in oxygen saturation have not been specifically investigated in the context of neonatal conditions such as sepsis or NEC (Sullivan and Fairchild 2015).

The use of plethysmography-derived signals within the NICU context has been limited predominantly to time domain techniques, though with varying degrees of effectiveness. Renner *et al* evaluated the usefulness of respiratory variations in velocity time integral, peak flow velocity as well as the plethysmography variability index (Cannesson *et al* 2008) for predicting fluid responsiveness in infants (Renner *et al* 2011). There is also potential for exploring features extracted from the raw plethysmography signal, such as respiration rate or those pertaining to cardiovascular and neural fluctuations. The fluctuations from each manifest in this plethysmography signal at different frequencies, allowing each to be distinguished by their sinusoidal components (Nilsson 2013). This would be particularly beneficial for cases where ECG signals may not be available.

The presence of artefact may also confound clinical interpretation, from movement of the patient, mains contamination to the detachment of electrodes. Signal quality measures can thus help to contextualise the interpretation of these analyses. These can be applied in the form of a threshold for analysis of the selected window or as a means of adjusting the probability of prediction, as was used in a neonatal apnoea detection algorithm (Monasterio *et al* 2012). Pre-processing may also play a key role, where removal of cardiac artefact from the CI signal improved neonatal apnoea detection (Lee *et al* 2011) and the removal of non-linear trends facilitated valid interpretation of fluctuation analyses (Bryce and Sprague 2012). Removal of baseline wander from ECG similarly allows for accurate heart rate acquisition.

Nyquist's theorem stipulates that the sampling frequency must be at least 2 times the highest frequency component of the signal, though the sampling frequency required for accurate and non-biased variability analysis may be higher. Motivated by computational efficiency, which is especially pertinent for real-time applications, a considerable number of studies have investigated the influence of sampling frequency on variability analysis (Merri *et al* 1990, Garcia-Gonzalez *et al* 2004, Ziemssen *et al* 2008, Bhatia *et al* 2010, Ellis *et al* 2015, Choi and Shin 2017). Some have proposed analysis-driven minimum bounds for sampling frequency, such as the >25 Hz for pulse rate analysis from photoplethysmographic signals (Choi and Shin 2017), >100 Hz for frequency-domain blood pressure variability (Bhatia *et al* 2010) and the caution against sampling at <12 Hz for HRV spectral analysis (Garcia-Gonzalez *et al* 2004).

3. Clinical applications

Clinical applications of variability analysis are often aimed at predicting the onset of critical conditions, leveraging technology and signal processing to assist healthcare providers (Griffin *et al* 2005) as well as prompting further bedside evaluation and consideration of appropriate tests or treatments (Fairchild *et al* 2013). A summary of these conditions and their current incidence rates is provided in table 1.

The above techniques have been applied to a range of neonatal intensive care applications, from HRV analysis for the prediction of sepsis (Griffin and Moorman 2001, Griffin *et al* 2003, Moorman *et al* 2006, Sullivan and Fairchild 2015) to early developments in IVH and mortality risk prediction (Gilmore *et al* 2011, da Costa *et al* 2015) to ECG and EEG analysis for seizure detection (Greene *et al* 2007, Greene de Chazal *et al* 2007). The following sections offer an overview of variability analysis that has been applied in the context of the NICU-relevant conditions summarised in table 1.

3.1. Neonatal sepsis and necrotising enterocolitis

Neonatal sepsis is a systemic infection which remains an important cause of morbidity and mortality in preterm infants (Stoll *et al* 2002). Prolonged hospitalisation may predispose

22

preterm infants to late-onset sepsis due to indwelling vascular catheters and incompletely developed immunity (Moorman *et al* 2011). Late-onset sepsis, hereinafter sepsis, is characterised by an onset of at least 3 d following birth, often originating from the hospital or NICU environment (Boghossian *et al* 2013). Risk of sepsis increases with decreasing birthweight and gestational age (Stoll *et al* 2002) and is associated with PDA, prolonged mechanical ventilation, NEC and bronchopulmonary dysplasia (Neu *et al* 2008).

It is diagnosed via blood culture, though this may be limited by the low sampling volume ($<0.5-1$ ml) for neonates (Goldstein *et al* 2005, Connell *et al* 2007) and has been associated with false positives and negatives (Volante *et al* 2004, Dong and Speer 2014). Diagnosis may also not occur until after substantial deterioration or haemodynamic compromise (Sullivan and Fairchild 2015). Antibiotics are generally administered in response to confirmed sepsis or suspected sepsis on the basis of concerning signs such as increased apnoea episodes, inactivity or feeding intolerance (Sullivan and Fairchild 2015). These signs are neither very sensitive nor specific to the disease, overlapping with other conditions such as NEC.

Necrotising enterocolitis (NEC) affects the gastrointestinal tract, occurring most commonly in the small and large bowel (Henry and Moss 2009). In preterm infants, bacterial colonisation of the gut is abnormal and often delayed, which, in combination with external stressors and ischemic or hypoxic insults, may lead to mucosal injury. This may also be exacerbated by the invasion of certain pathogens, amplifying the inflammatory response (Papillon *et al* 2013). Key risk factors include prematurity and formula feeding though its pathogenesis is not well-understood (Caplan 2008, Neu *et al* 2008). Clinical stages can be defined according to Bell's criteria (I to III), with the latter two stages involving significant intestinal distension and ileus as confirmed via abdominal radiography (Bell *et al* 1978).

The premise for the early detection or prior to manifest clinical signs lies in the elevated levels of the cytokines which initiate sepsis or sepsis-like illnesses. These were reported by Kuster *et al* to be present up to 2 days preceding clinical diagnosis (Küster *et al* 1998). The general trend observed in these disease states has been a reduction in regularity, a term used interchangeably with complexity. Lake *et al* observed a reduction of SampEn in the lead up to clinical diagnosis of neonatal sepsis, while acknowledging its susceptibility to noisy or large spkies in the data (Lake *et al* 2002). In response to this, it was suggested that regularity based on the SampEn statistic requires interpretation alongside additional measures of complexity and non-stationarity (Lake *et al* 2002).

Griffin and Moorman observed a reduction in HRV and transient decelerations in heart rate (HR) in association with neonatal sepsis and sepsis-like illnesses which preceded clinical signs and abrupt deterioration. In some cases, this lead time was as much as 24 h (Griffin *et al* 2005), motivating further exploration of this approach. An earlier publication of this work explored the use of time domain techniques such as the skewness of the asymmetrical RR interval histograms in discerning sepsis cases. A multicentre randomised trial using a novel HRC monitoring system was conducted, detecting both variability and transient decelerations in heart rate. Moorman *et al* reported a significant reduction in mortality rates from 10.2% to 8.1% when this variability information was displayed alongside other signals (Moorman *et al* 2011). A retrospective review of this study was also conducted on NEC cases in a subset of infants, with reports of significant increases in the HRC index prior to clinical diagnosis of NEC (Stone *et al* 2013). This elevation from the patient-specific baselines were observed from 16 hours prior to diagnosis of NEC cases requiring surgical intervention and from 6 hours for the remaining cases (Stone *et al* 2013).

Recent work in this area has also demonstrated the potential utility of cross-correlation of vital signs in early detection of sepsis and NEC (Fairchild *et al* 2017). The cross-correlation between HR and SpO_2 was one of the best independent predictors of sepsis and NEC,

increasing prior to diagnosis. This increase was also reported to be greater in NEC cases than in those with sepsis. The reported model including this statistic, was combined with HRC index from the HeRO monitor and found to significantly improve predictive ability (Fairchild *et al* 2017).

3.2. Intraventricular haemorrhage

Intraventricular haemorrhage (IVH) stands as the most prevalent intracranial haemorrhage in preterm neonates and remains a major concern in extremely preterm infants. The severity of the haemorrhage is defined by Grades I to IV of the Papile system, reflecting whether the area of bleeding is confined to the subependymal germinal matrix (I) or extends to the ventricular system (II to IV) (Papile *et al* 1978). Along with IVH, periventricular haemorrhagic infarction (PVHI) and cerebellar infarction are also important conditions for neurodevelopmental prognosis. With increasing survival of infants born preterm, there is a growing number of those at risk of permanent brain injury, cerebral palsy or adverse neurodevelopmental outcomes (Volpe 2008, 2009).

The preferred method for imaging the neonatal brain is cranial ultrasonography (van Wezel-Meijler *et al* 2010); it is highly sensitive in the detection of IVH and/or PVHI and cerebellar infarction. It allows the location and extent of the injury to be visualised (Volpe 1989a). Currently, no treatment for IVH exists, though studies identifying risk factors and pathophysiology help to guide clinical care for prevention. One such study identified a strong association between improvements in blood pressure and increased in low-frequency variability in blood pressure within infants exposed to inotropes, often administered to improve mean arterial blood pressure (Vesoulis *et al* 2017). This variability was quantified using spectral analysis of certain low-frequency (0.005–0.16 Hz) components of the blood pressure signal and may be used as an adjunctive measure of response to these medications and potentially mitigating risk of hypertension-induced IVH (Bel *et al* 1987) through more optimal titration (Vesoulis *et al* 2017).

Pressure passivity has been observed in cases of intraventricular haemorrhage, where the cerebral circulation is affected by fluctuations in system circulation. This autoregulation was quantified by Gilmore *et al* (2011) using COx, a cerebral oximetry index calculated from a running window correlation of between the arterial blood pressure (ABP) and cortical reflectance oximetry. A similar approach was employed by da Costa *et al* (2015) in their definition of TOHRx, the running correlation between the tissue oxygenation index (TOI) as measured from near-infrared spectroscopy and heart rate (HR). Reduced complexity, as characterised by SampEn calculations over multiple time scales (multiscale entropy) was also observed in preterm infants who subsequently developed IVH (Sortica da Costa *et al* 2017).

In retrospective analyses of premature infants following IVH, altered autonomic functions have been observed (van Ravenswaaij-Arts *et al* 1991, Hanna *et al* 2000), as reflected by HRV analysis. The potential for predicting IVH has been explored using DFA to quantify the fractal dynamics of heart rate (Tuzcu *et al* 2009) as well as beat-to-beat mean arterial and systolic blood pressure (Zhang *et al* 2013). Both reported on increased short-term scaling exponent α_1 in infants who subsequently developed IVH.

The impact of impaired cerebral autoregulation associated with IVH and long-term neurodevelopmental outcomes has been reported in literature (Futagi *et al* 2006, Bolisetty *et al* 2013). The dynamic nature and multiple-time-scale (MTS) properties (Zhang *et al* 1998, 2000) of cebral autoregulation has recently been characterised using MTS correlation analysis of the mean arterial pressure and regional cerebral tissue oxygen saturation measured using

R266

24

NIRS. Chalak *et al* reported associations of both in-phase short-term (15 mins) and long-term antiphase (4 h) time scales with abnormal neurodevelopmental outcomes in a small number (*n* = 5) of newborns (Chalak *et al* 2016).

3.3. Patent ductus arteriosus

Patent ductus arteriosus (PDA) is characterised by persistence of the ductus, resulting in shunting from the systemic to the pulmonary circulation. Though this is normal at birth, functional closure of the foetal ductus generally occurs within 24 to 48 h of birth (Schneider and Moore 2006). Diagnosis is usually through cardiac ultrasound, allowing the degree of constriction and nature of the shunt to be characterised. Observational studies of PDA have shown a consistent association with various adverse outcomes such as NEC and IVH. Other studies have also reported on associations with neonatal chronic lung disease and death (Evans and Kluckow 1996, Kluckow and Evans 2000, Noori *et al* 2009).

Several studies have been focused on how and when to treat PDA (Laughon *et al* 2004, Benitz 2010, Evans 2015), more so than its effect on physiological parameters such as HRV. Although the overarching question of treatment continues, variability analysis of PDA-specific data may help to characterise other conditions with which it is correlated.

3.4. Neonatal apnoea and seizures

Apnoea, the cessation of breathing, is highly prevalent among very low birthweight infants (<1500 g) and often accompanied by bradycardia (HR < 100 beats per minute) and/or oxygen desaturation (SpO_2 < 80%). They are considered clinical events if cessation of breathing exceeds 20 s or 10 s in the presence of one of the above (Finer *et al* 2006). An assessment of central apnoea in preterm infants supported the inverse relationship between apnoea events and gestational and postmenstrual age (Fairchild *et al* 2016). This study also reported an increase in apnoea events in the presence of bradycardia and desaturation prior to diagnosis of sepsis and NEC. Interestingly, such events were not more frequent in infants with IVH after adjustment for gestational age (Fairchild *et al* 2016).

Two main signals pertinent for neonatal apnoea detection are the chest impedance (CI) signal and the pulse-oximeter derived signal for oxygen saturation, with previous studies exploring the relationship between desaturation and apnoea (Haider *et al* 1995, 1996). Effective detection of apnoea based on respiration rate has often been confounded by inaccuracy of the CI signal, the presence of cardiac artefact and the fact that many infants are on mechanical ventilation (Lee *et al* 2011, Sullivan and Fairchild 2015). It thus depends on the capacity to identify and filter out these artefacts as well as the algorithm for classification of relevant events. Validation of one such algorithm showed over 90% agreement with clinical experts (Lee *et al* 2011). This approach involved both the CI and ECG signals, where cardiac artefact was filtered from the CI interval and residual fluctuations were subsequently analysed using a running standard deviation (SD). Cessation of breathing was identified in excerpts below a defined variance threshold (Lee *et al* 2011). Reduced HRV was also reported for preterm infants with apnoea of prematurity compared with term neonates (Henslee *et al* 1997).

Seizures frequently occur within the neonatal period and are often characterised by rapid changes in heart rate, respiration rate and blood pressure (Bassan *et al* 2008). Other subtle seizure phenomena can include altered behaviour, motor manifestations such as peculiar limb movements and apnoea. Although there have been small-scale studies on the detection of seizures based on variability, specifically using heart rate based features, the precise nature of these effects have not been clarified. Vaugh *et al* observed a reduction in mean heart rate

during seizure (Vaughn *et al* 1995), while the converse was reported by Greene *et al* in their investigations (Greene de Chazal *et al* 2007). The approaches to ECG-based seizure detection have utilised features spanning multiple domains (Greene de Chazal *et al* 2007, Malarvili and Mesbah 2009). These included the conventional time domain features such as mean, SD, coefficients of variation and Hjorth parameters for quantifying activity, mobility and complexity (Hjorth 1973, Malarvili and Mesbah 2009, Oh *et al* 2014). Various frequency domain coefficients of power spectral density were also extracted (Greene *et al* 2008, Temko *et al* 2011).

Although EEG-based approaches have been explored for adult seizures (Gardner *et al* 2006, Srinivasan *et al* 2007), similar studies have only recently been conducted on the neonatal population, with EEG not routinely monitored in the NICU. An example of this is the early work by Lommen *et al* in developing an algorithm for automated neonatal seizure detection. They used frequency analysis for identifying artefacts and reported a rise in the lower boundary of the amplitude-integrated EEG signal during seizures in the studied newborns ($n = 13$) (Lommen *et al* 2007). Continuous changes inherent in the developing brain may render EEG results difficult to interpret at this stage and there are additional technical considerations when recording from a neonatal (therefore, small) scalp, ranging from skin resistance to the pathologically-bound states of activity.

There also remains the question of treatment and early interventions following detection, with potential neurotoxic risks of commonly used seizure medications (Bittigau *et al* 2002), especially given the lack of definitive evidence linking reduced seizure burden with neurodevelopmental outcomes (Glass *et al* 2012).

4. Remaining questions and future directions

A summary of the parameters relevant for each condition and evidence for observed patterns in variability is presented in tables 1 and 2.

4.1. Multidimensional approaches for identifying key features and classifying disease

Studies and reviews on variability analysis techniques continue to emphasise the need for multiple techniques rather than a consensus on the use of a single approach (Seely and Macklem 2004). Once these individual features have been extracted, it is necessary to evaluate their relevance (feature selection) as well as consider approaches for combining this information (feature classification).

Feature selection methods can be employed to improve model performance as well as mitigate risks of overfitting to the data. These algorithms can be broadly categorised as filter, wrapper and embedded approaches which have been reviewed in detail elsewhere (Saeys *et al* 2007, Hira and Gillies 2015). Their use must also take into consideration computational efficiency, specificity to the selected classification algorithm as well as the nature and number of features extracted.

In terms of classification, approaches ranges from multivariable logistic regression models and Bayesian models to decision trees, support vector machines and artificial neural networks, the details of which have been extensively reviewed (Michie *et al* 1994, Kotsiantis *et al* 2007).

At this stage, a considerable proportion of NICU-specific variability analysis has focused on single-parameter models, with an expanding breadth of application for HRV analysis. This is particularly true of approaches to identifying sepsis, NEC and IVH cases. A natural path for development and improved robustness in this area lies in multivariate models and the combination of features which have independently been shown to be predictive of such conditions.

Table 2. Parameters and their use in intensive care and predictive monitoring. Use of various signals including arterial blood pressure (ABP), tissue oxygenation index (TOI) and electroencephalography (EEG).

Technique	Parameter	Signals applied to	Physiological interpretation; clinical significance
Simple statistical measures	Standard deviation (SD)	HRV (RR intervals)	Reduced SDRR interpreted as reduced variability of RR
	Percentage of absolute difference between consecutive RR > x ms	HRV (RR intervals)	Reduced pNNx indicates lower variability; used to monitor cardiac parasympathetic activity (Ewing et al 1984)
Hjorth parameters	Variance of the time domain or signal power in the frequency domain (σ^2)	HRV, EEG (Malarvili and Mesbah 2009, Oh et al 2014)	Measures activity; used in neonatal seizure detection from ECG-based HRV (Malarvili and Mesbah 2009)
	Ratio of SDs from the first derivative and original signal (σ_d/σ)	HRV, EEG (Malarvili and Mesbah 2009, Oh et al 2014)	Measures mobility; used in neonatal seizure detection from ECG-based HRV (Malarvili and Mesbah 2009)
	Change in frequency ($\sigma_{dd}\sigma_d/\sigma_d\sigma$)	HRV, EEG (Malarvili and Mesbah 2009, Oh et al 2014)	Measure complexity as deviation from pure sinusoidal wave; used in neonatal seizure detection from ECG-based HRV (Malarvili and Mesbah 2009)
Poincaré/ recurrence plots	Standard deviations perpendicular to and along the line of identity, reflecting short and long-term variability (SD1, SD2)	HRV (RR intervals)	Ratio between SD1/SD2 possible surrogate measure of sympathovagal balance (Hoshi et al 2013)
Frequency/ power spectrum	Integral of the power spectrum from Fourier or wavelet transform, partitioned into very low (VLF), low (LF) and high frequency (HF) power bands	HRV	VLF of HRV associated with mortality (Bigger et al 1992); LF with both sympathetic and parasympathetic inputs; HF of parasympathetic activity
Entropy	Approximate entropy (ApEn) quantifying likelihood that similar patterns remain close in adjacent comparisons (Pincus 1991)	HRV, EEG (Beuchée et al 2009, Srinivasan et al 2007)	In HRV, decreased ApEn used in identifying sepsis in premature infants (Richman and Moorman 2000, Beuchée et al 2009), though with reference to other parameters

(Continued)

Table 2. (*Continued*)

Technique	Parameter	Signals applied to	Physiological interpretation; clinical significance
	Sample entropy (SampEn), quantifying self-similarity in matches; similar to ApEn but excluding self-matches (Richman and Moorman 2000)	HRV (Lake *et al* 2002)	In HRV, SampEn observed to decrease prior to clinical signs of neonatal sepsis (Lake *et al* 2002), though may be reflective of increased non-stationarity (Cao *et al* 2004)
Detrended fluctuation analysis	Scaling exponent (α) quantifying fractal-like self-similarity in fluctuations across time scales	HRV, ABP (Tuzcu *et al* 2009, Zhang *et al* 2013)	Fractal patterns demontrated in heart rate and blood pressure variability (Tuzcu *et al* 2009, Zhang *et al* 2013, Fairchild *et al* 2014)
Running correlation	Correlation coefficient (r) between two signals applied on a running window basis	ABP and oximetry (Gilmore *et al* 2011), TOI and HR (da Costa *et al* 2015), MAP and cerebral tissue oxygenation saturation (Chalak *et al* 2016)	In oximetry, used to characterise cerebral autoregulation changes, associated with IVH (Gilmore *et al* 2011). In TOI versus HR, used to define optimal, patient-specific blood pressure ranges for potentially identifying morbidity/mortality risk (da Costa *et al* 2015). In MAP versus cerebral tissue oxygenation saturation, used to characterise cerebral autoregulation over time (Chalak *et al* 2016)
Phase synchronisation	Degree to which two signals oscillate together (λ)	HRV and respiratory rate variability (Nguyen *et al* 2012)	Used to characterise cardiorespiratory coupling/ decoupling; possible indicator of stability in critical conditions (Nguyen *et al* 2012)

This is further motivated by the understanding these conditions are systemic and manifest in ways that are not limited to a single physiological parameter, spanning cardiac signals to respiration rate, arterial blood pressure and oxygen saturation. Their application however is not without unique consideration—respiration rate, for example, may be confounded by mechanical ventilation in the NICU context.

HRV and respiration rate variability have been independently shown to offer predictive potential in the NICU context and may offer improved prediction performance through combination with each other or other informative features. This approach was adopted by Green *et al* in the discernment of adult organ dysfunction cases and severity characterisation (Green *et al* 2013). ECG and EEG-based features were also combined and reported to improve automatic neonatal seizure detection (Greene *et al* 2007).

From a classification standpoint, multiclass classifiers also present opportunities to characterise the severity of these conditions. IVH for example, is a suitable candidate for such classification, given the staged approach to its diagnosis using the four-grade Papile system (Papile *et al* 1978). Similar approaches to identifying the severity of septic shock in the case of neonatal sepsis may also be possible. Among the available approaches for classification, there are also varying degrees of complexity. These range from simpler logistic regression models to the integration with support vector machines (Temko *et al* 2011, 2012) and neural networks (Srinivasan *et al* 2007).

4.2. Clinical interpretation

There are a number discussion points surrounding the clinical applications of variability analysis. Although variability studies may have demonstrated potential to distinguish or classify disease states on this basis, there remains the challenge of physiological interpretation of these differences. Changes in indices of variability may not necessarily indicate a physiological change in system control.

One such example is found in the interpretation of the frequency components of HRV which was highlighted early on by the task force paper (Malik 1996). The higher frequency component has been associated with parasympathetic input, though there remains conjecture around the interpretation of the lower frequency component. It is thought to be reflective of both sympathetic and parasympathetic activity (Eckberg 1997, Longin *et al* 2005, Billman 2007, Shaffer *et al* 2014) by some, and an indicator of sympathetic excitation by others (Malliani *et al* 1991, Montano *et al* 2009), though this evidence mostly stems from investigations of orthostatic tilt effects on HRV (Reyes del Paso *et al* 2013). The interpretation of such HRV changes also requires consideration that these components may be influenced by both non-neural and autonomic mechanisms and their valid interpretation may be reliant on controlled studies and specific populations (Lombardi 2011).

Another example is the scaling exponent α from DFA, which can be defined for short and long-term (α_1 and α_2, respectively) ranges. Modelling has aided the understanding of how changes in sympatho-vagal activity can impact these exponents calculated for HRV (Rojo-Álvarez *et al* 2007), while studies using vagal blockade by atropine have demonstrated an increase in the short-term scaling exponent of blood pressure fluctuations (Castiglioni *et al* 2011).

These studies have helped to clarify the mechanisms that may influence parameters from advanced analyses, although in practice, clinical application and interpretation is often difficult. Associations between variability parameters and respective pathophysiology may be confounded by external influences, such as medically-indicated interventions altering the underlying dynamics of the data or mechanical ventilation in the interpretation of respiratory

R271

29

variability parameters. Another factor is limitations in the data acquired, e.g. lack of precision in heartbeat detection for HRV or air flow or breath timing for respiratory variability analyses, missing data or insufficient length of monitoring, presence of artefacts, linear or non-linear trends or violation of assumptions behind the analysis techniques. Finally, there is also the possibility that these techniques are simply unable to capture the underlying mechanisms for changes in pathophysiology.

Although the exact mechanisms of influence may be unknown and under investigation, we may continue to develop an understanding of how variability patterns and their links to the onset of conditions such as neonatal sepsis, seizures and IVH.

4.3. Other emerging applications

Characterising physiological variability can also extend beyond neonatal conditions and the onset of diseases; it has also been applied in the prediction of extubation-readiness in neonates, with Kaczmarek *et al* reporting significant HRV reductions in those who failed extubation (Kaczmarek *et al* 2013). This becomes particularly important when considering the increased incidence of neurodevelopmental problems linked to prolonged mechanical ventilation (Walsh *et al* 2005). A cardiorespiratory approach was also used in this area (Precup *et al* 2012), where frequency-based features of HRV and parameters measured using respiratory inductive plethysmography were combined for training support vector machine classification.

Several monitoring systems and software have been developed, grounded in the work on clinical/physiological variability analysis. The previously described HeRO monitor represents a more neonate-specific development, although recent years have seen the introduction of continuous individualised multiorgan variability analysis (CIMVA) software. This computes features (such as HRV and respiratory rate variability) using a range of variability analysis techniques and accepting input from multiple physiological waveforms (Bravi 2013). Systems such as these highlight the potential to support clinical decision-making and risk assessment, though there remains the question of the extent to which they impact and are responsible for the final decision. While specific medical claims are not made in relation to these technologies, they may still be introduced to these settings as objective tools for clinicians to improve prognostication and guide interventions (Seely and Macklem 2004).

5. Conclusion

The neonatal intensive care environment presents many unique challenges to monitoring and management of its patients, from the criticality of the conditions faced, the abrupt and sudden nature of changes in system state as well as the long-term implications of the treatments done at this stage of life. The availability of physiological signals and monitoring systems in the NICU establishes a foundation for further signal processing. Variability analysis techniques can be utilised to capture short- and long-term data trends, develop markers for risk as well as assist understanding of the patterns associated with the onset of illness and other critical conditions. In an appropriate context and with sufficient validation, these techniques offer caregivers more time in responding to and treating critically-ill patients.

ORCID iDs

J Huvanandana ● https://orcid.org/0000-0001-7683-9070
C Thamrin ● https://orcid.org/0000-0003-3284-3558

References

Ahmad S *et al* 2009a Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults *PLoS One* **4** e6642

Ahmad S, Tejuja A, Newman K D, Zarychanski R and Seely A J 2009b Clinical review: a review and analysis of heart rate variability and the diagnosis and prognosis of infection *Crit. Care* **13** 232

Baird T M, Goydos J M and Neuman M R 1992 Optimal electrode location for monitoring the ECG and breathing in neonates *Pediatric Pulmonol.* **12** 247–50

Bassan H, Bental Y, Shany E, Berger I, Froom P, Levi L and Shiff Y 2008 Neonatal seizures: dilemmas in workup and management *Pediatric Neurol.* **38** 415–21

Bel F, Bor M, Stijnen T, Baan J and Ruys J H 1987 Aetiological rôle of cerebral blood-flow alterations in development and extension of peri-intraventricular haemorrhage *Dev. Med. Child Neurol.* **29** 601–14

Bell M J, Ternberg J L, Feigin R D, Keating J P, Marshall R, Barton L and Brotherton T 1978 Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging *Ann. Surg.* **187** 1

Benitz W 2010 Treatment of persistent patent ductus arteriosus in preterm infants: time to accept the null hypothesis? *J. Perinatol.* **30** 241–52

Beuchée A, Carrault G, Bansard J Y, Boutaric E, Bétrémieux P and Pladys P 2009 Uncorrelated randomness of the heart rate is associated with sepsis in sick premature infants *Neonatology* **96** 109–14

Bhatia V, Rarick K R and Stauss H M 2010 Effect of the data sampling rate on accuracy of indices for heart rate and blood pressure variability and baroreflex function in resting rats and mice *Physiol. Meas.* **31** 1185

Bigger J T, Fleiss J L, Steinman R C, Rolnitzky L M, Kleiger R E and Rottman J N 1992 Frequency domain measures of heart period variability and mortality after myocardial infarction *Circulation* **85** 164–71

Billman G E 2007 The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Front. Physiol.* **4** 26

Billman G E, Huikuri H V, Sacha J and Trimmel K 2015 An introduction to heart rate variability: methodological considerations and clinical applications *Front. Physiol.* **6** 55

Bittigau P *et al* 2002 Antiepileptic drugs and apoptotic neurodegeneration in the developing brain *Proc. Natl Acad. Sci.* **99** 15089–94

Boghossian N S *et al* 2013 Late-onset sepsis in very low birth weight infants from singleton and multiple-gestation births *J. Pediatrics* **162** 1120–4

Bohanon F J, Mrazek A A, Shabana M T, Mims S, Radhakrishnan G L, Kramer G C and Radhakrishnan R S 2015 Heart rate variability analysis is more sensitive at identifying neonatal sepsis than conventional vital signs *Am. J. Surg.* **210** 661–7

Bolisetty S *et al* 2013 Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants *Pediatrics* **13** 55–62

Bravi A 2013 CIMVA Core Description manual (http://ohridal.org/cimva/CIMVA-Core-Description.pdf) (Accessed: 18 Sept 2017)

Bravi A, Longtin A and Seely A J 2011 Review and classification of variability analysis techniques with clinical applications *Biomed. Eng. Online* **10** 90

Brennan M, Palaniswami M and Kamen P 2001 Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng.* **48** 1342–7

Bryce R and Sprague K 2012 Revisiting detrended fluctuation analysis *Sci. Rep.* **2** 315

Cannesson M, Slieker J, Desebbe O, Bauer C, Chiari P, Hénaine R and Lehot J J 2008 The ability of a novel algorithm for automatic estimation of the respiratory variations in arterial pulse pressure to monitor fluid responsiveness in the operating room *Anesthesia Analgesia* **106** 1195–200

Cao H, Griffin M, Lake D and Moorman J 2004 Increased nonstationarity of neonatal heart rate prior to sepsis and systemic inflammatory response syndrome *Ann. Biomed. Eng.* **32** 233–44

Caplan M S 2008 Neonatal necrotizing enterocolitis—introduction *Seminars in Perinatology* vol 32 (Philadelphia, PA: Saunders) p 69

Castiglioni P, Parati G, Di Rienzo M, Carabalona R, Cividjian A and Quintin L 2011 Scale exponents of blood pressure and heart rate during autonomic blockade as assessed by detrended fluctuation analysis *J. Physiol.* **589** 355–69

Chalak L F, Tian F, Tarumi T and Zhang R 2016 Cerebral hemodynamics in asphyxiated newborns undergoing hypothermia therapy: pilot findings using a multiple-time-scale analysis *Pediatric Neurol.* **55** 30–6

Choi A and Shin H 2017 Photoplethysmography sampling frequency: pilot assessment of how low can we go to analyze pulse rate variability with reliability? *Physiol. Meas.* **38** 586

ChuDuc H, NguyenPhan K and NguyenViet D 2013 A review of heart rate variability and its applications *APCBEE Proc.* **7** 80–5

Collins M P, Lorenz J M, Jetton J R and Paneth N 2001 Hypocapnia and other ventilation-related risk factors for cerebral palsy in low birth weight infants *Pediatric Res.* **50** 712–9

Connell T G, Rele M, Cowley D, Buttery J P and Curtis N 2007 How reliable is a negative blood culture result? Volume of blood submitted for culture in routine practice in a children's hospital *Pediatrics* **119** 891–6

Costa M, Goldberger A L and Peng C K 2002 Multiscale entropy analysis of complex physiologic time series *Phys. Rev. Lett.* **89** 068102

Costa M, Goldberger A L and Peng C K 2005 Multiscale entropy analysis of biological signals *Phys. Rev.* E **71** 021906

da Costa C S, Czosnyka M, Smielewski P, Mitra S, Stevenson G N and Austin T 2015 Monitoring of cerebrovascular reactivity for determination of optimal blood pressure in preterm infants *J. Pediatrics* **167** 86–91

Dong Y and Speer C P 2014 Late-onset neonatal sepsis: recent developments *Archives of Disease in Childhood-Fetal and Neonatal Edition* **100** F257–F263

Eckberg D L 1997 Sympathovagal balance *Circulation* **96** 3224–32

Eckmann J P and Ruelle D 1985 Ergodic theory of chaos and strange attractors *Rev. Mod. Phys.* **57** 617

Ellis R J, Zhu B, Koenig J, Thayer J F and Wang Y 2015 A careful look at ECG sampling frequency and r-peak interpolation on short-term measures of heart rate variability *Physiol. Meas.* **36** 1827

Erickson S, Grauaug A, Gurrin L and Swaminathan M 2002 Hypocarbia in the ventilated preterm infant and its effect on intraventricular haemorrhage and bronchopulmonary dysplasia *J. Paediatrics Child Health* **38** 560–2

Evans N 2015 Preterm patent ductus arteriosus: a continuing conundrum for the neonatologist? *Seminars in Fetal and Neonatal Medicine* vol **20** (Amsterdam: Elsevier) pp 272–7

Evans N and Kluckow M 1996 Early ductal shunting and intraventricular haemorrhage in ventilated preterm infants *Arch. Dis. Child.-Fetal Neonatal Ed.* **75** F183–6

Evans N, Malcolm G, Osborn D and Kluckow M 2004 Diagnosis of patent ductus arteriosus in preterm infants *NeoReviews* **5** e86–97

Ewing D J, Neilson J and Travis P 1984 New method for assessing cardiac parasympathetic activity using 24 h electrocardiograms *Br. Heart J.* **52** 396–402

Fairchild K D, Lake D E, Kattwinkel J, Moorman J R, Bateman D A, Grieve P G, Isler J R and Sahni R 2017 Vital signs and their cross-correlation in sepsis and NEC: a study of 1065 very low birth weight infants in two nicus *Pediatric Res.* **81** 315

Fairchild K D *et al* 2013 Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial *Pediatric Res.* **74** 570–5

Fairchild K, Mohr M, Paget-Brown A, Tabacaru C, Lake D, Delos J, Moorman J R and Kattwinkel J 2016 Clinical associations of immature breathing in preterm infants: part 1—central apnea *Pediatric Res.* **80** 21–7

Fairchild K, Sinkin R, Davalian F, Blackman A, Swanson J, Matsumoto J, Lake D, Moorman J and Blackman J 2014 Abnormal heart rate characteristics are associated with abnormal neuroimaging and outcomes in extremely low birth weight infants *J. Perinatol.* **34** 375–9

Finer N N, Higgins R, Kattwinkel J and Martin R J 2006 Summary proceedings from the apnea-of-prematurity group *Pediatrics* **117** S47–51

Frey U, Silverman M, Barabasi A and Suki B 1998 Irregularities and power law distributions in the breathing pattern in preterm and term infants *J. Appl. Physiol.* **85** 789–97

Funk D, Sebat F and Kumar A 2009 A systems approach to the early recognition and rapid administration of best practice therapy in sepsis and septic shock *Curr. Opin. Crit. Care* **15** 301–7

Furlan R, Guzzetti S, Crivellaro W, Dassi S, Tinelli M, Baselli G, Cerutti S, Lombardi F, Pagani M and Malliani A 1990 Continuous 24 h assessment of the neural regulation of systemic arterial pressure and RR variabilities in ambulant subjects *Circulation* **81** 537–47

Futagi Y, Toribe Y, Ogawa K and Suzuki Y 2006 Neurodevelopmental outcome in children with intraventricular hemorrhage *Pediatric Neurol.* **34** 219–24

Garcia-Gonzalez M, Fernandez-Chimeno M and Ramos-Castro J 2004 Bias and uncertainty in heart rate variability spectral indices due to the finite ECG sampling frequency *Physiol. Meas.* **25** 489

Gardner A B, Krieger A M, Vachtsevanos G and Litt B 2006 One-class novelty detection for seizure analysis from intracranial EEG *J. Mach. Learn. Res.* **7** 1025–44

Gilmore M, Stone B, Shepard J, Czosnyka M, Easley R and Brady K 2011 Relationship between cerebrovascular dysautoregulation and arterial blood pressure in the premature infant *J. Perinatol.* **31** 722–9

Glass H C, Kan J, Bonifacio S L and Ferriero D M 2012 Neonatal seizures: treatment practices among term and preterm infants *Pediatric Neurol.* **46** 111–5

Glass L 1999 Chaos and heart rate variability *J. Cardiovascular Electrophysiol.* **10** 1358–60

Goldberger A L 1997 Fractal variability versus pathologic periodicity: complexity loss and stereotypy in disease *Perspect. Biol. Med.* **40** 543–61

Goldstein B *et al* 2005 International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics *Pediatric Crit. Care Med.* **6** 2–8

Grassberger P and Procaccia I 1983 Measuring the strangeness of strange attractors *Phys. D: Nonlinear Phenom.* **9** 189–208

Green G C, Bradley B, Bravi A and Seely A J 2013 Continuous multiorgan variability analysis to track severity of organ failure in critically ill patients *J. Crit. Care* **28** 879

Greene B, Faul S, Marnane W, Lightbody G, Korotchikova I and Boylan G 2008 A comparison of quantitative EEG features for neonatal seizure detection *Clin. Neurophysiol.* **119** 1248–61

Greene B R, Boylan G B, Reilly R B, de Chazal P and Connolly S 2007 Combination of EEG and ECG for improved automatic neonatal seizure detection *Clin. Neurophysiol.* **118** 1348–59

Greene B R, de Chazal P, Boylan G B, Connolly S and Reilly R B 2007 Electrocardiogram based neonatal seizure detection *IEEE Trans. Biomed. Eng.* **54** 673–82

Griffin M P and Moorman J R 2001 Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis *Pediatrics* **107** 97–104

Griffin M P, Lake D E, Bissonette E A, Harrell F E, O'Shea T M and Moorman J R 2005 Heart rate characteristics: novel physiomarkers to predict neonatal infection and death *Pediatrics* **116** 1070–4

Griffin M P, O'Shea T M, Bissonette E A, Harrell F E, Lake D E and Moorman J R 2003 Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness *Pediatric Res.* **53** 920–6

Haider A, Rehan V, Al-Saedi S, Alvaro R, Kwiatkowski K, Cates D and Rigatto H 1995 Effect of baseline oxygenation on the ventilatory response to inhaled 100% oxygen in preterm infants *J. Appl. Physiol.* **79** 2101–5

Haider A Z, Rehan V, Alvaro R, Al-Saedi S, Kwiatkowski K, Cates D and Rigatto H 1996 Low baseline oxygenation predisposes preterm infants to mixed apneas during inhalation of 100% oxygen *Am. J. Perinatol.* **13** 363–9

Hanna B, Nelson M, White-Traut R C, Silvestri J, Vasan U, Rey P M, Patel M and Comiskey E 2000 Heart rate variability in preterm brain-injured and very-low-birth-weight infants *Neonatology* **77** 147–55

Henry M C and Moss R L 2009 Necrotizing enterocolitis *Ann. Rev. Med.* **60** 111–24

Henslee J, Schechtman V, Lee M and Harper R 1997 Developmental patterns of heart rate and variability in prematurely-born infants with apnea of prematurity *Early Human Dev.* **47** 35–50

Heuchan A, Evans N, Smart D H and Simpson J 2002 Perinatal risk factors for major intraventricular haemorrhage in the australian and new zealand neonatal network, 1995–97 *Arch. Dis. Child.-Fetal Neonatal Ed.* **86** F86–90

Hira Z M and Gillies D F 2015 A review of feature selection and feature extraction methods applied on microarray data *Adv. Bioinform.* **2015** 198363

Hjorth B 1973 The physical significance of time domain descriptors in EEG analysis *Electroencephalogr. Clin. Neurophysiol.* **34** 321–5

Horsfield K and Cumming G 1968 Morphology of the bronchial tree in man *J. Appl. Physiol.* **24** 373–83

Hoshi R A, Pastre C M, Vanderlei L C M and Godoy M F 2013 Poincaré plot indexes of heart rate variability: relationships with other nonlinear variables *Autonomic Neurosci.* **177** 271–4

Imhoff M and Kuhls S 2006 Alarm algorithms in critical care monitoring *Anesthesia Analgesia* **102** 1525–37

Ivanov P C *et al* 1996 Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis *Nature* **383** 323

Kaczmarek J, Chawla S, Marchica C, Dwaihy M, Grundy L and Sant'Anna G M 2013 Heart rate variability and extubation readiness in extremely preterm infants *Neonatology* **104** 42–8

Kane E A and Higham T E 2015 Complex systems are more than the sum of their parts: using integration to understand performance, biomechanics, and diversity *Integr. Comparative Biol.* **55** 146–65

Kluckow M and Evans N 2000 Ductal shunting, high pulmonary blood flow, and pulmonary hemorrhage *J. Pediatrics* **137** 68–72

Kotsiantis S B, Zaharakis I and Pintelas P 2007 Supervised machine learning: a review of classification techniques *Informatica* **31** 249–68

Kristoufek L 2014 Measuring correlations between non-stationary series with DCCA coefficient *Physica A* **402** 291–8

Küster H, Weiss M, Willeitner A E, Detlefsen S, Jeremias I, Zbojan J, Geiger R, Lipowsky G and Simbruner G 1998 Interleukin-1 receptor antagonist and interleukin-6 for early diagnosis of neonatal sepsis 2 days before clinical manifestation *Lancet* **352** 1271–7

Kwapień J, Oświęcimka P and Drożdż S 2015 Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations *Phys. Rev.* E **92** 052815

Lahra M M, Beeby P J and Jeffery H E 2009 Intrauterine inflammation, neonatal sepsis, and chronic lung disease: a 13-year hospital cohort study *Pediatrics* **123** 1314–9

Lake D E, Richman J S, Griffin M P and Moorman J R 2002 Sample entropy analysis of neonatal heart rate variability *Am. J. Physiol.-Regulatory, Integr. Comparative Physiol.* **283** R789–97

Lanska M J, Lanska D J, Baumann R J and Kryscio R J 1995 A population-based study of neonatal seizures in Fayette county, Kentucky *Neurology* **45** 724–32

Laughon M M, Simmons M A and Bose C L 2004 Patency of the ductus arteriosus in the premature infant: is it pathologic? Should it be treated? *Curr. Opin. Pediatrics* **16** 146–51

Lee H C, Bennett M V, Schulman J and Gould J B 2013 Accounting for variation in length of NICU stay for extremely low birth weight infants *J. Perinatol.* **33** 872–6

Lee H *et al* 2011 A new algorithm for detecting central apnea in neonates *Physiol. Meas.* **33** 1

Lemons J A *et al* 2001 Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 through December 1996 *Pediatrics* **107** e1

Lin P W and Stoll B J 2006 Necrotising enterocolitis *Lancet* **368** 1271–83

Lombardi F 2011 Origin of heart rate variability and turbulence: an appraisal of autonomic modulation of cardiovascular function *Front. Physiol.* **2** 95

Lommen C, Pasman J, Van Kranen V, Andriessen P, Cluitmans P, Van Rooij L and Bambang Oetomo S 2007 An algorithm for the automatic detection of seizures in neonatal amplitude-integrated EEG *Acta Paediatrica* **96** 674–80

Longin E, Schaible T, Lenz T and König S 2005 Short term heart rate variability in healthy neonates: normative data and physiological observations *Early Hum. Dev.* **81** 663–71

Luig M and Lui K 2005 Epidemiology of necrotizing enterocolitis—part ii: risks and susceptibility of premature infants during the surfactant era: a regional study *J. Paediatrics Child Health* **41** 174–9

Malarvili M B and Mesbah M 2009 Newborn seizure detection based on heart rate variability *IEEE Trans. Biomed. Eng.* **56** 2594–603

Malik M 1996 Heart rate variability *Ann. Noninvasive Electrocardiol.* **1** 151–81

Malliani A, Pagani M, Lombardi F and Cerutti S 1991 Cardiovascular neural regulation explored in the frequency domain *Circulation* **84** 482–92

Mansier P, Clairambault J, Charlotte N, Médigue C, Vermeiren C, LePape G, Carré F, Gounaropoulou A and Swynghedauw B 1996 Linear and non-linear analyses of heart rate variability: a minireview *Cardiovascular Res.* **31** 371–9

McClure C, Jang S Y and Fairchild K 2016 Alarms, oxygen saturations, and spo2 averaging time in the NICU *J. Neonatal-Perinatal Med.* **9** 357–62

Merri M, Farden D C, Mottley J G and Titlebaum E L 1990 Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability *IEEE Trans. Biomed. Eng.* **37** 99–106

Michie D, Spiegelhalter D J and Taylor C C 1994 *Machine Learning, Neural and Statistical Classification* (Upper Saddle River, NJ: Ellis Horwood)

Monasterio V, Burgess F and Clifford G D 2012 Robust classification of neonatal apnoea-related desaturations *Physiol. Meas.* **33** 1503

Montano N, Porta A, Cogliati C, Costantino G, Tobaldini E, Casali K R and Iellamo F 2009 Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior *Neurosci. Biobehavioral Rev.* **33** 71–80

Moody C, Callahan T J, Aldrich H, Gance-Cleveland B and Sables-Baus S 2017 Early initiation of newborn individualized developmental care and assessment program (NIDCAP) reduces length of stay: a quality improvement project *J. Pediatric Nurs.* **32** 59–63

Moorman J R *et al* 2011 Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial *J. Pediatrics* **159** 900–6

Moorman J R, Lake D E and Griffin M P 2006 Heart rate characteristics monitoring for neonatal sepsis *IEEE Trans. Biomed. Eng.* **53** 126–32

Neu J, Mshvildadze M and Mai V 2008 A roadmap for understanding and preventing necrotizing enterocolitis *Curr. Gastroenterol. Rep.* **10** 450–7

Nguyen C D, Dakin C, Yuill M, Crozier S and Wilson S 2012 The effect of sigh on cardiorespiratory synchronisation in healthy sleeping infants *Sleep* **35** 1643–50

Nilsson L M 2013 Respiration signals from photoplethysmography *Anesthesia Analgesia* **117** 859–65

Noori S, McCoy M, Friedlich P, Bright B, Gottipati V, Seri I and Sekar K 2009 Failure of ductus arteriosus closure is associated with increased mortality in preterm infants *Pediatrics* **123** e138–44

Oh S H, Lee Y R and Kim H N 2014 A novel EEG feature extraction method using Hjorth parameter *Int. J. Electron. Electr. Eng.* **2** 106–10

Ohlsson A, Walia R and Shah S S 2013 Ibuprofen for the treatment of patent ductus arteriosus in preterm and/or low birth weight infants *Cochrane Database Syst. Rev.* **2** CD003481

Oświęcimka P, Drożdż S, Forczek M, Jadach S and Kwapień J 2014 Detrended cross-correlation analysis consistently extended to multifractality *Phys. Rev.* E **89** 023305

Papile L A, Burstein J, Burstein R and Koffler H 1978 Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1 500 g *J. Pediatrics* **92** 529–34

Papillon S, Castle S L, Gayer C P and Ford H R 2013 Necrotizing enterocolitis *Adv. Pediatrics* **60** 263–79

Peng C K, Havlin S, Stanley H E and Goldberger A L 1995 Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series *Chaos* **5** 82–7

Pincus S and Singer B H 1996 Randomness and degrees of irregularity *Proc. Natl Acad. Sci.* **93** 2083–8

Pincus S M 1991 Approximate entropy as a measure of system complexity *Proc. Natl Acad. Sci.* **88** 2297–301

Pincus S M, Mulligan T, Iranmanesh A, Gheorghiu S, Godschalk M and Veldhuis J D 1996 Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males *Proc. Natl Acad. Sci.* **93** 14100–5

Precup D, Robles-Rubio C A, Brown K A, Kanbar L, Kaczmarek J, Chawla S, Sant'Anna G and Kearney R E 2012 Prediction of extubation readiness in extreme preterm infants based on measures of cardiorespiratory variability *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* (IEEE) pp 5630–3

Renner J, Broch O, Gruenewald M, Scheewe J, Francksen H, Jung O, Steinfath M and Bein B 2011 Non-invasive prediction of fluid responsiveness in infants using pleth variability index *Anaesthesia* **66** 582–9

Reyes del Paso G A, Langewitz W, Mulder L J, Roon A and Duschek S 2013 The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies *Psychophysiology* **50** 477–87

Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy *Am. J. Physiol. Heart Circ. Physiol.* **278** H2039–49

Robertson N and Rennie J 1992 *Textbook of Neonatology* (London: Churchill) pp 984–5

Rojo-Álvarez J L, Sanchez-Sanchez A, Barquero-Perez O, Goya-Esteban R, Everss E, Mora-Jimenez I and Garcia-Alberola A 2007 Analysis of physiological meaning of detrended fluctuation analysis in heart rate variability using a lumped parameter model *Computers in Cardiology* (IEEE) pp 25–8

Ronen G M, Penney S and Andrews W 1999 The epidemiology of clinical neonatal seizures in Newfoundland: a population-based study *J. Pediatrics* **134** 71–5

Saeys Y, Inza I and Larrañaga P 2007 A review of feature selection techniques in bioinformatics *Bioinformatics* **23** 2507–17

Saliba R M, Annegers J F, Waller D K, Tyson J E and Mizrahi E M 1999 Incidence of neonatal seizures in Harris county, Texas, 1992–4 *Am. J. Epidemiol.* **150** 763–9

Schmidt G, Malik M, Barthel P, Schneider R, Ulm K, Rolnitzky L, Camm A J, Bigger J T and Schömig A 1999 Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction *Lancet* **353** 1390–6

Schneider D J and Moore J W 2006 Patent ductus arteriosus *Circulation* **114** 1873–82

Seely A J and Macklem P T 2004 Complex systems and the technology of variability analysis *Crit. Care* **8** R367

Shaffer F, McCraty R and Zerr C L 2014 A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability *Frontiers Psychol.* **5** 1040

Shen C H, Li C L and Si Y L 2015 A detrended cross-correlation analysis of meteorological and API data in Nanjing, China *Phys. A: Stat. Mech. Appl.* **419** 417–28

Sortica da Costa C, Placek M M, Czosnyka M, Cabella B, Kasprowicz M, Austin T and Smielewski P 2017 Complexity of brain signals is associated with outcome in preterm infants *J. Cerebral Blood Flow Metab.* accepted (http://doi.org/10.1177/0271678X16687314)

Srinivasan V, Eswaran C and Sriraam N 2007 Approximate entropy-based epileptic EEG detection using artificial neural networks *IEEE Trans. Inf. Technol. Biomed.* **11** 288–95

Stanley H, Amaral L N, Goldberger A, Havlin S, Ivanov P C and Peng C K 1999 Statistical physics and physiology: monofractal and multifractal approaches *Phys. A: Stat. Mech. Appl.* **270** 309–24

Stoll B J *et al* 2002 Late-onset sepsis in very low birth weight neonates: the experience of the NICHD neonatal research network *Pediatrics* **110** 285–91

Stone M L, Tatum P M, Weitkamp J H, Mukherjee A B, Attridge J, McGahren E D, Rodgers B M, Lake D E, Moorman J R and Fairchild K D 2013 Abnormal heart rate characteristics before clinical diagnosis of necrotizing enterocolitis *J. Perinatol.* **33** 847–50

Sullivan B A and Fairchild K D 2015 Predictive monitoring for sepsis and necrotizing enterocolitis to prevent shock *Seminars in Fetal and Neonatal Medicine* vol 20 (Amsterdam: Elsevier) pp 255–61

Taeusch H W, Ballard R A, Gleason C A and Avery M E 2005 *Avery's Diseases of the Newborn* (Amsterdam: Elsevier)

Temko A, Lightbody G, Thomas E M, Boylan G B and Marnane W 2012 Instantaneous measure of EEG channel importance for improved patient-adaptive neonatal seizure detection *IEEE Trans. Biomed. Eng.* **59** 717–27

Temko A, Thomas E, Marnane W, Lightbody G and Boylan G 2011 EEG-based neonatal seizure detection with support vector machines *Clin. Neurophysiol.* **122** 464–73

Thamrin C and Stern G 2010 New methods: what do they tell us? Fluctuation analysis of lung function *Eur. Respiratory Monogr.* **47** 310–24

Thuraisingham R A and Gottwald G A 2006 On multiscale entropy analysis for physiological data *Physica* A **366** 323–32

Torrence C and Compo G P 1998 A practical guide to wavelet analysis *Bull. Am. Meteorol. Soc.* **79** 61–78

Tröger B *et al* 2014 Risk for late-onset blood-culture proven sepsis in very-low-birth weight infants born small for gestational age: a large multicenter study from the german neonatal network *Pediatric Infectious Dis. J.* **33** 238–43

Tuzcu V, Nas S, Ulusar U, Ugur A and Kaiser J R 2009 Altered heart rhythm dynamics in very low birth weight infants with impending intraventricular hemorrhage *Pediatrics* **123** 810–5

van Ravenswaaij-Arts C M, Hopman J C, Kollée L A, van Amen J P, Stoelinga G B and van Geijn H P 1991 The influence of respiratory distress syndrome on heart rate variability in very preterm infants *Early Hum. Dev.* **27** 207–21

van Wezel-Meijler G, Steggerda S J and Leijser L M 2010 Cranial ultrasonography in neonates: role and limitations *Semin. Perinatol.* **34** Elsevier 28–38

Vaughn B V, Quint S, Messenheimer J and Robertson K R 1995 Heart period variability in sleep *Electroencephalogr. Clin. Neurophysiol.* **94** 155–62

Vergales B D *et al* 2014 Accurate automated apnea analysis in preterm infants *Am. J. Perinatol.* **31** 157–62

Vesoulis Z A, Hao J, McPherson C, El Ters N and Mathur A M 2017 Low-frequency blood pressure oscillations and inotrope treatment failure in premature infants *J. Appl. Physiol.* **123** 55–61

Villar J *et al* 2006 Preeclampsia, gestational hypertension and intrauterine growth restriction, related or independent conditions? *Am. J. Obstetrics Gynecol.* **194** 921–31

Volante E, Moretti S, Pisani F and Bevilacqua G 2004 Early diagnosis of bacterial infection in the neonate *J. Maternal-Fetal Neonatal Med.* **16** 13–6

Volpe J J 1989a Intraventricular hemorrhage in the premature infant-current concepts. Part I *Ann. Neurol.* **25** 3–11

Volpe J J 1989b Neonatal seizures: current concepts and revised classification *Pediatrics* **84** 422–8

Volpe J J 2008 *Neurology of the Newborn* 5th edn (Philadelphia, PA: Saunders)

Volpe J J 2009 Cerebellum of the premature infant: rapidly developing, vulnerable, clinically important *J. Child Neurol.* **24** 1085–104

Voss A, Kurths J, Kleiner H, Witt A, Wessel N, Saparin P, Osterziel K, Schurath R and Dietz R 1996 The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death *Cardiovascular Res.* **31** 419–33

Walsh M C *et al* 2005 Extremely low birthweight neonates with protracted ventilation: mortality and 18-month neurodevelopmental outcomes *J. Pediatrics* **146** 798–804

Weibel E R 1962 Architecture of the human lung. Use of quantitative methods establishes fundamental relations between size and number of lung structures *Science* **137** 577–85

Xia J and Shang P 2012 Multiscale entropy analysis of financial time series *Fluctuation Noise Lett.* **11** 1250033

Zebende G 2011 DCCA cross-correlation coefficient: quantifying level of cross-correlation *Phys. A: Stat. Mech. Appl.* **390** 614–8

Zhang R, Zuckerman J H and Levine B D 2000 Spontaneous fluctuations in cerebral blood flow: insights from extended-duration recordings in humans *Am. J. Physiol. Heart Circ. Physiol.* **278** H1848–55

Zhang R, Zuckerman J H, Giller C A and Levine B D 1998 Transfer function analysis of dynamic cerebral autoregulation in humans *Am. J. Physiol. Heart Circ. Physiol.* **274** H233–41

Zhang Y, Chan G S, Tracy M B, Hinder M, Savkin A V and Lovell N H 2013 Detrended fluctuation analysis of blood pressure in preterm infants with intraventricular hemorrhage *Med. Biol. Eng. Comput.* **51** 1051–7

Ziemssen T, Gasch J and Ruediger H 2008 Influence of ECG sampling frequency on spectral analysis of RR intervals and baroreflex sensitivity using the eurobavar data set *J. Clin. Monit. Comput.* **22** 159–68

2.2 Concluding remarks

This chapter reviews several time and frequency domain techniques for variability analysis as well as current applications in neonatal intensive care, from predictive monitoring to characterising physiological patterns associated with specific treatments and conditions. In the following three chapters, we explore the potential for early identification of infants at risk of IVH (Chapter 3), characterise the impact of caffeine therapy on cardiovascular variability (Chapter 4), and describe the dynamics of cerebral autoregulation (Chapter 5).

Chapter 3

Identification of preterm infants at risk of IVH

This chapter presents the early work on model development for the identification of infants at high risk of developing IVH. This chapter addresses the following research questions:

2: Can fluctuation analysis and multivariable models be applied to improve the early detection of IVH in the NICU setting?

3: Can pre-processing techniques improve the performance of DFA?

The work on this chapter utilises detrended fluctuation analysis (DFA), described in Chapter 2, to characterise blood pressure and respiratory variability in preterm infants. The content presented in this chapter is published [26] as:

Huvanandana, J., Nguyen, C., Thamrin, C., Tracy, M., Hinder, M. and McEwan, A.L., 2017. Prediction of intraventricular haemorrhage in preterm infants using time series analysis of blood pressure and respiratory signals. *Scientific reports*, 7, p.46538.

3.1 Prediction of intraventricular haemorrhage in preterm infants using time series analysis of blood pressure and respiratory signals

Statement of Contributions of Joint Authorship

- Jacqueline Huvanandana (Candidate): corresponding author, completed the analysis, writing, reviewing and editing of the manuscript

- Chinh Nguyen: aided in the analysis, reviewing and editing the manuscript

- Cindy Thamrin: provided technical advice on the analysis, reviewing and editing the manuscript

- Mark Tracy: data collection, provided the main idea, reviewing and editing the manuscript

- Murray Hinder: data collection, reviewing and editing the manuscript

- Alistair McEwan (Principal Supervisor): provided the main idea, reviewing and editing the manuscript

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Alistair McEwan **Date:** 24 May 2018

SCIENTIFIC REP🞱RTS

Prediction of intraventricular haemorrhage in preterm infants using time series analysis of blood pressure and respiratory signals

Jacqueline Huvanandana[1], Chinh Nguyen[2], Cindy Thamrin[2], Mark Tracy[3,4], Murray Hinder[1,3] & Alistair L. McEwan[1]

Despite the decline in mortality rates of extremely preterm infants, intraventricular haemorrhage (IVH) remains common in survivors. The need for resuscitation and cardiorespiratory management, particularly within the first 24 hours of life, are important factors in the incidence and timing of IVH. Variability analyses of heart rate and blood pressure data has demonstrated potential approaches to predictive monitoring. In this study, we investigated the early identification of infants at a high risk of developing IVH, using time series analysis of blood pressure and respiratory data. We also explore approaches to improving model performance, such as the inclusion of multiple variables and signal pre-processing to enhance the results from detrended fluctuation analysis. Of the models we evaluated, the highest area under receiver-operator characteristic curve (5th, 95th percentile) achieved was 0.921 (0.82, 1.00) by mean diastolic blood pressure and the long-term scaling exponent of pulse interval (PI α_2), exhibiting a sensitivity of >90% at a specificity of 75%. Following evaluation in a larger population, our approach may be useful in predictive monitoring to identify infants at high risk of developing IVH, offering caregivers more time to adjust intensive care treatment.

Intraventricular Haemorrhage (IVH) remains a serious threat to survival for preterm infants and neurodevelopmental outcomes[1]. Despite advances in modern neonatal care such as antenatal steroids, artificial surfactant treatment and the use of neuroprotective agents such as magnesium sulphate given to mothers in labour, rates of IVH, particularly high grade, remain unchanged. Prematurity, respiratory-distress syndrome and mechanical ventilation are among the factors that may predispose infants to IVH. Recent studies have also suggested an association between IVH and cerebral pressure passivity, that is, where changes in cerebral blood flow correspond to changes in blood pressure[2]. The need for resuscitation and cardiorespiratory management of preterm infants within the first 24 hours of life play an important role in the development and timing of IVH[3,4], where the majority of these cases can be detected at their full extent by the end of the first postnatal week[5]. The potential to identify infants at high risk of developing IVH is thus, particularly important.

Retrospective studies of premature infants after the diagnosis of IVH have highlighted altered autonomic functions which are reflected by heart rate variability analysis[6,7]. In particular, one study showed that these differences could be detected using electrocardiogram data from the first 24 hours of life[8]. Variability of beat-to-beat systolic blood pressure and mean arterial pressure has also been shown to offer useful information in distinguishing infants who later developed IVH from those who did not[9]. Such distinctions were demonstrated using detrended fluctuation analysis (DFA), a non-linear time domain technique that is able to quantify long-range power law correlations in a given time series. Its application is characterised by a scaling exponent (α) which can be calculated over different time scales and indicates the corresponding degree of correlation[10,11].

More recent work in this area by Fairchild et al. has demonstrated associations between a heart rate characteristic index and adverse neurodevelopmental outcomes or white matter damage[12]. Models for early prediction of IVH have explored either clinical risk factors, as in the case of Luque et al.[13] with an AUC of 0.79, or employed

[1]School of Electrical and Information Engineering, University of Sydney, Sydney, Australia. [2]Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia. [3]Westmead Hospital, Sydney, Australia. [4]School of Paediatrics and Child Health, University of Sydney, Sydney, Australia. Correspondence and requests for materials should be addressed to J.H. (email: j.huvanandana@gmail.com)

41

Variable	IVH (n = 7)	Non-IVH (n = 20)	p
Gestational age (weeks)	26.8 ± 1.2	26.9 ± 1.8	0.781
Birthweight (g)	1120 ± 282	1029 ± 293	0.580
Sex (% male)	57.1 ± 49.5	65.0 ± 47.7	0.741
CRIBII	9 ± 1	9 ± 2	1.000
PDA (%)	85.7 ± 35.0	80.0 ± 40.0	0.774
RDS	1.0 ± 0.0	1.0 ± 0.0	1.000
Apgar 1-min	6 ± 1	6 ± 2	0.696
Apgar 5-min	7 ± 1	7 ± 1	0.421
MAP (mmHg)	32.5 ± 6.1	35.2 ± 4.7	0.422
DBP (mmHg)	25.0 ± 3.9	29.0 ± 4.6	0.050
MAP^c (mmHg)	32.1 ± 5.6	35.5 ± 4.2	0.234
DBP^c (mmHg)	24.6 ± 3.5	29.4 ± 4.1	0.019

Table 1. Comparison of physiological variables between infants who later developed intraventricular haemorrhage (IVH) and those who did not (non-IVH). Values are reported as mean ± SD. ^{c}Denotes detrended features, CRIBII is the Clinical Risk Index for Babies score II, PDA is Patent Ductus Arteriosus and RDS is Respiratory Distress Syndrome. *p* values are derived from a two-sided Mann-Whitney U-test where significance is defined as $p < 0.05$.

time series analysis techniques and physiological signals as in Tuzcu *et al.* who reported a sensitivity of 70% and specificity of 79% for their model using heart rate variability[8].

The objective of this study was twofold; firstly, to explore means of improving prediction of IVH from DFA through pre-processing, and secondly, to evaluate the potential of multivariable or multimodal models in the prediction of IVH. The latter objective focused specifically on combinations of blood pressure and respiratory features and inherently involved an evaluation of robustness as applied in a clinical context. The features evaluated comprised of the mean (μ) as well as short- and long-term scaling exponents derived from DFA (α_1 and α_2, respectively), extracted from five different time series. These were: mean arterial pressure (MAP), systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse interval (PI) series as derived from the arterial blood pressure data, as well as the interbreath intervals (IBI) from the respiratory air flow data.

Results

Study Population.
The study cohort consisted of 27 low birth-weight (<1500 g) infants, 8 of which subsequently developed IVH. We examined the differences in physiological characteristics and other metadata between the two groups, as summarised in Table 1. Although the IVH and non-IVH groups did not exhibit significant differences in the collected metadata, certain clinical characteristics, namely, the mean DBP values were observed to be significantly different for the two groups after non-linear trend removal ($p < 0.05$). This established a foundation for fitting the univariate logistic regression models, though evaluation of their overall performance requires reference to leave-one-out cross-validation (LOOCV) results.

Effect of Detrending.
In the initial stages of model fitting, we observed that in certain instances, the arterial blood pressure was subject to non-linear drift. We examined the effect of detrending the arterial blood pressure signal and noted that prediction performance of the fitted logistic regression models could be improved. This detrending affected the linear blood pressure features in particular, which also propagated to the beat-to-beat blood pressure values and thus the scaling exponents derived from DFA (α_1 and α_2, the short- and long-term scaling exponents, respectively). The Mann-Whitney U-test comparisons of the non-detrended features are shown in Table 2, where the effect of detrending was characterised by the changes in AUC and *p* values from the two-sided Mann-Whitney U-test. For example, the AUC scores of the mean DBP model improved from 0.757 to 0.807 subsequent to detrending. A similar increase from 0.757 to 0.771 was also observed for the univariate model of the long-term scaling exponent of DBP (DBP α_2), motivating the inclusion of this pre-processing step for the subsequent analyses. The histograms of mean DBP and DBP α_1 are also shown in Fig. 1.

Univariate Predictors of IVH.
We fitted univariate logistic regression models using various linear and DFA features, taking the mean across all qualifying time windows of data for each subject. We evaluated the AUC, the 95% confidence interval (5th, 95th percentile) according to the Delong method for determining standard error[14], as well as the positive likelihood ratio and threshold corresponding to a specificity of 75%. These results are summarised in Table 3. Overall, the short-term fractal exponents (α_1) derived from the MAP, SBP and DBP signals as well as the mean DBP yielded the highest AUCs. Respiratory variables were not found to be strongly predictive.

Multivariable Predictors of IVH.
It was also observed that model performance could also be improved through combination of the predictors extracted for the univariate models, as shown in Table 4. Predictor combinations that were significantly correlated were excluded to mitigate effects of collinearity. Many of the multivariable models exhibited higher AUC scores than univariate models, with the highest being the combination of mean DBP and PI α_2.

42

Variable	IVH	Non-IVH	AUC	AUC^c	p	p^c
MAP						
μ	32.5 ± 6.1 mmHg	35.2 ± 4.7 mmHg	0.607	0.657	0.422	0.234
α_1	0.96 ± 0.17	0.78 ± 0.19	0.779	0.779	0.033	0.033
α_2	1.10 ± 0.06	1.00 ± 0.18	0.671	0.65	0.194	0.257
SBP						
μ	41.9 ± 9.7 mmHg	42.7 ± 5.4 mmHg	0.564	0.55	0.638	0.719
α_1	0.83 ± 0.11	0.69 ± 0.15	0.764	0.771	0.043	0.038
α_2	1.04 ± 0.08	0.97 ± 0.16	0.643	0.664	0.281	0.213
DBP						
μ	25.0 ± 3.9 mmHg	29.0 ± 4.6 mmHg	0.757	0.807	0.05	0.019
α_1	0.85 ± 0.12	0.68 ± 0.20	0.786	0.807	0.029	0.019
α_2	1.05 ± 0.06	0.93 ± 0.16	0.757	0.771	0.05	0.038

Table 2. Effect of Detrending. Values are reported as mean $\pm$ SD, p values are from Mann-Whitney U-tests from the non-detrended data. AUC is the area under the ROC curve for prediction of IVH. Note that AUC^c and p^c are obtained from the detrended data.
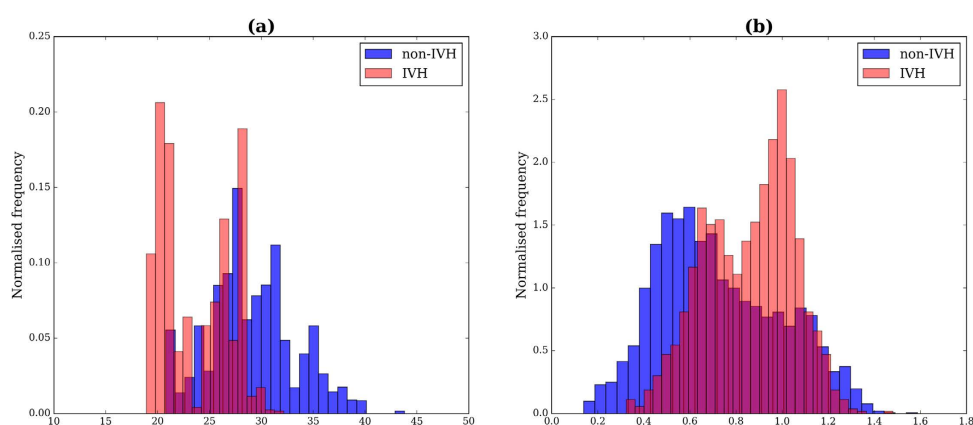

Figure 1. Normalised histograms of (**a**) mean DBP and (**b**) DBP α_1 for IVH and non-IVH groups. The distributions for each group were based on features extracted from all individual windows which met the quality criteria.

Similar results were obtained when evaluating all qualifying windows and with the inclusion of gestational age. Note however that the comparisons with the best performing univariate model (i.e. mean DBP, Table 3) were not statistically significant ($p > 0.05$) due to high degree of overlap. The ROC curves for a number of these models are displayed in Fig. 2, where the non-linear detrending process to obtain the DBP^c μ feature is illustrated in Fig. 3.

Leave-One-Out Cross-Validation. The models were further evaluated using LOOCV, where the probability estimates of each testing sample were used to construct a receiver-operator characteristic (ROC) curve[15,16], as summarised in the latter column of Table 4. Delong comparison of these LOOCV ROC curves with the corresponding LOOCV mean DBP model did not exhibit statistically-significant results ($p > 0.05$), with the exception of the mean DBP and SBP α_1 combination.

Sensitivity Analysis. The models used to obtain these results were based on the mean feature(s) calculated from all qualifying 10 min windows for each subject, with a 30 sec overlap. To evaluate how robust these results were, we examined the effect of including all individual windows rather than the mean feature per subject, of using non-overlapping windows and of including gestational age. The first two cases involved use of a mixed-model allowing for repeated measures while the latter involved the addition of gestational age as a predictor in the existing multivariable models. In all three cases, we found that mean DBP and PI α_2 remained the most predictive combination for IVH (where AUC $= 0.88$, 0.86 and 0.85 for the three cases, respectively, compared to the AUC $= 0.92$ reported in Table 4).

Discussion
Summary of findings. This study evaluated the blood pressure and respiratory features in discerning infants with IVH from those without. The highest AUC achieved was 0.921 (95% CI 0.82, 1.00) by the model fitted with PI α_2 and mean DBP. The results of cross-validation also supported this, with an AUC$_{LOOCV}$ of 0.821 (0.66, 0.99). This model exhibited a sensitivity of $>90\%$ at a specificity of 75% which is greater than that reported for the heart rate variability-based model from Tuzcu *et al.* (70% sensitivity, 79% specificity)[8]. This latter cohort was of a similar

Model	AUC (95% CI)	p	Threshold	LR
MAP				
μ	0.657 (0.37, 0.95)	0.218	31.72 mmHg	2.29
α_1	0.779 (0.60, 0.96)	0.359	0.92	2.86
α_2	0.650 (0.44, 0.86)	0.839	1.08	2.40
SBP				
μ	0.550 (0.20, 0.90)	0.389	37.96 mmHg	2.29
α_1	0.771 (0.58, 0.96)	0.382	0.81	2.86
α_2	0.664 (0.43, 0.90)	0.792	0.94	1.60
DBP				
μ	0.807 (0.62, 0.99)	0.022	26.34 mmHg	2.86
α_1	0.807 (0.64, 0.97)	0.278	0.79	3.43
α_2	0.771 (0.59, 0.95)	0.415	1.02	2.80
PI				
μ	0.543 (0.25, 0.83)	0.759	50.10 ms	1.40
α_1	0.607 (0.38, 0.84)	1.000	0.42	1.40
α_2	0.707 (0.45, 0.97)	0.709	1.08	2.29
IBI				
μ	0.707 (0.46, 0.96)	0.643	115.88 ms	2.40
α_1	0.500 (0.25, 0.75)	0.568	0.52	1.14
α_2	0.557 (0.26, 0.85)	0.813	0.45	0.40

Table 3. Univariate Logistic Regression models. Models were fitted with mean (μ), short- and long-term scaling exponents (α_1 and α_2, respectively) for five time series: mean arterial (MAP), systolic (SBP) and diastolic (DBP) blood pressure, as well as pulse (PI) and interbreath (IBI) intervals. Positive likelihood ratios (LR) and corresponding thresholds are reported at a specificity of 75%. 95% confidence intervals (CI) and p values reported for the AUC are derived from the Delong approach[14] for determining standard error and comparison with the reference ROC of the non-detrended mean MAP model.

size (n = 24), though it was limited to very low birthweight infants (<1000 g) as opposed to our low birthweight cohort, potentially contributing to the difference in IVH representation observed (41.7% compared to 29.6%). Although the univariate model fitted with mean DBP exhibited an AUC of 0.807 in the initial analysis, results from LOOCV cautioned its use as a sole predictor, with an AUC of 0.607 and a non-significant 95% confidence interval of (0.38, 0.88).

Effect of detrending. Out of all the factors we examined, detrending as part of the pre-processing phase of analysis resulted in the greatest improvement in prediction of IVH. It rendered both mean DBP and DBP α_2 significantly different ($p < 0.05$) between the IVH and non-IVH groups, where the AUC scores for two univariate models fitted with these features increased from 0.757 to 0.807 and 0.771, respectively (Tables 2 and 3). These improvements suggest that long-term drift and/or baseline wander of the blood pressure signals, among other aspects of signal quality, may confound the results from DFA as well as linear parameters. The significant contribution of mean DBP is of particular importance, given previous assertions that linear features alone were not an informative proxy of cerebral perfusion pressure[17] and reports that such features were not significantly different between the IVH and non-IVH groups[9]. This relationship to DBP may also have been a reflection of a widened pulse pressure, seen in symptomatic patent ductus arteriosus. It is necessary to note however that this feature was not explicitly evaluated in the previous studies. Aside from the detrending, this analysis differed in other aspects including the quality control measures for window selection and the evaluation process; the features extracted represent the mean of all qualifying 10 min windows across the recording, rather than a single segment. In the clinical application of DFA to monitored signals, we strongly recommend examination of signal data to determine whether overall detrending is necessary prior to analysis.

Use of bivariate models. Subsequent to detrending, a further increase in AUC achieved through fitting of bivariate models, where the combination of mean DBP with MAP α_1 and PI α_2, for example, exhibited respective scores of 0.871 and 0.921. This would suggest that relevant, non-redundant information may be captured using linear and DFA-based approaches in the prediction of IVH. The short-term scaling exponents for the beat-to-beat MAP and SBP, along with mean DBP were shown to be relatively robust markers in prediction of IVH for the studied cohort. Although all of the studied infants had triggered ventilation modes (synchronised intermittent positive pressure ventilation), breathing termination was not employed and so the potential for adverse patient ventilator interaction was possible. Previous work examining the impact of patient-ventilator asynchrony indicates the significant potential for IVH with this phenomenon[18,19]. These results align with those previously reported[9] and the altered vagal nerve activity in infants with IVH[20]. From further evaluation of the bivariate models, it was clear that the initial analysis did not necessarily translate to robust and consistent performance in leave-one-out cross-validation. The model with the highest AUC in the initial analysis achieved an AUC_{LOOCV} of 0.821 (0.66, 0.99), though it did not exhibit a statistically significant improvement on the univariate

Feature 1	Feature 2	AUC (95% CI)	p	LR	AUC$_{LOOCV}$
SBP α_1	DBP μ	0.843 (0.68, 1.01)	0.009	2.86	0.721 (0.50, 0.94)*
PI α_1	DBP μ	0.843 (0.69, 1.00)	0.014	2.86	0.643 (0.40, 0.89)
DBP α_1	DBP μ	0.864 (0.72, 1.00)	0.022	2.86	0.750 (0.56, 0.94)
PI α_2	MAP μ	0.864 (0.72, 1.01)	0.068	3.43	0.679 (0.46, 0.89)
MAP α_1	DBP μ	0.871 (0.74, 1.01)	0.027	3.43	0.743 (0.55, 0.94)
PI α_2	DBP μ	0.921 (0.82, 1.02)	0.035	4.00	0.821 (0.66, 0.99)

Table 4. Multivariable logistic regression models. Features included mean (μ), short- and long-term scaling exponents (α_1 and α_2, respectively) for mean arterial (MAP), systolic (SBP) and diastolic (DBP) blood pressure, as well as pulse interval (PI) time series. LR denotes the positive likelihood ratios, the 95% confidence intervals (CI) are reported for the AUC. p values are derived from the Delong comparison[14] with the non-detrended mean MAP model. The corresponding AUCs from leave-one-out cross-validation (AUC$_{LOOCV}$) are also reported, where *denotes a statistically-significant ($p < 0.05$) difference from the Delong comparison of the LOOCV mean DBP model[14].
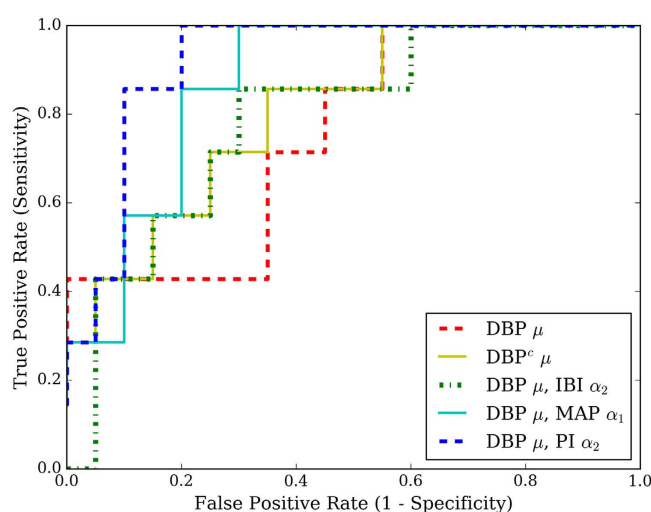

Figure 2. Receiver-Operator Characteristic Curves. These show the ROC curves for the non-detrended univariate mean DBP (DBP μ) model, the impact of detrending this feature (DBP^c μ), the addition of a respiratory feature (IBI α_2) as well as two of the highest-scoring models (DBP μ combined with MAP α_1 and PI α_2, respectively).

reference model (mean DBP)[14]. It was interesting to note the inclusion of pulse interval-based features in the highest-scoring model in both the initial analysis and cross-validation, given its use as an estimate of heart rate variability and the reported high correlation between the two[21,22]. The accuracy of this estimation however, has not been clarified, particularly in the neonatal context, though electrocardiogram-based heart rate variability has been found to offer useful information in distinguishing infants with and without IVH[8].

Addition of respiratory signals. In this study, we found that the addition of respiratory signals did not considerably improve model performance. The fractal dynamics of respiration have been applied in the context of preterm infants[23,24], though not with respect to IVH. As for the models fitted with interbreath interval (IBI) features, mechanical ventilation may have contributed to their observed lack of prediction capacity ($p > 0.05$), despite the relevance of respiration mechanics in the development of IVH[18]. It is also possible that the IBI-based features were not suited to characterising patient-ventilator asynchrony.

Clinical significance and application. Hypercarbia, high ventilator pressure and patency of the ductus arteriosus are among the factors and events that may influence the fluctuation of blood pressure of preterm infants in the neonatal intensive care unit[25]. Infants who later developed IVH exhibited lower mean DBP and a higher DBP α_2 ($p < 0.05$) across the entire recording in this study. Recent studies have reported a range of observations pertaining to blood pressure and IVH, with the main focus on characterising cerebral perfusion. These include reports of IVH being associated with the elevated diastolic closing margin[17] and significant deviation above a defined optimal MAP value in infants who later developed IVH[26].

This approach may be applied to a clinical context in a manner similar to that shown in Fig. 4, offering examples of both correct and incorrect classification of IVH from the studied cohort. A threshold may be defined according to the dashed line in each of the cases (a) to (d), where calculated probabilities exceeding this threshold
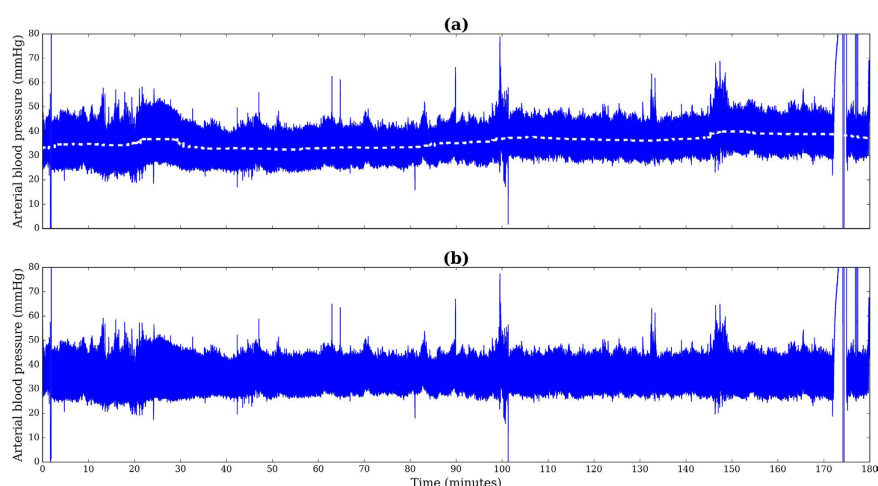
Figure 3. Detrending of overall segment. (**a**) shows the original signal and the corresponding non-linear trend, while (**b**) displays the signal after removal of this trend.

could flag infants at high-risk of developing IVH. Further model evaluation requires validation on a larger and more balanced cohort to estimate the prediction error and support its potential application in a clinical context.

Limitations.   We acknowledge that this study was limited by the size of the dataset (n = 27) as well as representation of IVH (29.6%), slightly lower than the referenced 35–45% of incidence reported in neonatal care facilities[27]. Model evaluation was also limited by the low number of IVH cases (n = 8), though our LOOCV and sensitivity analyses showed the main findings to be consistent. Another limiting factor was the signal quality of the recordings which was managed by implementing quality control measures as part of the feature extraction process.

Conclusion

In conclusion, this study found mean DBP and short-term scaling exponents from beat-to-beat MAP, DBP and SBP to be useful markers in the prediction of IVH in preterm infants. Non-linear trend removal and the inclusion of additional features such as the short-term scaling exponent (α_1) of MAP was able to improve model performance. Of the models evaluated, the one that performed consistently in both the initial analysis and cross-validation was fitted with mean DBP and PI α_2. In a clinical context, such an approach to signal processing and predictive monitoring could be applied, where a running 10 min window could continuously evaluate the relevant features from qualifying segments of data. Following evaluation in a larger population, these features may be helpful in identifying infants at high-risk of developing IVH, offering caregivers more time to adjust intensive care treatment.

Methods

Data Collection.   Physiological data was collected from the infants within 1–3 hours of birth as part of a prospective clinical investigation at a large tertiary neonatal intensive care unit in Sydney, Australia. The study was approved by the Sydney West Area Health Service Human Research and Ethics and conducted according to the World Medical Association Declaration of Helsinki. Informed parental consent was obtained in all cases.

Inclusion criteria for the cohort comprised low birthweight (<1500 g), gestational age (<30 weeks) and an absence of significant congenital anomalies. Of the 46 infants enrolled, 27 infants had arterial blood pressure and air flow wave recordings with sufficiently long, artefact-free segments. The average (SD) length of recording was 156 (34) mins. Intra-arterial blood pressure was measured using an umbilical or peripheral arterial catheter, following single-point calibration to atmospheric pressure, collected using a bedside patient monitor (Philips Agilent Systems, Philip Healthcare, North Ryde, Australia), while the raw air flow wave was acquired from a ventilator (Babylog 8000, Drägerwerk, Lübeck, Germany). Both signals were sampled at 1 kHz and recorded by an analog data acquisition system (ADInstruments, Sydney, Australia). Cranial ultrasounds were performed at 2, 12, 24 and 36 hours then daily for the first week. The presence and grade of IVH was determined according to the Papile system[27].

Signal Processing and Data Analysis.   Signal processing and feature extraction was completed in Python (Python Software Foundation, version 2.7. https://www.python.org/). Each of the arterial blood pressure and air flow signals were down-sampled to 125 Hz prior to analysis for computational efficiency. This frequency was sufficient for peak detection in both respiratory and blood pressure signals. From the downsampled signals, the following time series were extracted; the beat-to-beat MAP, SBP, DBP and PI, as derived from arterial blood pressure, as well as IBIs derived from air flow data. The signal quality constraints of the air flow data limited extraction of other respiratory features such as peak flow.

Only the arterial blood pressure signal was found to exhibit significant drift, defined by non-linear trends in the diastolic and systolic blood pressure ranges. Thus, the detrending was applied solely to this signal, as
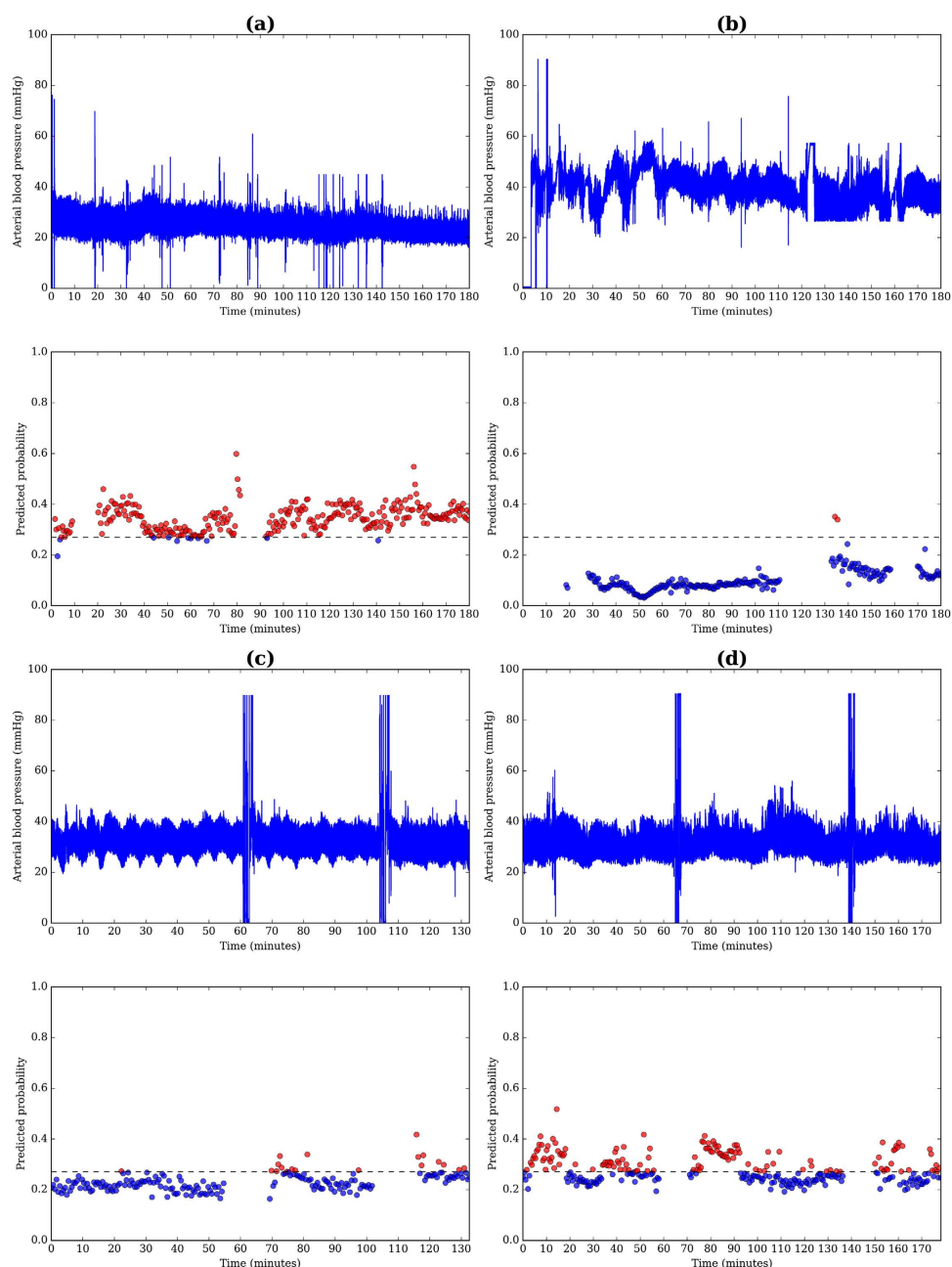
Figure 4. Arterial blood pressure data and the predicted probability for IVH using the highest scoring model, mean DBP and PI α_2 for correct classification of (**a**) IVH and (**b**) non-IVH, as well as incorrect classification of (**c**) IVH and (**d**) non-IVH. The threshold for classifying IVH, designated by the dashed line was defined at 90% specificity and 85% sensitivity. Red and blue markers represent windows that exceeded and did not exceed the threshold, respectively.

shown in Fig. 3. Such a correction would also minimally impact the derivation of the IBI-based features from the air flow signal. The overall trend of each signal was determined using a large-window median filter (window width = 1000 ms) on a further downsampled signal and the mean-centred trend was subsequently removed from the original arterial blood pressure signal. This approach was similar to the baseline wander removal that has been applied widely prior to feature extraction from the electrocardiogram signal[28]. An example of this detrending process is shown in Fig. 3.

The features used in IVH prediction were extracted from a running 10 min window of arterial blood pressure and air flow data, shifted in 30 sec increments across the total recording length. This approach was adopted to simulate the application of these techniques in a clinical setting, where windows which fulfilled the quality criteria were included for feature extraction. This criteria comprised defined ranges for the allowable number of beats and breaths in a given window (40–250 beats per minute and >20 breaths per minute), a maximum limit for an absence of detected beats (15 sec) as well as an absence of large spikes in the arterial blood pressure signal (range

47

of beat-to-beat SBP <30 mmHg). For each respective time series, outliers were removed by imposing a maximum change of 150% from the previous data point and also a maximum loss of 30% for each window. Features including the mean (μ), short- and long-term scaling exponents (α_1, α_2, respectively) from DFA of the five time series (MAP, SBP, DBP, PI and IBI) were subsequently extracted.

Developed by Peng and co-workers[10], DFA is able to quantify long-range power law correlations and accommodate for confounding non-stationarities often seen in real-world signals. It does this through the detrending, that is, linear trend removal, step prior to calculating the root-mean squared fluctuation as defined in equation 1.

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n]^2}$$

(1)

where $y(k)$ is any given time series, $y_n(k)$ the local linear trend for a given segment, and N the number of data points in the series for a given round of analysis. The application of DFA is further explained by Thamrin *et al.*[11].

The scaling exponent α is calculated from the gradient of the line fitted to the Log-Log relationship between n and F(n). In this case, the short-term scaling exponent was defined over 4–15 beats, as aligned with similar observations of this data[9] and similarly defined for heart rate variability analysis of preterm infants[8]. The long-term scaling exponent was determined across 15–50 beats.

Model Fitting and Evaluation.
Statistical analysis was completed using R 3.3.1 software[29]. Logistic regression models were used to fit the mean of extracted predictors across all qualifying 10 min windows, while the AUC was used to assess accuracy in predicting IVH. Fitted models were evaluated using leave-one-out cross-validation, where the predicted probability of each test sample was subsequently compiled and used to generate a ROC curve for performance comparison.

References
1. Bolisetty, S. *et al.* Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants. *Pediatrics* peds–2013 (2013).
2. O'Leary, H. *et al.* Elevated cerebral pressure passivity is associated with prematurity-related intracranial hemorrhage. *Pediatrics* **124,** 302–309 (2009).
3. Fanaroff, A. A. *et al.* Trends in neonatal morbidity and mortality for very low birthweight infants. *American journal of obstetrics and gynecology* **196,** 147–e1 (2007).
4. Landmann, E., Misselwitz, B., Steiss, J. O. & Gortner, L. Mortality and morbidity of neonates born at less than 26 weeks of gestation (1998–2003). A population-based study. *Journal of perinatal medicine* **36,** 168–174 (2008).
5. Ment, L. R. *et al.* Risk factors for early intraventricular hemorrhage in low birth weight infants. *The Journal of pediatrics* **121,** 776–783 (1992).
6. van Ravenswaaij-Arts, C. M. *et al.* The influence of respiratory distress syndrome on heart rate variability in very preterm infants. *Early human development* **27,** 207–221 (1991).
7. Hanna, B. *et al.* Heart rate variability in preterm brain-injured and very-low-birth-weight infants. *Neonatology* **77,** 147–155 (2000).
8. Tuzcu, V., Nas, S., Ulusar, U., Ugur, A. & Kaiser, J. R. Altered heart rhythm dynamics in very low birth weight infants with impending intraventricular hemorrhage. *Pediatrics* **123,** 810–815 (2009).
9. Zhang, Y. *et al.* Detrended fluctuation analysis of blood pressure in preterm infants with intraventricular hemorrhage. *Medical & biological engineering & computing* **51,** 1051–1057 (2013).
10. Peng, C.-K., Havlin, S., Stanley, H. E. & Goldberger, A. L. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **5,** 82–87 (1995).
11. Thamrin, C. & Stern, G. New methods: what do they tell us? Fluctuation analysis of lung function. *Eur Respir Mon* **47,** 310–324 (2010).
12. Fairchild, K. *et al.* Abnormal heart rate characteristics are associated with abnormal neuroimaging and outcomes in extremely low birth weight infants. *Journal of Perinatology* **34,** 375–379 (2014).
13. Luque, M. *et al.* A risk prediction model for severe intraventricular hemorrhage in very low birth weight infants and the effect of prophylactic indomethacin. *Journal of Perinatology* **34,** 43–48 (2014).
14. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 837–845 (1988).
15. LeDell, E., Petersen, M. & van der Laan, M. Computationally efficient confidence intervals for cross-validated area under the roc curve estimates. *Electronic journal of statistics* **9,** 1583 (2015).
16. Gönen, M. *Analyzing receiver operating characteristic curves with SAS* (SAS Institute, 2007).
17. Rhee, C. J. *et al.* Elevated diastolic closing margin is associated with intraventricular hemorrhage in premature infants. *The Journal of pediatrics* (2016).
18. Perlman, J. M., McMenamin, J. B. & Volpe, J. J. Fluctuating cerebral blood-flow velocity in respiratory distress syndrome: relation to the development of intraventricular hemorrhage. *New England Journal of Medicine* **309,** 204–209 (1983).
19. Perlman, J. M., Goodman, S., Kreusser, K. L. & Volpe, J. J. Reduction in intraventricular hemorrhage by elimination of fluctuating cerebral blood-flow velocity in preterm infants with respiratory distress syndrome. *New England Journal of Medicine* **312,** 1353–1357 (1985).
20. Koketsu, N., Moskowitz, M. A., Kontos, H. A., Yokota, M. & Shimizu, T. Chronic parasympathetic sectioning decreases regional cerebral blood flow during hemorrhagic hypotension and increases infarct size after middle cerebral artery occlusion in spontaneously hypertensive rats. *Journal of Cerebral Blood Flow & Metabolism* **12,** 613–620 (1992).
21. Schäfer, A. & Vagedes, J. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *International journal of cardiology* **166,** 15–29 (2013).
22. Bulte, C. S., Keet, S. W., Boer, C. & Bouwman, R. A. Level of agreement between heart rate variability and pulse rate variability in healthy individuals. *European Journal of Anaesthesiology (EJA)* **28,** 34–38 (2011).
23. Baldwin, D. N. *et al.* Effect of sighs on breathing memory and dynamics in healthy infants. *Journal of Applied Physiology* **97,** 1830–1839 (2004).
24. Frey, U., Silverman, M., Barabasi, A. & Suki, B. Irregularities and power law distributions in the breathing pattern in preterm and term infants. *Journal of Applied Physiology* **85,** 789–797 (1998).
25. Ballabh, P. Intraventricular hemorrhage in premature infants: mechanism of disease. *Pediatric research* **67,** 1–8 (2010).
26. da Costa, C. S. *et al.* Monitoring of cerebrovascular reactivity for determination of optimal blood pressure in preterm infants. *The Journal of pediatrics* **167,** 86–91 (2015).

27. Papile, L.-A., Burstein, J., Burstein, R. & Koffler, H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. *The Journal of pediatrics* **92,** 529–534 (1978).
28. Łeski, J. M. & Henzel, N. Ecg baseline wander and powerline interference reduction using nonlinear filter bank. *Signal processing* **85,** 781–793 (2005).
29. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2016).

Acknowledgements

Author Contributions

A.M. and M.T. conceived the experiment(s), M.T. and M.H. collected the data, J.H., C.N. analyzed the results with technical advice from C.T. All authors reviewed and provided scientific input on the manuscript.

Additional Information

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Huvanandana, J. *et al.* Prediction of intraventricular haemorrhage in preterm infants using time series analysis of blood pressure and respiratory signals. *Sci. Rep.* **7**, 46538; doi: 10.1038/srep46538 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.2   Concluding remarks

The work presented in this chapter addresses the second aim of the dissertation to explore the potential for applying signal processing and variability analysis in the early identification of preterm infants at risk of IVH. In the NICU setting, previous work has uncovered the value in sophisticated time series analysis of physiological signals for the prediction of IVH within the first few hours of birth [20, 27]. In this chapter, the thesis builds on the work by demonstrating that prediction of IVH using DFA can be improved by careful pre-processing of the signals and combining both linear and non-linear measures of blood pressure variability, while showing that the inclusion of respiratory features has limited benefit. The predictive performance of developed models also represents an improvement on the initial non-detrended counterparts (Appendix I).

The analysis and model combining mean diastolic blood pressure (DBP μ) with the long-term pulse interval scaling exponent (PI α_2) requires validation on a larger and more balanced dataset. There are currently no interventions for IVH: following diagnosis, preterm infants may receive monitoring for hydrocephalus, haemodynamic and respiratory support as well as supportive care [28]. While such a model may help to identify high-risk infants and potential predictive variables, a greater understanding of the mechanisms leading to injury is required.

In this chapter, the thesis found that the combination of both linear and non-linear metrics holds predictive value in identifying infants at risk of later brain injury. The techniques for analysing physiological signals are both diverse in approach and application; they may assist in detection of a functional outcome or clinically-important events, though they may also facilitate an understanding of the mechanisms behind certain treatments and therapies routinely used in the NICU. Pharmacological [29] and animal studies [30] have facilitated the interpretation of scaling exponents from DFA for both heart rate and beat-to-beat blood pressure time series. The metrics from Poincare analysis have similarly been shown to reflect parasympathetic and sympathetic influences on these time series [31, 32]. In the next chapter, we apply similar techniques to characterise the acute cardiovascular effects of intravenous caffeine, one of the most commonly administered drugs in the NICU [33].

Chapter 4

Characterising changes in cardiovascular dynamics following caffeine therapy

The prematurity of the patients admitted to the NICU means that their brain regulated respiratory systems are often insufficiently developed, predisposing them to significant risk of apnoea and complex interactions between the resultant treatments and concurrent respiratory conditions. While the incidence of respiratory distress syndrome has decreased markedly with routine use of surfactant, apnoea of prematurity and recurrent hypoxic events are frequently associated with bradycardia/tachycardia and pathological swings in blood pressure. These factors may place infants at risk of later morbidity, including white matter injury and IVH.

The long-term clinical outcomes of caffeine therapy [11, 12, 21, 22] and the acute effects on linear measures of variability [23,25] have been previously studied, though there has been limited work using non-linear analysis techniques. These techniques may be more suited to evaluating complex cardiovascular behaviour. This chapter addresses the following research question:

4: What are the acute affects of a loading dose of intravenous caffeine on the dynamics of heart rate and arterial blood pressure?

We employ a range of advanced analysis techniques reviewed in Chapter 2 to characterise the cardiovascular impact of caffeine therapy on preterm infants.

The content presented in this chapter is currently published [34] as:

Huvanandana, J., Thamrin, C., Hinder, M., McEwan, A. and Tracy, M. 2018. Cardiovascular impact of intravenous caffeine in preterm infants. *Acta Paediatrica*

4.1   Cardiovascular impact of intravenous caffeine in preterm infants

Statement of contributions of joint authorship

- Jacqueline Huvanandana (Candidate): data and statistical analyses, writing, reviewing and editing of the manuscript

- Cindy Thamrin: provided technical advice on the analysis, reviewing and editing the manuscript

- Alistair McEwan (Principal Supervisor): provided technical advice on the analysis, reviewing and editing the manuscript

- Murray Hinder: data collection, reviewing and editing the manuscript

- Mark Tracy: corresponding author, provided the main idea, data collection, reviewing and editing the manuscript

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Alistair McEwan   **Date:** 24 May 2018

Cardiovascular impact of intravenous caffeine in preterm infants

Jacqueline Huvanandana[1,2], Cindy Thamrin[2], Alistair L. McEwan[1,3], Murray Hinder[1,4], Mark B. Tracy (mark.tracy@sydney.edu.au)[4,5] (iD)

1. School of Electrical and Information Engineering, University of Sydney, Sydney, NSW, Australia
2. Woolcock Institute of Medical Research, University of Sydney, Sydney, NSW, Australia
3. Cerebral Palsy Alliance, Sydney, NSW, Australia
4. Neonatal Intensive Care Unit, Westmead Hospital, Sydney, NSW, Australia
5. School of Paediatrics and Child Health, University of Sydney, Sydney, NSW, Australia

ABSTRACT

Aim: To evaluate the acute effect of intravenous caffeine on heart rate and blood pressure variability in preterm infants.

Methods: We extracted and compared linear and nonlinear features of heart rate and blood pressure variability at two time points: prior to and in the two hours following a loading dose of 10 mg/kg caffeine base.

Results: We studied 31 preterm infants with arterial blood pressure data and 25 with electrocardiogram data, and compared extracted features prior to and following caffeine administration. We observed a reduction in both scaling exponents (α_1, α_2) of mean arterial pressure from detrended fluctuation analysis and an increase in the ratio of short- (SD1) and long-term (SD2) variability from Poincare analysis (SD1/SD2). Heart rate variability analyses showed a reduction in α_1 (mean (SD) of 0.92 (0.21) to 0.86 (0.21), p < 0.01), consistent with increased vagal tone. Following caffeine, beat-to-beat pulse pressure variability (SD) also increased (2.1 (0.64) to 2.5 (0.65) mmHg, p < 0.01).

Conclusion: This study highlights potential elevation in autonomic nervous system responsiveness following caffeine administration reflected in both heart rate and blood pressure systems. The observed increase in pulse pressure variability may have implications for caffeine administration to infants with potentially impaired cerebral autoregulation.

INTRODUCTION

Caffeine therapy is prescribed predominantly to prevent and treat apnoea of prematurity in neonatal intensive care. In the short term, it reduces the frequency of apnoea, intermittent hypoxia and use of assisted ventilation (1,2). Part of the methylxanthine group, caffeine is a nonspecific inhibitor of adenosine receptors (A1 and A2a receptors) (3). However, there are conflicting reports on the effects of caffeine on cardiac function: some studies report increased (4) or no significant changes (5,6) to left ventricular output. The Caffeine for Apnoea of Prematurity (CAP) trial was a large, randomised, placebo-controlled trial to evaluate the outcomes of caffeine therapy for apnoea of prematurity in very low birthweight infants over the short and long term (7,8). Caffeine improves neurocognitive outcomes at corrected 18–21 months and reduces incidence of cerebral palsy and risk of motor impairment (8). These differences

did not persist at five years follow-up (9), although at 11 years, caffeine therapy was associated with a reduced risk of motor impairment (10).

Our group examined the effects of a 10 mg/kg loading dose of caffeine base on cerebral oxygenation and cerebral blood flow velocity in a group of 40 preterm neonates in 2010. We showed significant reductions in Doppler cerebral blood flow velocity and cerebral tissue oxygenation measured via near-infrared spectroscopy (6).

Previous work has examined the clinical outcomes of caffeine therapy (8,11), the impact of timing and dosage (12–14), as well as its acute effect on various physiological

Abbreviations
ABP, Arterial blood pressure; CAP, Caffeine for apnoea of prematurity; DFA, Detrended fluctuation analysis; ECG, Electrocardiogram; HRV, Heart rate variability; MAP, Mean arterial pressure; PI, Pulse interval; PP, Pulse pressure; SD, Standard deviation; SE, Standard error.

Key Notes
- Caffeine administration increases beat-to-beat pulse pressure variability
- A standard loading dose of caffeine alters nonlinear dynamics of heart rate and blood pressure variability, increasing SD1/SD2 of mean arterial pressure and decreasing α_1 of heart rate
- Caffeine likely increases autonomic nervous system responsiveness as shown by detrended fluctuation analysis and Poincare analysis

1

variables (5,6,15). These effects are generally quantified using linear statistics such as mean and standard deviation (SD) which may not adequately capture the complex and nonlinear behaviour of physiological systems. Ulanovsky et al. (15) found no effect of 15–20 mg/kg loading dose of caffeine citrate on nonlinear dynamics of the heart rate of 21 preterm infants. This finding may well represent a type II error due to the small sample size.

Pharmacological studies (16) have contributed to the understanding of autonomic cardiac control on heart rate and blood pressure dynamics. In the context of heart rate variability (HRV), high frequency components (or short time windows) of heart rate are mainly modulated by parasympathetic activity. Both sympathetic and parasympathetic activity can be characterised using measures obtained from the Poincare plot (a plot of each data point n against the consecutive point n + 1 in a time series) (17). These measures include variability in the direction perpendicular (SD1) and parallel (SD2) to the line of identity, reflecting short- and long-term variability, respectively. The ratio SD1/SD2 is thought to reflect sympathovagal balance (17). Detrended fluctuation analysis (DFA) is one method of nonlinear time series analysis developed to characterise fluctuations over a range of scales (18). This technique has been applied to characterising heart rhythm dynamics during development (19) and prior to impending intraventricular haemorrhage (20). Studies using selective autonomic blockade in adult human (16) subjects have helped to clarify the effect of sympathetic and parasympathetic activity on the scaling exponents from DFA.

In this study, we aimed to characterise the cardiovascular impact of a 10 mg/kg loading dose of caffeine base using both linear and nonlinear variability analysis, newly applied to our published data set (6). We hypothesise that caffeine administration would have acute effects on the autonomic nervous system and blood pressure control, which are evident from changes in DFA and Poincare measures of heart rate and blood pressure variability.

PATIENTS AND METHODS
Data collection
This study was approved by the Western Sydney Area Health Service Human Research and Ethics Committee (ethics number: 06/062) and informed parental consent was obtained in all cases. Physiological data from preterm infants (gestational age <34 weeks) requiring caffeine therapy were collected between August and December 2006. Eligibility criteria required that caffeine therapy was initiated for either weaning from mechanical ventilation, reducing extubation failure risk or treatment of apnoea of prematurity. Infants with significant congenital anomalies and high-grade peri-intraventricular haemorrhage at the time of study were excluded.

A loading dose of 10 mg/kg caffeine base was delivered intravenously to all enrolled infants over half an hour. Data collection commenced at least 20 minutes prior to caffeine administration, and concluded at least four hours following

dose completion. Time series analysis focused on two time points (i) prior to dose administration and (ii) in the two hours following dose completion, during which plasma concentrations of caffeine is presumed to be greatest (21).

After single-point calibration to atmospheric pressure, intra-arterial blood pressure data were measured via an arterial line (umbilical or peripheral catheter) and collected using a bedside patient monitor (NCMS Philips Agilent System, Philips Healthcare, North Ryde, Australia). ECG data were acquired using the same patient monitor. ECG was not monitored for certain infants to preserve skin integrity.

Preprocessing
Signal preprocessing, feature extraction and subsequent analyses were undertaken in Python (Python Software Foundation, version 2.7 https://www.python.org/). The arterial blood pressure signal was down-sampled to 200 Hz. While DFA is generally robust against nonstationarities, it can be susceptible to longer term trends. Thus, we corrected for baseline drift characterised by a series of median filters (widths of 100, 300, 500 and 1000 ms) and the analysis techniques were then applied to the detrended signal. The ECG signal was similarly down-sampled to 200 Hz for computational efficiency and preprocessed using successive median filters of widths 200 and 600 ms to characterise the baseline wander (22).

Feature extraction
From the arterial blood pressure signal, we extracted the beat-to-beat mean arterial pressure, pulse pressure (systolic$_n$-diastolic$_n$ blood pressure for beat n) and the interval between successive systolic beats (pulse interval). From the ECG signal, we extracted the RR intervals following R-peak identification with Hilbert transform-based QRS detector algorithm (22). From each of these time series, the linear (mean, SD) and nonlinear (DFA: α_1 and α_2, Poincare: SD1, SD2 and SD1/SD2) features were extracted, as explained further below.

Detrended fluctuation analysis characterises the absence or presence of long-range correlations in a signal (18) and is well-suited to cope with nonstationarities (that is, constantly changing statistical properties), which are often present in physiological signals (23). DFA takes the cumulative sum of the mean-centred signal y, divides it into boxes of equal size n, detrends each box k then determines the root mean square fluctuation within the box (Equation 1). This process is repeated for varying box sizes n. The scaling exponent α is defined as gradient of the log–log relationship between box size n and corresponding root mean square fluctuation F(n).

$$F(n) = \left(\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2 \right)^{\frac{1}{2}} \tag{1}$$

A scaling exponent $\alpha = 0.5$ is characteristic of a completely random series and an increasing α indicates

increasing long-range correlations. For DFA of the blood pressure time series, we partitioned the relationship into short (α_1) and long-term (α_2) scaling exponents, defined respectively as $2 \leq n \leq 30$ and $35 \leq n \leq 200$, consistent with previous work reported for blood pressure (24). For HRV, these were defined as $2 \leq n \leq 16$ and $32 \leq n \leq 200$, respectively.

Poincare analyses offer a means of visualising short and long-term variability, displaying each data point n in a time series with respect to its neighbouring data point n + 1, for all points. These recurrence plots can be quantified in terms of SD1 (the SD in the direction perpendicular to the line of identity, a measure of short-term variability) and SD2 (the SD in the direction along it, a measure of long-term variability (17)), while the SD1/SD2 ratio reflects the balance between short- and long-term variability.

These features were extracted from a running 10-minute window of the respective time series, shifted in increments of 30 seconds. Quality criteria included (i) mean pulse rate of 40–250 beats per minute, (ii) a maximum of 10 seconds without detected peaks and (iii) a maximum loss of 20% for each time series following removal of outliers (>150% change from previous data point). For eligible windows, outlier removal affected <2.5% of all data points. Two additional criteria were applied to the arterial blood pressure signal: (iv) a maximum period of 10 seconds with sub-zero signal and (v) a maximum of 20% of the signal showing evidence of clipping (<0.05 mV change between subsequent systolic blood pressure points).

Statistical analysis was performed using R version 3.3.1 (R Core Team, 2012) and lme4 (Bates, Maechler & Bolker, 2012). We used a linear mixed-effects model to analyse the relationship between the extracted features and caffeine, adjusting for gestational age and birthweight Z-score calculated from Fenton growth charts (25) (fixed effects) and including intercepts for subjects as a random effect. Residual plots showed no curvature or pattern, with no obvious deviations from normality. Statistically significant variables (defined as $p < 0.05$) were confirmed with likelihood ratio tests of the full model against the model without the effect in question.

Sensitivity analyses
There were 11 subjects with arterial blood pressure data and six subjects with ECG data that had insufficient high-quality data at both time points (prior to and following caffeine administration). As a sensitivity analysis, we restricted the data set to the 20 and 19 subjects with complete arterial blood pressure and ECG data, respectively. We then determined the mean feature for each subject and compared pre- and postcaffeine values using a paired t-test or Wilcoxon test, depending on confirmed normality (Shapiro–Wilk test, p > 0.05).

RESULTS
Of the 40 infants enrolled in the previously published study (6), 31 had available arterial blood pressure data and 25 had

ECG data (16 infants with both signals available). The characteristics of both subsets are summarised in Table 1. Mean (SD) length of data collected for all infants was 365 (176) minutes.

Table 2 summarises the mean and SD from all infants for features from linear statistical analysis, Poincare analysis and DFA. Figure 1 summarises the α_1 and SD1/SD2 ratio changes across each of the four physiological time series. Linear mixed-model coefficients are summarised in Table S1. Following caffeine administration, mean arterial pressure α_2 decreased significantly, while SD1/SD2 from Poincare analysis increased. An example of increased an increased SD1/SD2 ratio is shown in Figure S1. These changes were driven predominantly by an increased SD1, a marker of short-term variability. Mean arterial pressure α_1 also exhibited a significant albeit mild decrease following caffeine.

Mean pulse pressure was not altered significantly following caffeine administration, though SD increased. We similarly observed an increase in SD1 and SD2 with their ratio not significantly altered. For HRV, α_1 was lower at the postcaffeine time point. Figure S2 shows plots from DFA and Poincare analysis of a single analysis window of ECG data. Our cohort of preterm infants exhibited similar albeit slightly lower scaling exponents for HRV and blood pressure than those reported previously (15).

Following a sensitivity analysis, the direction of change of all variability parameters in this subset remained consistent with those reported in Table 2. Statistical significance also held for all parameters except for pulse pressure α_1 (paired t-test p = 0.063).

DISCUSSION
Caffeine is widely used in the neonatal intensive care unit as a respiratory stimulant to reduce the frequency of apnoea and to aid weaning from mechanical ventilation (1). It is thought to act as an adenosine antagonist at the A_1 and A_{2a} receptors, stimulating the central respiratory centres and increasing chemoreceptor responsiveness to hypercapnia

Table 1 Cohort characteristics for preterm infants with arterial blood pressure data (ABP subset) and with electrocardiogram data (ECG subset)

Variable	ABP subset (n = 31)	ECG subset (n = 25)
Gestational age (weeks)	27.0 (23.6–33.3)	29.0 (24.3–33.3)
Birthweight (g)	934 (552–2100)	1265 (681–2100)
Postnatal age (days)	2.0 (0.1–7.8)	1.7 (0.1–7.8)
Sex (M/F)	17/14 (54.8%:45.2%)	15/10 (60%:40%)
Intubated	30 (96.8%)	15 (60%)
CPAP	1 (3.2%)	3 (12%)
No respiratory support	0 (0%)	7 (28%)
RDS	28 (90.3%)	19 (76%)
Died	1 (3.2%)	0 (0%)

Data are presented as median (range) or number (%).
CPAP = Continuous positive airway pressure, RDS = Respiratory distress syndrome.

Table 2 Summary of extracted features of heart rate and blood pressure variability prior to caffeine administration (precaffeine) and in the two hours following dose completion (postcaffeine)

Feature	Precaffeine	Postcaffeine	Coefficient (SE)
Mean arterial pressure			
Mean (mm Hg)	40 (7.2)	41 (7.1)	0.26 (0.39)
SD (mm Hg)	2.4 (0.81)	2.5 (0.68)	0.08 (0.11)
α_1	1.12 (0.21)	1.07 (0.16)	−0.06 (0.02)*
α_2	1.0 (0.14)	0.96 (0.12)	−0.04 (0.02)*
SD1	0.99 (0.49)	1.2 (0.46)	0.21 (0.06)**
SD2	3.2 (1.1)	3.3 (0.92)	0.06 (0.15)
SD1/SD2	0.34 (0.15)	0.37 (0.14)	0.05 (0.02)**
Pulse pressure			
Mean (mm Hg)	22 (5.5)	21 (4.7)	−0.60 (0.64)
SD (mm Hg)	2.1 (0.64)	2.5 (0.65)	0.34 (0.10)**
α_1	0.68 (0.18)	0.73 (0.14)	0.05 (0.02)*
α_2	0.95 (0.14)	0.95 (0.13)	−0.01 (0.02)
SD1	1.7 (0.71)	1.9 (0.7)	0.21 (0.07)*
SD2	2.4 (0.77)	2.9 (0.79)	0.44 (0.12)**
SD1/SD2	0.76 (0.32)	0.71 (0.27)	−0.04 (0.03)
Pulse interval			
Mean (ms)	420 (29)	410 (25)	−1.40 (0.83)
SD (ms)	22 (8.2)	26 (9.4)	0.80 (0.43)
α_1	0.79 (0.26)	0.76 (0.22)	−0.04 (0.03)
α_2	0.96 (0.13)	1 (0.12)	0.06 (0.03)
SD1	2.9 (0.98)	3.3 (1.3)	0.59 (0.27)*
SD2	5.5 (2.4)	6.4 (2.5)	0.95 (0.55)
SD1/SD2	0.64 (0.27)	0.58 (0.19)	−0.03 (0.04)
Heart rate interval			
Mean (ms)	88 (6)	86 (5.4)	−1.7 (0.99)
SD (ms)	7.6 (3.6)	8.9 (3.5)	1.3 (0.8)
α_1	0.92 (0.21)	0.86 (0.21)	−0.06 (0.02)**
α_2	0.87 (0.1)	0.91 (0.17)	0.03 (0.03)
SD1	4.9 (3.1)	6.1 (3.7)	1.10 (0.69)
SD2	9.3 (4.2)	11 (3.9)	1.40 (0.93)
SD1/SD2	0.51 (0.18)	0.54 (0.2)	0.03 (0.02)

Mean (SD) features for all subjects at each time point are presented for both linear and nonlinear analyses. Linear mixed-model coefficients for the caffeine variable and standard error (SE) are also included, and statistical significance from likelihood ratio tests denoted by *p < 0.05, **p < 0.01.

(1,26). Loading of caffeine is generally commenced when the preterm infant is still invasively ventilated and requiring positive pressure ventilation, often from soon after delivery. Alterations in inspiratory and expiratory pressures with subsequent changes in mean airway pressure and blood gases impact significantly on intracranial pressure (27). The potential for additive negative impact with airway pressure changes and caffeine is not to be overlooked. Considerable uncertainty remains regarding the effect of caffeine, with conflicting evidence of cardiac, blood pressure and cerebral blood flow effects (4–6) and increased vagal or sympathetic tone quantified by HRV analyses (15,28). Caffeine may increase vagally mediated HRV, although these findings are largely reported in adults with studies with heterogeneous study design and demographics (29).

We compared both linear and nonlinear measures of heart rate and blood pressure variability at two time points;

prior to a loading dose of caffeine and in the two hours following dose completion. Using linear mixed modelling, we found that the linear measures of variability (mean, SD) for mean arterial blood pressure and heart rate did not change significantly following caffeine (Table 2). While our findings are consistent with previous studies (5,15), it may be that the linear metrics do not adequately capture altered dynamics in heart rate and blood pressure control. The decrease in α_1 of HRV following caffeine administration from a mean (SD) of 0.92 (0.21) to 0.86 (0.21) (linear mixed-effects coefficient −0.06, standard error 0.02, p < 0.01) suggests weaker long-range correlations, consistent with an increase in parasympathetic activity (16).

We observed a mild increase in variability (SD) of pulse pressure, which may be an important consideration in infants with impaired cerebrovascular autoregulation: discordance between systemic and cerebral blood flow may contribute to brain injury. Analysis of long-term outcomes from the CAP trial has nevertheless shown no adverse neurodevelopmental outcomes in infants receiving caffeine therapy rather than placebo (7,9,10).

The observed changes in heart rate and blood pressure dynamics may represent a range of influences and conditions. For example, the increase in vagal tone may simply represent a gradual relaxation of the subjects over the monitoring period or maturation of the autonomic nervous system, rather than the direct effect of caffeine. Nakamura et al. (19) reported correlations between postnatal age and the HRV scaling exponents (α_1, α_2) over several days to weeks. Given the two-hour window over which features were extracted for the second time point, the latter effect may be minor, although it can nevertheless be disregarded as the study did not include a control group of infants not receiving caffeine. These findings suggest an enhanced reactivity of the autonomic nervous system (29) following a loading dose of caffeine in preterm infants, as characterised by increased vagally mediated heart rate and blood pressure variability.

Ulanovsky et al. (15) showed nonlinear HRV metrics were unchanged in a cohort of preterm infants following a loading dose of caffeine. They compared similar metrics from a 10-minute window extracted from each time point for each subject: prior to dose administration and in the one to two hours following completion. Comparisons showed no significant alterations in these parameters which may represent a type II error. Our subset of 31 infants with arterial blood pressure data had a median (range) gestational age of 27 (23.6–33.3) weeks (6), notably more preterm than the mean (SD) 30.3 (2.5) weeks gestation in the 21 infants reported by Ulanovsky et al. (15). This earlier study also compared a single window at each time point, which may not adequately account for natural variability within each subject.

To our knowledge, this is the first study to examine changes in nonlinear dynamics of blood pressure variability following caffeine administration in preterm infants. The decrease in mean arterial pressure α_1 (1.12 (0.21) to 1.07 (0.16), linear mixed-effects coefficient −0.06, standard error
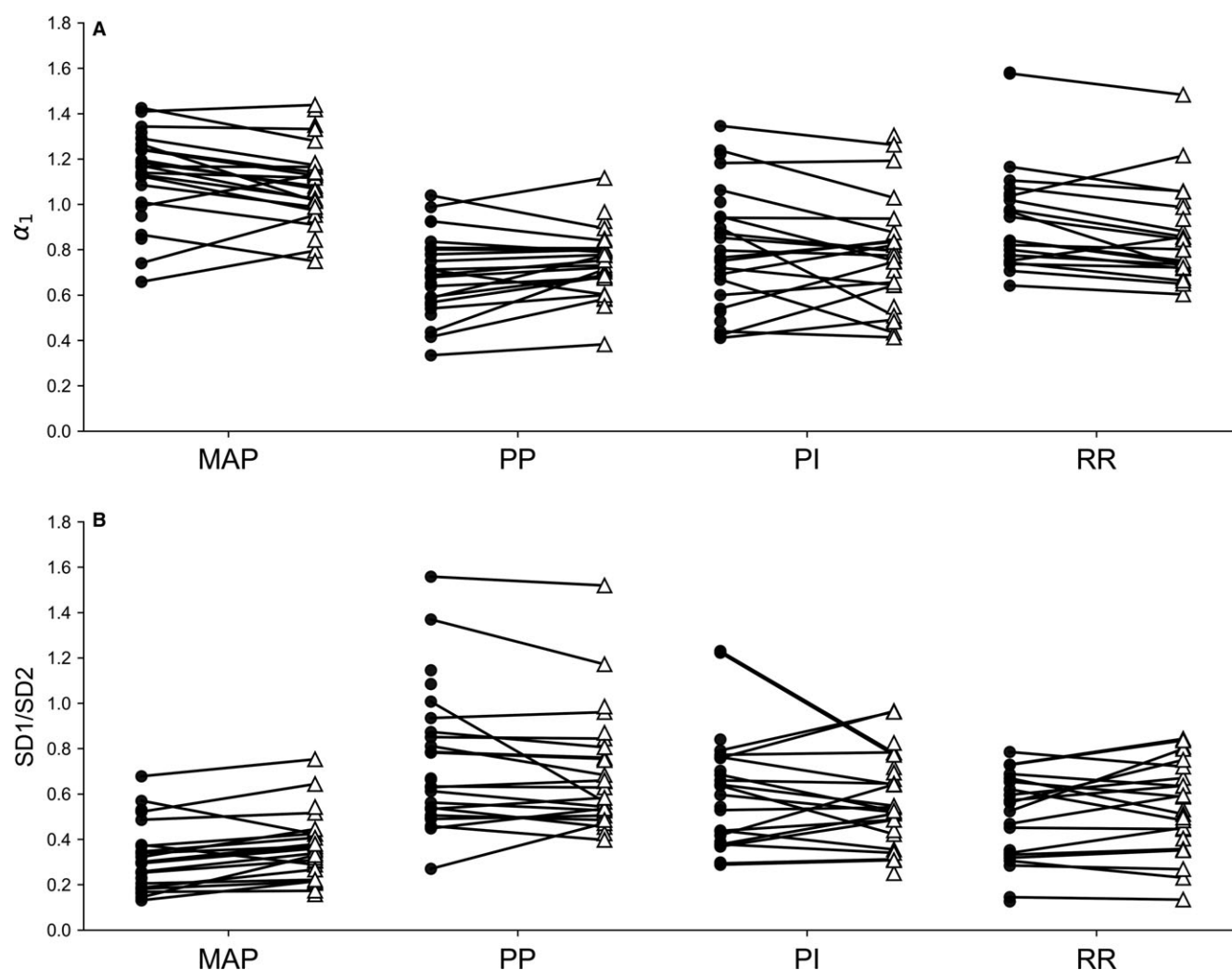
Figure 1 Precaffeine and postcaffeine plots of (A) short-term scaling exponents from detrended fluctuation analyses and (B) the SD1/SD2 ratio from Poincare analyses of the following time series: the beat-to-beat mean arterial pressure (MAP), pulse pressure (PP), pulse interval (PI) and heart rate intervals (RR). Values plotted are pre- and post-mean values of all qualifying windows for each subject. Closed circles and open triangles denote pre and postcaffeine administration, respectively.

0.02, p < 0.05) suggests an altered complexity that may also be associated with increased vagal tone. This finding was also reflected by Poincare analysis with an elevated SD1/SD2 ratio for mean arterial pressure (17,30). In adult subjects, vagal blockade by atropine has led to an increased short-term scaling exponent for the beat-to-beat blood pressure time series, while sympathetic inhibition by clonidine conversely reduced these exponents (16). It is also possible that the changes in the blood pressure scaling exponents were driven by those in heart rate: altered RR intervals influence the period of diastolic decay in the arterial pressure pulse via the Windkessel mechanism and may alter systolic blood pressure by shortening or lengthening the period of diastolic filling (16).

The altered heart rate dynamics in this cohort of infants were more clearly discerned using DFA than by Poincare analysis. The improved sensitivity of DFA over Poincare analysis for time rather than amplitude-based features also remained true of pulse interval which is often used as a proxy of heart rate, despite susceptibility to cardiorespiratory coupling. It is possible that DFA, which specifically quantifies correlations over time, may be more suited to temporal metrics such as pulse and heart rate intervals, whereas Poincare analysis, which quantifies variability, may better describe amplitude-based metrics such as mean arterial and pulse pressures.

One of the limitations of this study was the signal quality of the arterial blood pressure data, where artefacts and evidence of clipping may have influenced the extracted time series. We sought to mitigate this by applying quality control criteria and adopting maximal thresholds for abrupt variations in beat-to-beat values (quality criteria iv). Another limitation was that not all subjects had sufficiently high-quality data at both time points, although sensitivity analyses of the results with the restricted subset of data showed consistent results. We also used a linear mixed-effects model for the statistical analysis to mitigate the impact of missing values. As there was no control group of

infants not receiving caffeine therapy, there is also potential that other factors or postnatal development may have influenced the findings. This study was also limited by the sample size and the concurrent availability of ECG and arterial blood pressure data. The statistical analysis of HRV features was based on a different albeit overlapping subset of infants.

CONCLUSION
This study demonstrated that with a standard loading dose of caffeine, preterm infants showed enhanced autonomic nervous system responsiveness, reflected by indices of parasympathetic activity from both heart rate and blood pressure systems. Our findings improve our understanding of the mechanisms behind caffeine therapy. The use of advanced nonlinear analyses of heart rate and blood pressure variability used in this study may be applicable in other contexts of cardiovascular control. Our observation of increased pulse pressure variability may also hold implications for caffeine administration to infants with potentially impaired cerebral autoregulation.

FUNDING

CONFLICT OF INTERESTS
The authors declare no conflict of interest.

References

1. Henderson-Smart DJ, De Paoli AG. Methylxanthine treatment for apnoea in preterm infants. *Cochrane Library* 2010; (12): CD000140. https://doi.org/10.1002/14651858.CD000140. pub2.
2. Rhein LM, Dobson NR, Darnall RA, Corwin MJ, Heeren TC, Poets CF, et al. Effects of caffeine on intermittent hypoxia in infants born prematurely: a randomized clinical trial. *JAMA Pediatric* 2014; 168: 250–7.
3. Fredholm BB. Adenosine, adenosine receptors and the actions of caffeine. *Basic Clin Pharmacol Toxicol* 1995; 76: 93–101.
4. Soloveychik V, Bin-Nun A, Ionchev A, Sriram S, Meadow W. Acute hemodynamic effects of caffeine administration in premature infants. *J Perinatol* 2009; 29: 205–8.
5. Hoecker C, Nelle M, Poeschl J, Beedgen B, Linderkamp O. Caffeine impairs cerebral and intestinal blood flow velocity in preterm infants. *Pediatrics* 2002; 109: 784–7.
6. Tracy M, Klimek J, Hinder M, Ponnampalam G, Tracy S. Does caffeine impair cerebral oxygenation and blood flow velocity in preterm infants? *Acta Paediatr* 2010; 99: 1319–23.
7. Schmidt B, Roberts RS, Davis P, Doyle LW, Barrington KJ, Ohlsson A, et al. Caffeine therapy for apnea of prematurity. *N Engl J Med* 2006; 354: 2112–21.
8. Schmidt B, Roberts RS, Davis P, Doyle LW, Barrington KJ, Ohlsson A, et al. Long-term effects of caffeine therapy for apnea of prematurity. *N Engl J Med* 2007; 357: 1893–902.
9. Schmidt B, Anderson PJ, Doyle LW, Dewey D, Grunau RE, Asztalos EV, et al. Survival without disability to age 5 years after neonatal caffeine therapy for apnea of prematurity. *JAMA* 2012; 307: 275–82.
10. Schmidt B, Roberts RS, Anderson PJ, Asztalos EV, Costantini L, Davis PG, et al. Academic performance, motor function, and behavior 11 years after neonatal caffeine citrate therapy for apnea of prematurity: an 11-year follow-up of the CAP randomized clinical trial. *JAMA Pediatrics* 2017; 171: 564.
11. Patel R, Leong T, Carlton D, Vyas-Read S. Early caffeine therapy and clinical outcomes in extremely preterm infants. *J Perinatol* 2013; 33: 134.
12. Katheria AC, Sauberan JB, Akotia D, Rich W, Durham J, Finer NN. A pilot randomized controlled trial of early versus routine caffeine in extremely premature infants. *Am J Perinatol* 2015; 32: 879–86.
13. Katz S, Perenyi A, Parris RO, Stefanov DG. The influence of dosage and timing of caffeine administration on neurodevelopmental outcome of very preterm infants. *Neonat Pediatr Med* 2015; 1: 2.
14. Steer P, Flenady V, Shearman A, Charles B, Gray P, Henderson-Smart D, et al. High dose caffeine citrate for extubation of preterm infants: a randomised controlled trial. *Arch Dis Child Fetal Neonatal Ed* 2004; 89: F499–503.
15. Ulanovsky I, Haleluya N, Blazer S, Weissman A. The effects of caffeine on heart rate variability in newborns with apnea of prematurity. *J Perinatol* 2014; 34: 620.
16. Castiglioni P, Parati G, Di Rienzo M, Carabalona R, Cividjian A, Quintin L. Scale exponents of blood pressure and heart rate during autonomic blockade as assessed by detrended fluctuation analysis. *J Physiol* 2011; 589: 355–69.
17. Brennan M, Palaniswami M, Kamen P. Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Transact Biomed Eng* 2001; 48: 1342–7.
18. Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 1995; 5: 82–7.
19. Nakamura T, Horio H, Miyashita S, Chiba Y, Sato S. Identification of development and autonomic nerve activity from heart rate variability in preterm infants. *Biosystems* 2005; 79: 117–24.
20. Tuzcu V, Nas S, Ulusar U, Ugur A, Kaiser JR. Altered heart rhythm dynamics in very low birth weight infants with impending intraventricular hemorrhage. *Pediatrics* 2009; 123: 810–5.
21. Aranda JV, Cook CE, Gorman W, Collinge JM, Loughnan PM, Outerbridge EW, et al. Pharmacokinetic profile of caffeine in the premature newborn infant with apnea. *J Pediatric* 1979; 94: 663–8.
22. De Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Trans Biomed Eng* 2003; 50: 686–96.
23. Bryce R, Sprague K. Revisiting detrended fluctuation analysis. *Scientific Rep* 2012; 2: 315.
24. Galhardo C, Penna T, de Menezes MA, Soares P. Detrended fluctuation analysis of a systolic blood pressure control loop. *New J Phys* 2009; 11: 103005.
25. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr* 2013; 13: 59.
26. Comer AM, Perry CM, Figgitt DP. Caffeine citrate. *Paediatr Drugs* 2001; 3: 61–79.
27. Stewart A, Finer N, Peters K. Effects of alterations of inspiratory and expiratory pressures and lnspiratory/expiratory ratios on mean airway pressure, blood gases, and intracranial pressure. *Pediatrics* 1981; 67: 474–81.

28. Notarius CF, Floras JS. Caffeine enhances heart rate variability in middle-aged healthy, but not heart failure subjects. *J Caffeine Res* 2012; 2: 77–82.
29. Koenig J, Jarczok MN, Kuhn W, Morsch K, Schäfer A, Hillecke TK, et al. Impact of caffeine on heart rate variability: a systematic review. *J Caffeine Res* 2013; 3: 22–37.
30. Woo MA, Stevenson WG, Moser DK, Trelease RB, Harper RM. Patterns of beat-to-beat heart rate variability in advanced heart failure. *Am Heart J* 1992; 123: 704–10.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Examples of SD1/SD2 from Poincare analysis of mean arterial pressure showing broadening of ellipse following caffeine administration in a single infant.
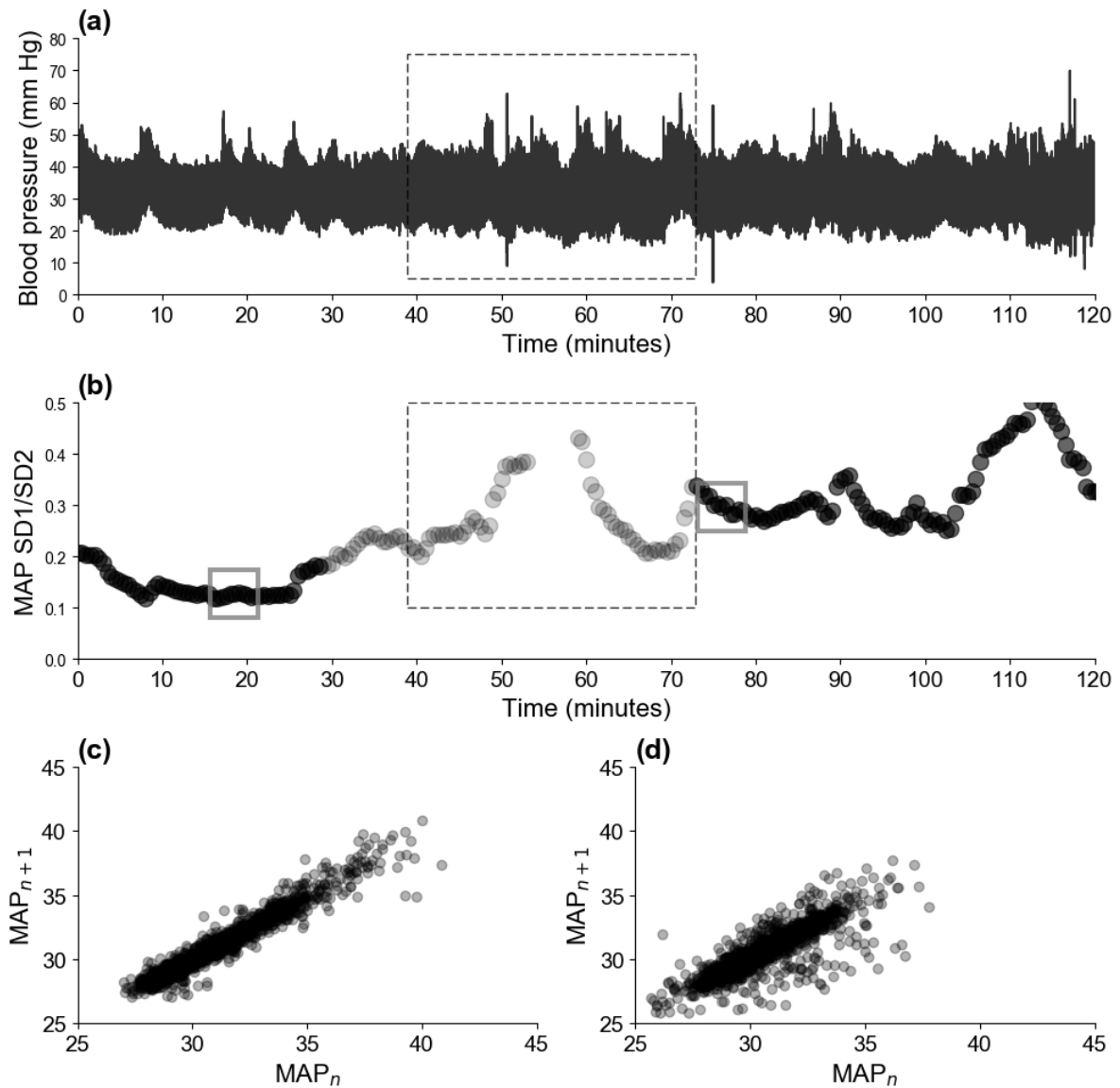
Figure S2 Examples of (a) detrended fluctuation analysis and (b) Poincare analysis of heart rate for a single analysis window.

Table S1 Summary of linear mixed model coefficients (standard error) and statistical significance for gestational age in weeks, birthweight Z-scores and caffeine (pre-caffeine = 0, post-caffeine = 1). Statistical significance is denoted by *p < 0.05, **p < 0.01.

Supplementary Table S1 Summary of linear mixed model coefficients (standard error) and statistical significance for gestational age in weeks, birthweight Z-scores and caffeine (pre-caffeine = 0, post-caffeine = 1). Statistical significance is denoted by *p < 0.05, **p < 0.01.

Feature	Gestational age	Birthweight Z-score	Caffeine
Mean arterial pressure			
Mean	0.95 (0.43)*	2.3 (1)*	0.26 (0.39)
SD	0.086 (0.049)	0.028 (0.11)	0.081 (0.11)
α_1	-0.0038 (0.014)	0.009 (0.033)	-0.056 (0.025)*
α_2	-0.028 (0.0086)**	0.0015 (0.02)	-0.042 (0.016)*
SD1	0.069 (0.032)*	0.0021 (0.075)	0.21 (0.056)**
SD2	0.1 (0.068)	0.048 (0.15)	0.055 (0.15)
SD1/SD2	0.016 (0.011)	-0.0089 (0.025)	0.049 (0.015)**
Pulse pressure			
Mean	0.12 (0.42)	0.63 (0.99)	-0.6 (0.64)
SD	0.085 (0.045)	0.02 (0.1)	0.34 (0.095)**
α_1	-0.016 (0.012)	0.0053 (0.027)	0.047 (0.021)*
α_2	-0.012 (0.0091)	0.027 (0.021)	-0.011 (0.024)
SD1	0.11 (0.048)*	-0.037 (0.11)	0.21 (0.073)*
SD2	0.067 (0.055)	0.08 (0.12)	0.44 (0.12)**
SD1/SD2	0.03 (0.02)	-0.035 (0.047)	-0.042 (0.027)
Pulse interval			
Mean	0.91 (0.39)*	0.34 (0.9)	-1.4 (0.83)
SD	0.34 (0.11)**	0.056 (0.25)	0.8 (0.43)
α_1	0.014 (0.019)	0.011 (0.044)	-0.035 (0.032)
α_2	-0.014 (0.0095)	0.018 (0.021)	0.061 (0.029)
SD1	0.23 (0.095)*	-0.097 (0.21)	0.59 (0.27)*
SD2	0.42 (0.15)*	0.12 (0.34)	0.95 (0.55)
SD1/SD2	-0.00025 (0.016)	-0.012 (0.037)	-0.03 (0.038)
Heart rate interval			
Mean	0.48 (0.45)	0.13 (1.1)	-1.7 (0.99)
SD	-0.3 (0.23)	0.73 (0.57)	1.3 (0.8)
α_1	0.052 (0.015)**	0.016 (0.037)	-0.064 (0.021)**
α_2	0.028 (0.0085)**	0.012 (0.021)	0.029 (0.031)
SD1	-0.58 (0.21)*	0.26 (0.53)	1.1 (0.69)
SD2	-0.19 (0.26)	1 (0.66)	1.4 (0.93)
SD1/SD2	-0.045 (0.013)**	-0.025 (0.031)	0.03 (0.025)

Supplementary Figure S1 Examples of SD1/SD2 from Poincare analysis of mean arterial pressure showing broadening of ellipse following caffeine administration in a single infant. Panels (a) shows the arterial blood pressure data and (b) the calculated SD1/SD2 ratio from beat-to-beat mean arterial pressure (MAP) from each 10-minute running window, where dashed lines delineate the period of caffeine base administration. Solid grey lines prior to and following caffeine correspond to the SD1/SD2 ratio extracted from Poincare plots in panels (c) and (d), respectively.

Supplementary Figure S2 Examples of (a) detrended fluctuation analysis and (b) Poincare analysis of heart rate for a single analysis window. Closed circles and open triangles denote pre- and post-caffeine base administration, respectively.
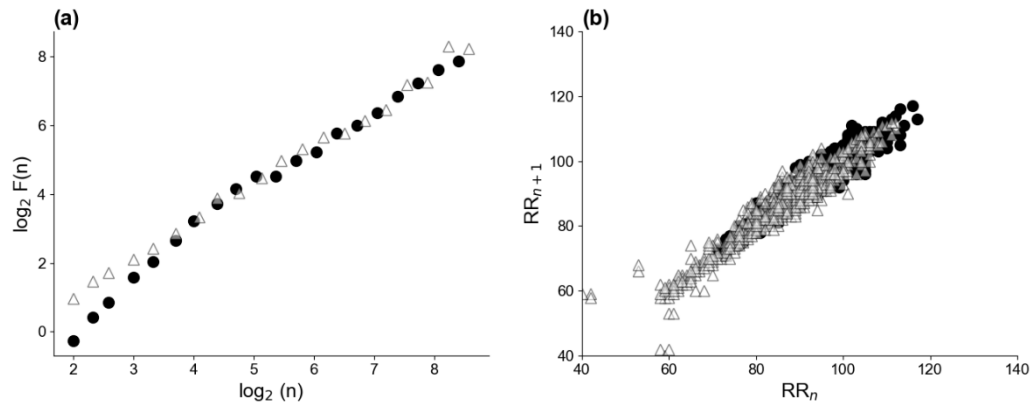
4.2 Concluding remarks

The work presented in this chapter addresses the third aim of the dissertation to characterise the cardiovascular impact of caffeine therapy in preterm infants.

Caffeine is widely prescribed as a respiratory stimulant in neonatal intensive care, with considerable work on long-term clinical outcomes [12, 21, 22] and the acute impact on blood pressure or cerebral blood flow [23, 25]. Work using non-linear analysis techniques has been limited, and where they have been applied, yielded negative results [35]. The thesis found altered dynamics in both heart rate and arterial blood pressure, consistent with increased vagal tone [31, 32]. This is important for two reasons; firstly, many extremely low birthweight infants being treated for apnoea of prematurity have impaired cerebrovascular autoregulation, thus caffeine induced variability in cardiovascular function may be deleterious. Secondly, the large multi-centre trial of caffeine versus placebo (the Caffeine for Apnoea of Prematurity (CAP) trial) showed neurodevelopmental advantage at 18-21 months [12], which largely dissipated by 5 years of age [21]. These observations in a small cohort warrant further investigation in a larger dataset. In addition to comparisons of long-term outcomes, advanced analysis of physiological variability aids our understanding of the acute effects of caffeine, and in clarifying its potential benefits or harms in preterm infants. The thesis may thus contribute to our understanding of the mechanisms behind caffeine therapy.

Both this chapter and Chapter 3 have identified the importance of both linear and non-linear approaches to characterising the behaviour of complex physiological systems such as the dynamics of heart rate and arterial blood pressure control. Multivariable techniques such as correlation analysis incorporate multiple signals, and represent another path of development in this area [28] (Chapter 2).

In the next chapter, the thesis builds on these techniques and examine the utility of correlation analysis in characterising cerebral autoregulation in preterm infants, adding to our understanding of the effects of caffeine therapy.

Chapter 5

Characterising cerebral autoregulation in preterm infants

Cerebral autoregulation is the mechanism by which cerebral blood flow is maintained relatively constant over a range of perfusion pressures. This dynamic autoregulation is sensitive to changes in arterial pressure, and has been characterised using techniques including transfer function analysis [36], coherence analysis [37] and time domain correlation [38].

Dynamic cerebral autoregulation in preterm infants is an important contributor to risk of brain injury and adverse neurodevelopmental outcomes; autoregulatory dysfunction has been postulated as a causal pathway of brain injury such as IVH in this population [39]. Near-infrared spectroscopy (NIRS) offers a means of non-invasively monitoring cerebral oxygenation, which, coupled with continuous measurements of systemic blood pressure, allows this dynamic relationship to be described. Monitoring of this relationship may play an important role in identifying infants at risk of autoregulatory dysfunction and later brain injury. It may also facilitate a greater understanding of commonly administered treatments such as caffeine therapy in the NICU, the long-term clinical outcomes of which have been previously studied in the Caffeine for Apnoea of Prematurity (CAP) trial [12, 21, 22]. This chapter addresses the following research questions:

  5: Can detrended cross-correlation analysis, a relatively novel method to jointly describe the time patterns of two signals, be used to characterise cerebral autoregulation?

  6: How does caffeine affect variability of cerebral autoregulation?

We apply a range of correlation analysis techniques to characterise cerebral autoregulation in a novel manner.

The content presented in this chapter is currently under review with Acta Paediatrica.

5.1 The effect of caffeine loading on cerebral autoregulation in preterm infants

Statement of contributions of joint authorship

- Jacqueline Huvanandana (Candidate): data and statistical analyses, writing, reviewing and editing of the manuscript

- Cindy Thamrin: provided technical advice on the analysis, reviewing and editing the manuscript

- Mark Tracy: provided the main idea, data collection, reviewing and editing the manuscript

- Murray Hinder: data collection, reviewing and editing the manuscript

- Alistair McEwan (Principal Supervisor): provided technical advice on the analysis, reviewing and editing the manuscript

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Alistair McEwan   **Date:** 24 May 2018

The effect of caffeine loading on cerebral autoregulation in preterm infants

Jacqueline Huvanandana[1,2], Cindy Thamrin[2], Murray Hinder[1,3], Alistair McEwan[1,5] Mark Tracy[3,4]

[1]School of Electrical and Information Engineering, University of Sydney, Sydney, Australia

[2]Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia

[3]Westmead Hospital, Sydney, Australia

[4]School of Paediatrics and Child Health, University of Sydney, Sydney, Australia

[5]Cerebral Palsy Alliance, Sydney, Australia

Short title: Cerebrovascular effect of caffeine in preterm infants

Corresponding author

Dr Mark Tracy

PhD MSc (Epi) MBBS FRACP CCPU

Senior Staff Specialist

Westmead Hospital Neonatal Intensive Care

Clinical Senior Lecturer Department of Paediatrics and Child Health

The University of Sydney

Phone: +61 2 9845 8911/6568

mailto:mark.tracy@sydney.edu.au

Abstract

Aim: To evaluate cerebral autoregulation changes in preterm infants receiving a loading dose of caffeine base.

Methods: Cerebral autoregulation was determined with correlation analyses of the mean arterial pressure (MAP) and tissue oxygenation index (TOI) time series. Data from two timepoints were used: prior to and in the 2 hours following administration of 10 mg/kg caffeine base (equivalent to 20 mg/kg caffeine citrate).

Results: Time domain analyses showed a reduced correlation between MAP and TOI following caffeine administration. These reductions were observed across the cerebral oximetry index (linear mixed-model coefficient -0.093, standard error 0.04; p = 0.028) and the detrended cross-correlation coefficients (ρ_1 coefficient -0.061, standard error 0.029; p = 0.046, ρ_2 coefficient -0.12, standard error 0.04; p = 0.006 and ρ_5 coefficient -0.13, standard error 0.055; p = 0.025), and suggested an acute improvement in cerebral autoregulation. Features from detrended cross-correlation analysis exhibited greater discriminative value than other methods between pre- and post-caffeine timepoints.

Conclusion: We observed a reduced correlation between MAP and TOI from near-infrared spectroscopy. These findings suggest an acute enhanced capacity for cerebral autoregulation following a loading dose of caffeine in preterm infants, contributing to our understanding of the physiological impact of caffeine therapy.

Key notes:

- Following caffeine base administration, preterm infants exhibit falls in TOI and cerebral blood flow velocity.

- Preterm infants exhibited reduced correlation between arterial blood pressure and cerebral oxygenation, suggestive of enhanced cerebral autoregulation which may counter the direct effects of vasoconstriction

- Detrended cross-correlation analysis may be used to describe cerebral autoregulation in preterm infants

Introduction

Cerebral autoregulation is a mechanism by which cerebral blood flow is maintained relatively constant over a range of perfusion pressures. It can be thought of as the dynamic relationship between the arterial blood pressure (the system input) and cerebral blood flow (the system output). Impaired cerebral autoregulation is common in very preterm infants (1) and is considered a risk factor for brain injury. In particular, it has been proposed as a potential contributor to the development of intraventricular haemorrhage (IVH) (2).

Varying approaches have been adopted for quantifying the relationship between mean arterial pressure (MAP) and the tissue oxygenation index (TOI), a commonly-used surrogate for cerebral blood flow. These techniques range from time domain correlation (1, 3) to frequency domain coherence (4, 5) and transfer function analyses (6, 7). Some preterm infant studies have also used the pressure passivity index (1, 8) which denotes the percentage of time with high correlation or concordant changes between the two signals, though this approach requires dichotomization to 'intact' and 'impaired' cerebral autoregulation.

There has been limited work evaluating time and frequency domain features of cerebral autoregulation in preterm infants following caffeine loading. Caffeine given to preterm infants, many extremely premature to reduce apnoea of prematurity and prevent the need for intubation, are often in vunerable cardiovascular states. Many factors such as infection, mechanical ventilation and hypotension require inotrope support and may result in

significantly labile blood pressure. A study by Eriksen et al. compared cerebral oximetry index (COx), a metric of autoregulation, and frequency domain coherence, reporting poor correlation between them (9). Vesoulis et al. (6) have also applied transfer function to characterise the capacity of preterm infants to dampen fluctuations in MAP as a measure of cerebral autoregulation.

The Caffeine for Apnoea of Prematurity (CAP) trial showed that caffeine improved neurodevelopmental outcomes at a corrected 18-21 months of age (10) and reduced risk of motor impairment at 11 years follow-up (11), compared to placebo. Following a loading dose of caffeine, our group previously showed reduced cerebral perfusion using Doppler blood flow velocity and TOI (12), and increased pulse pressure variability from continuous arterial blood pressure data (13). The potential deleterious or beneficial impacts of these acute changes to cerebral autoregulation required further evaluation. The aim of this study was thus to evaluate and compare the effect of caffeine loading on cerebral autoregulation in preterm infants using a range of correlation techniques.

Patients and Methods

Data collection

Physiological data was collected as of a study approved by the Western Sydney Area Health Service Human Research and Ethics and conducted according to the World Medical Association Declaration of Helsinki. Informed parental consent was obtained in all cases. The examined cohort comprised of infants with gestational age < 34 weeks who required caffeine therapy for any of the following reasons: weaning from mechanical ventilation, reducing risk of extubation failure, and treatment of apnoea of prematurity. Infants with significant congenital anomalies and high-grade peri-intraventricular haemorrhage at time of study were excluded.

Thirty infants had concurrently available MAP and TOI data. Arterial blood pressure data were collected via an umbilical or peripheral arterial catheter, following single point

calibration to atmospheric pressure. Cerebral near-infrared spectroscopy (NIRS) data were collected simultaneously via the NIRO-300 system (Hamamatsu Photonics, Hamamatsu City, Japan), with smoothed mean values acquired at 6 Hz. NIRS offers a measure of cerebral oxygenation by emitting and receiving NIR light at specific wavelengths (775, 825, 850 and 904 nm) to determine a range of variables: oxygenated haemoglobin (HbO_2) and deoxygenated haemoglobin (HHb). TOI is defined as:

$$TOI = \frac{HbO_2}{HbO_2 + HHb}$$

MAP and TOI data were extracted from two timepoints relative to intravenous administration of 10 mg/kg of caffeine base: 20-30 minutes prior to start of dose (pre-caffeine) and 1-2 hours following dose completion (post-caffeine).

Pre-processing

Signal processing was completed in (Python Software Foundation, version 2.7. https://www.python.org/). We completed beat-to-beat extraction of MAP from the arterial blood pressure data down-sampled to 100 Hz. For running 10-minute windows shifted in 5-minute intervals with sufficient pulses (40-250 beats per minute), we extracted a range of features to describe cerebral autoregulation. The techniques applied to the aligned time-series of MAP and TOI are described in the following sections.

Coherence analysis

Magnitude-squared coherence (MSC) is defined as:

$$MSC(f) = \frac{|S_{xy}(f)|^2}{[S_{xx}(f)S_{yy}(f)]}$$

where S_{xx} is the autospectrum of changes in mean arterial pressure, S_{yy} that of changes in cerebral oxygenation and S_{xy} the cross-spectrum between the two signals. Magnitude-squared coherence approaching 1 suggests an increasing linear relationship while

coherence approaching 0 suggests a loss of this relationship or a potentially non-linear relationship between the given input (MAP) and output (TOI).

MAP and TOI signals were first down-sampled to 2 Hz (9), and mean coherence determined in the very low frequency (Coh_{VLF}; <0.003 Hz) and low frequency (Coh_{LF}; 0.003 – 0.004 Hz) ranges.

Time domain correlation analysis

The COx index, proposed initially by Brady et al. (3) is defined as the moving linear correlation coefficient between cerebral perfusion pressure and cerebral oximeter waveforms. They defined a sliding window of 5 minutes in length and shifted it in increments of 1 minute to obtain 6 correlation coefficient *r* values for a 10-minute epoch, with the mean of these being the calculated COx value for the given epoch (3).

We down-sampled the MAP and TOI signals to 0.1 Hz and determined the corresponding COx and regression coefficient (slope the fitted line) for each qualifying 10-minute window.

Cross-correlation analysis

The cross-correlation coefficient quantifies the degree of linear correlation between two signals, following adjustment for a time delay. Like COx, potential coefficient values range from -1 to 1, denoting perfectly negative and positive correlation, respectively. A study evaluating cerebral autoregulation in adult patients has applied cross-correlation to arterial blood pressure and cerebral blood flow velocity signals (14), though limited work exists in preterm infants.

From the 0.1 Hz down-sampled MAP and TOI time series, we determined the cross-correlation coefficient *r* and time delay τ (± 30 seconds which encompasses the 10 seconds or less of the time lag associated with normal autoregulation (15)).

Detrended cross-correlation analysis

Detrended cross-correlation analysis is a generalisation of the detrended fluctuation analysis method (16, 17) which has been applied widely to physiological data. The corresponding cross-correlation coefficient ρ_{DCCA} has been proposed by Zebende et al. (16) to quantify the degree of cross-correlation between nonstationary time series. Like the cross-correlation coefficient, it is bound by -1 and 1. It is calculated as follows: the two series $y_1(k)$ and $y_2(k)$ with same length N are first mean-centred and integrated to obtain $R_1(k)$ and $R_2(k)$, respectively, where $k = 1, \ldots, N$. For a given box size n, the time series are divided into N-n overlapping boxes and the covariance of the residuals is determined: $f^2_{DCCA}(n, i) = 1/(n+1) \sum_{k=i}^{i+n}(R_1(k) - \tilde{R}_{1,i}(k))(R_2(k) - \tilde{R}_{2,i}(k))$ where $\tilde{R}_{1,i}(k)$ and $\tilde{R}_{2,i}(k)$ are the local linear trend in each box beginning at i. The covariance of residuals across all boxes is averaged to determine $F^2_{DCCA}(n)$. The coefficient ρ_{DCCA} is then expressed as:

$$\rho_{DCCA} = \frac{F^2_{DCCA}(n)}{F_{DFA1}(n)F_{DFA2}(n)}$$

where $F^2_{DCCA}(n)$ is the detrended covariance function, and $F_{DFA1}(n)$ and $F_{DFA2}(n)$ are the detrened variance functions for $y_1(i)$ and $y_2(i)$, respectively. Figure 1 shows two excerpts of raw traces corresponding to a low and high absolute value of *ρ*.

To the best of our knowledge, this technique has not been applied for characterising cerebral autoregulation. We extracted detrended cross-correlation coefficient for 3 window sizes from the 0.1 Hz down-sampled MAP and TOI signals: 1 minute (ρ_1), 2 minutes (ρ_2) and 5 minutes (ρ_5).

Statistical analysis

R version 3.4 (R Core Team, 2012) and lme4 (Bates, Maechler & Bolker, 2012) were used for statistical analysis. Statistical significance for all models was defined as $p < 0.05$. To determine the mean weighted feature for a single subject, features were weighted according to the variability (SD) of the MAP time series (4, 9). Using univariate linear regression modelling, we first determined the relationship between weighted features and

the corresponding gestational age and birthweight z scores from Fenton growth charts (18). We also determined the Spearman rank correlation between the weighted features of cerebral autoregulation.

Hahn et al. found a minimum time of 1.3-3.7 hours was needed to discriminate between subjects using measures of coherence (4). Given the data available at both pre- and post-caffeine timepoints, all qualifying windows were taken into consideration via linear mixed modelling. We evaluated the effect of caffeine on the extracted features, adjusting for gestational age and birthweight z scores and including a random intercept for each subject. Statistical significance of independent variables was evaluated using likelihood ratio tests of the model with and without the variable in question (19).

Results

The cohort characteristics (n = 30) are summarised in Table 1. Infants had a mean (SD) gestational age of 27 (2.3) weeks, with birthweight 1080 (400) grams and postnatal age at evaluation of 2.6 (2.2) days.

Effect of caffeine therapy

Figure 2 provides examples of the change in weighted mean for a) a coherence feature (coh_{LF}), b) time-domain correlation analysis (COx) and c) a detrended cross-correlation feature (ρ_2). From mixed model analysis (Table 2), coherence features did not change significantly between timepoints (coh_{VLF}; p = 0.915 and coh_{LF}; p = 0.479). In contrast, caffeine base administration contributed significantly to the reduction in time-domain correlation metrics: COx (p = 0.028), ρ_1 (p = 0.046), ρ_2 (p = 0.006) and ρ_5 (p = 0.025).

Relationship between extracted features, gestational age and birthweight

Univariate linear regression modelling showed that coh_{LF} was mildly and inversely correlated with gestational age (linear coefficient -0.0073, standard error 00.0033; p = 0.037, supplementary table S1). In this dataset, COx and ρ_1 exhibited a negative linear relationship

with birthweight, independent of sex and gestational age. Examples of these relationships are also presented in supplementary figure S1.

Correlation between cerebral autoregulation features

Evaluation of the correlation between features identified significant associations between all time-domain features (Supplementary table S2). Autoregulation capacity as estimated by coherence measures at both very low and low frequencies were not strongly correlated with time domain features, nor each other. COx was strongly and linearly correlated with ρ_5, potentially due to the fact that both are based on correlations in the time domain, and to the similar running window size (5 minutes) over which the features were determined. The coefficient from cross-correlation was correlated with both COx and ρ_5, though the time delay was not correlated with any of the other metrics.

Discussion

In this study, we evaluated various time and frequency domain features to describe cerebral autoregulation in preterm infants following a loading dose of caffeine base. Caffeine is a non-specific inhibitor of adenosine receptors (20) and among the most frequency-used medication in neonatal intensive care (21). Evaluation of long-term outcomes from the CAP trial have shown improved neurodevelopmental outcomes at corrected 18-21 months of age (10), though differences were attenuated and no longer significant at 5 years follow-up (22). At 11 years follow-up however, infants in the caffeine therapy was associated with reduced risk of motor impairment (11), though little is known on its underlying mechanisms.

Our original study with the same cohort showed significant reduction in two different parameters of cerebral blood flow that being doppler cerebral blood velocity and reduced TOI (12). This finding suggested some cause for caution particularly in the most preterm unstable babies likely to exhibit impaired cerebral autoregulation. High correlation or concordance is often interpreted as an impaired cerebral autoregulation, where linear changes in arterial blood pressure correspond to similar changes in cerebral blood flow. This

has similarly been observed by Vesoulis et al. using a frequency domain approach to quantify dampening of the fluctuations in MAP and TOI signals (6). There is also evidence to suggest greater variability in cerebral oxygenation in preterm infants with a haemodynamically-significant PDA (8) and that treatment via surgical ligation may result in an acute impairment of cerebral autoregulation (23). Placing the original finding of this cohort with two parameters of reduced cerebral blood flow at 1 hour post-caffeine loading dose with these findings of potentially improved cerebral autoregulation post loading dose are reassuring that benefits accrue acutely with loading dose of caffeine in preterm newborns at risk of impaired cerebral autoregulation.

To our best knowledge, this is the first paper to examine changes in cerebral autoregulation following a loading dose of caffeine therapy in preterm infants, and the first to apply detrended cross-correlation analysis to describe these changes. We observed a reduction in time-domain correlation as characterised by COx, ρ_1, ρ_2 and ρ_5 in order of increasing sensitivity. Coherence in the LF range similarly trended towards this, though was not statistically-significant. This reduction (where changes in TOI are less correlated with those in MAP) is congruent with an improved capacity to autoregulate and may offer insight into the CAP trial findings (10, 24). The findings are in agreement with a recent study in adults (25) examining the effect of 200 mg caffeine which reported reduced cerebral blood flow with concurrent improved cerebral autoregulation, quantified by rate of regulation (26).

A possible explanation for these observations may be the caffeine-induced inhibition of the adenosine receptors: adenosine induces dilation of cerebral vessels (27), and may thus play a role in cerebral autoregulation (25, 28). The resulting cerebral vasoconstriction may then alter the dynamic autoregulation. Further work is required to understand if the observed effects are consistent in other cohorts and persist through maintenance dose administration, especially given the tolerance effect observed in adults (29).

Detrended cross-correlation analysis may offer a robust means of characterising cerebral autoregulation, particularly given its greater discriminative power in identifying

altered cross-correlation between MAP and TOI following caffeine base administration. The primary difference between detrended cross-correlation analysis and the other time domain techniques lies in the detrending which is not addressed by other techniques: it may potentially be more suited to monitoring longer-term changes in cerebral autoregulation where non-stationarities are likely to be present. This observation nevertheless requires further validation over larger monitoring periods.

The relatively poor discriminative value of the frequency-domain based features may have been due in part to the quantity of data available. Previous work using similar measurements have acknowledged the dependence of subject-specific estimations (4) on data quantity. These frequency-domain approaches are relatively insensitive to phase shifts between MAP and TOI signals. Phase shifts between signals have previously been proposed as markers of cerebral autoregulation (30) though has seldom been reported in previous studies. Despite earlier speculation that the time delay may hold autoregulatory information, this feature was not particularly discriminative, nor related to other measures of autoregulation. This may have been due in part to the sampling frequency (0.1 Hz) which may not have been sufficiently sensitive to capture these phase shifts.

Study limitations

Data quality and the concurrent availability of MAP and TOI signals limited the available data for analysis; not all subjects had arterial blood pressure lines and of those that did (n = 30), the presence of artefact necessitated the exclusion of certain pre- or post-caffeine timepoints. We cannot entirely attribute the observed effect to caffeine base administration, given the absence of a control group not receiving caffeine. The total evaluation time at each timepoint may also have been insufficient to discriminate between infants and thus, potential effects of caffeine. There may also have been other influencing factors which were not accounted for in statistical analysis: for example, the partial pressure of CO_2 (4) is an important regulator of cerebral blood flow and can thus affect the TOI signal.

Conclusion

We applied a range of time and frequency domain techniques to characterise cerebral autoregulation in preterm infants. We observed a reduction in COx and detrended cross-correlation coefficients over a range of time windows (ρ_1, ρ_2 and ρ_5), suggestive of an improved capacity of cerebral autoregulation following a loading dose of caffeine therapy. These observations help to clarify the underlying mechanisms and serve as a first step towards understanding the findings of the Caffeine for Apnoea of prematurity trial. These further observations of the caffeine cohort reported in this journal (12) help to clarify the underlying cerebrovascular physiological changes and potential mechanisms of harm and benefit. The discriminative value of the detrended cross-correlation coefficient also support its potential for cerebral autoregulation monitoring in preterm infants, an important step in understanding commonly-used treatments and risk stratification.

Funding and Conflict of interests

Table 1 Summary of cohort characteristics. Continuous variables are summarised as mean (SD) and binary variables are expressed as number (percentage, %) of the specified category.

Variable	Summary (n = 30)
Gestational age (weeks)	27 (2.3)
Birthweight (grams)	1080 (400)
Male sex	16 (53.3 %)
Postnatal age (days)	2.6 (2.2)
Continuous Positive Airway Pressure	1 (3.3 %)
Ventilated (%)	29 (96.7 %)
Respiratory distress syndrome (%)	27 (90 %)
Died (%)	1 (3.3 %)

Table 2 Linear mixed model coefficients (standard error) and statistical significance for characterising the impact of caffeine on multivariable features, adjusted for gestational age and birthweight z scores from Fenton charts. BW: birthweight, CC: cross-correlation. Statistical significance denoted by *p<0.05, **p<0.01.

	Caffeine	p	Gestational age	p	BW z scores	p
Coh_{VLF}	-0.0035 (0.033)	0.915	-0.0016 (0.0047)	0.734	-0.023 (0.011)	0.033*
Coh_{LF}	-0.0057 (0.0081)	0.479	-0.0029 (0.0021)	0.186	-0.0083 (0.0046)	0.083
COx	-0.093 (0.04)	0.028*	0.0029 (0.0089)	0.746	-0.056 (0.019)	0.007**
Reg	-0.05 (0.047)	0.287	-0.0089 (0.01)	0.398	-0.054 (0.022)	0.039*
CC r	0.017 (0.021)	0.414	-0.0044 (0.0056)	0.434	0.019 (0.012)	0.122
CC τ	0.024 (0.22)	0.913	0.035 (0.039)	0.377	0.012 (0.087)	0.889
ρ_1	-0.061 (0.029)	0.046*	0.0013 (0.0066)	0.84	-0.05 (0.014)	0.002**
ρ_2	-0.12 (0.04)	0.006**	-0.0015 (0.01)	0.885	-0.048 (0.021)	0.034*
ρ_5	-0.13 (0.055)	0.025*	0.003 (0.015)	0.845	-0.054 (0.032)	0.107

Figure legends

Figure 1 Example of traces for mean arterial pressure (MAP) and tissue oxygenation (TOI) time series corresponding to a lesser ρ (a) and greater ρ (b). Closed and open circles denote MAP and TOI time series, respectively.

Figure 2 Changes in cerebral autoregulation prior to (pre) and in the 2 hours following (post) caffeine administration, measured by a) coh_{LF}, b) COx and c) ρ_2. Features weighted according to MAP variability for each timepoint

References

1.   Soul JS, Hammer PE, Tsuji M, Saul JP, Bassan H, Limperopoulos C, et al. Fluctuating pressure-passivity is common in the cerebral circulation of sick premature infants. *Pediatric research* 2007; 61 4:467-73.

2.   Volpe JJ. Intraventricular hemorrhage in the premature infant—current concepts. Part I. *Annals of neurology* 1989; 25 1:3-11.

3.   Brady KM, Lee JK, Kibler KK, Smielewski P, Czosnyka M, Easley RB, et al. Continuous time-domain analysis of cerebrovascular autoregulation using near-infrared spectroscopy. *Stroke* 2007; 38 10:2818-25.

4.   Hahn GH, Christensen KB, Leung TS, Greisen G. Precision of coherence analysis to detect cerebral autoregulation by near-infrared spectroscopy in preterm infants. *Journal of biomedical optics* 2010; 15 3:037002--10.

5.   De Smet D, Jacobs J, Ameye L, Vanderhaegen J, Naulaers G, Lemmers P, et al. The partial coherence method for assessment of impaired cerebral autoregulation using near-infrared spectroscopy: potential and limitations. *Oxygen Transport to Tissue XXXI*: Springer, 2010.

6.   Vesoulis ZA, Liao SM, Trivedi SB, El Ters N, Mathur AM. A novel method for assessing cerebral autoregulation in preterm infants using transfer function analysis. *Pediatric research* 2015; 79 3:453-9.

7.   Zhang R, Zuckerman JH, Giller CA, Levine BD. Transfer function analysis of dynamic cerebral autoregulation in humans. *American Journal of Physiology-Heart and Circulatory Physiology* 1998; 274 1:H233-H41.

8.   Chock VY, Ramamoorthy C, Van Meurs KP. Cerebral oxygenation during different treatment strategies for a patent ductus arteriosus. *Neonatology* 2011; 100 3:233-40.

9.   Eriksen VR, Hahn GH, Greisen G. Cerebral autoregulation in the preterm newborn using near-infrared spectroscopy: a comparison of time-domain and frequency-domain analyses. *Journal of biomedical optics* 2015; 20 3:037009-.

10.   Schmidt B, Roberts RS, Davis P, Doyle LW, Barrington KJ, Ohlsson A, et al. Long-term effects of caffeine therapy for apnea of prematurity. *New England Journal of Medicine* 2007; 357 19:1893-902.

11.   Schmidt B, Roberts RS, Anderson PJ, Asztalos EV, Costantini L, Davis PG, et al. Academic performance, motor function, and behavior 11 years after neonatal caffeine citrate therapy for apnea of prematurity: an 11-year follow-up of the CAP randomized clinical trial. *JAMA pediatrics* 2017; 171 6:564-72.

12.   Tracy M, Klimek J, Hinder M, Ponnampalam G, Tracy S. Does caffeine impair cerebral oxygenation and blood flow velocity in preterm infants? *Acta Paediatrica* 2010; 99 9:1319-23.

13.   Jacqueline H, Cindy T, L MA, Murray H, B TM. Cardiovascular impact of intravenous caffeine in preterm infants. *Acta Paediatrica* 0 ja.

14.   Steinmeier R, Hofmann RP, Bauhuf C, Hübner U, Fahlbusch R. Continuous cerebral autoregulation monitoring by cross-correlation analysis. *Journal of neurotrauma* 2002; 19 10:1127-38.

15.   Elwell C, Cope M, Delpy D. An analytical method for determining cerebrovascular transit time using near infrared spectroscopy. *Oxygen Transport to Tissue XIX*: Springer, 1997.

16.   Zebende G. DCCA cross-correlation coefficient: quantifying level of cross-correlation. *Physica A: Statistical Mechanics and its Applications* 2011; 390 4:614-8.

17.   Podobnik B, Stanley HE. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physical review letters* 2008; 100 8:084102.

18.   Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC pediatrics* 2013; 13 1:59.

19.   Winter B. A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:13085499* 2013.

20.   Fredholm BB. Adenosine, adenosine receptors and the actions of caffeine. *Basic & Clinical Pharmacology & Toxicology* 1995; 76 2:93-101.

21.   Clark RH, Bloom BT, Spitzer AR, Gerstmann DR. Reported medication use in the neonatal intensive care unit: data from a large national data set. *Pediatrics* 2006; 117 6:1979-87.

22.   Schmidt B, Anderson PJ, Doyle LW, Dewey D, Grunau RE, Asztalos EV, et al. Survival without disability to age 5 years after neonatal caffeine therapy for apnea of prematurity. *Jama* 2012; 307 3:275-82.

23.   Chock VY, Ramamoorthy C, Van Meurs KP. Cerebral autoregulation in neonates with a hemodynamically significant patent ductus arteriosus. *The Journal of pediatrics* 2012; 160 6:936-42.

24.   Schmidt B, Roberts RS, Anderson PJ, Asztalos EV, Costantini L, Davis PG, et al. Academic Performance, Motor Function, and Behavior 11 Years After Neonatal Caffeine Citrate Therapy for Apnea of Prematurity: An 11-Year Follow-up of the CAP Randomized Clinical Trial. *Jama Pediatrics* 2017.

25.   Sasaki H, Hirasawa A, Washio T, Ogoh S. Acute effect of coffee drinking on dynamic cerebral autoregulation. *European journal of applied physiology* 2016; 116 5:879-84.

26.   Aaslid R, Lindegaard K-F, Sorteberg W, Nornes H. Cerebral autoregulation dynamics in humans. *Stroke* 1989; 20 1:45-52.

27.   Ngai AC, Winn HR. Effects of adenosine and its analogues on isolated intracerebral arterioles. Extraluminal and intraluminal application. *Circulation research* 1993; 73 3:448-57.

28.   Phillis J. Adenosine in the control of the cerebral circulation. *Cerebrovascular and brain metabolism reviews* 1989; 1 1:26-54.

29.   Addicott MA, Yang LL, Peiffer AM, Burnett LR, Burdette JH, Chen MY, et al. The effect of daily caffeine use on cerebral blood flow: how much caffeine can we tolerate? *Human brain mapping* 2009; 30 10:3102-14.

30.   Panerai RB. Assessment of cerebral pressure autoregulation in humans-a review of measurement methods. *Physiological measurement* 1998; 19 3:305.

Figure 1 Example of traces for mean arterial pressure (MAP) and tissue oxygenation (TOI) time series corresponding to a lesser ρ (a) and greater ρ (b). Closed and open circles denote MAP and TOI time series, respectively.
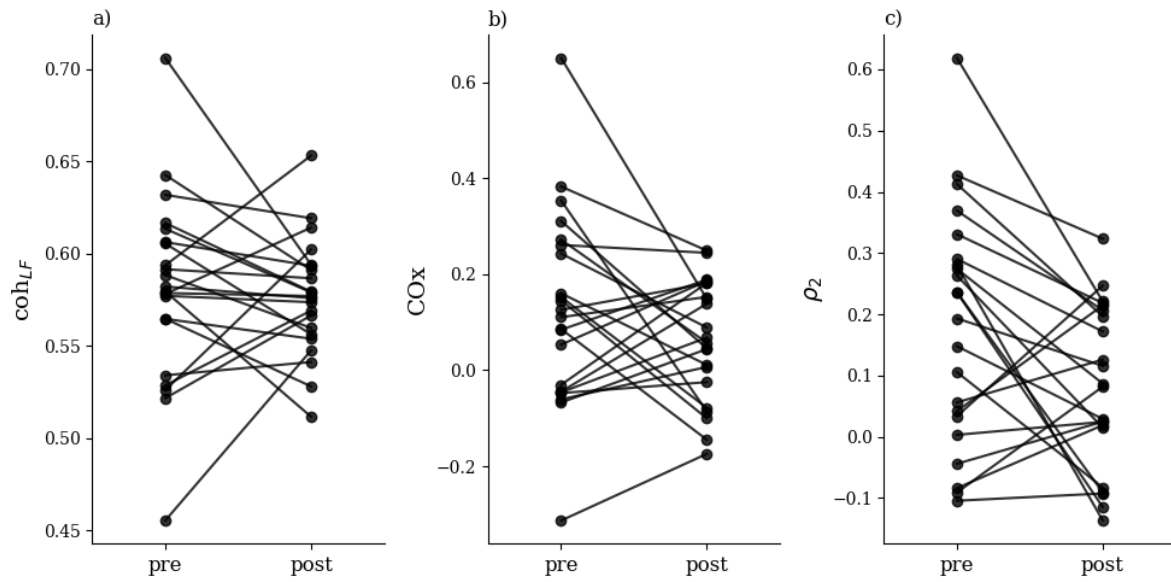
Figure 2 Changes in cerebral autoregulation prior to (pre) and in the 2 hours following (post) caffeine administration, measured by a) coh_{LF}, b) COx and c) ρ_2. Features weighted according to MAP variability for each timepoint,
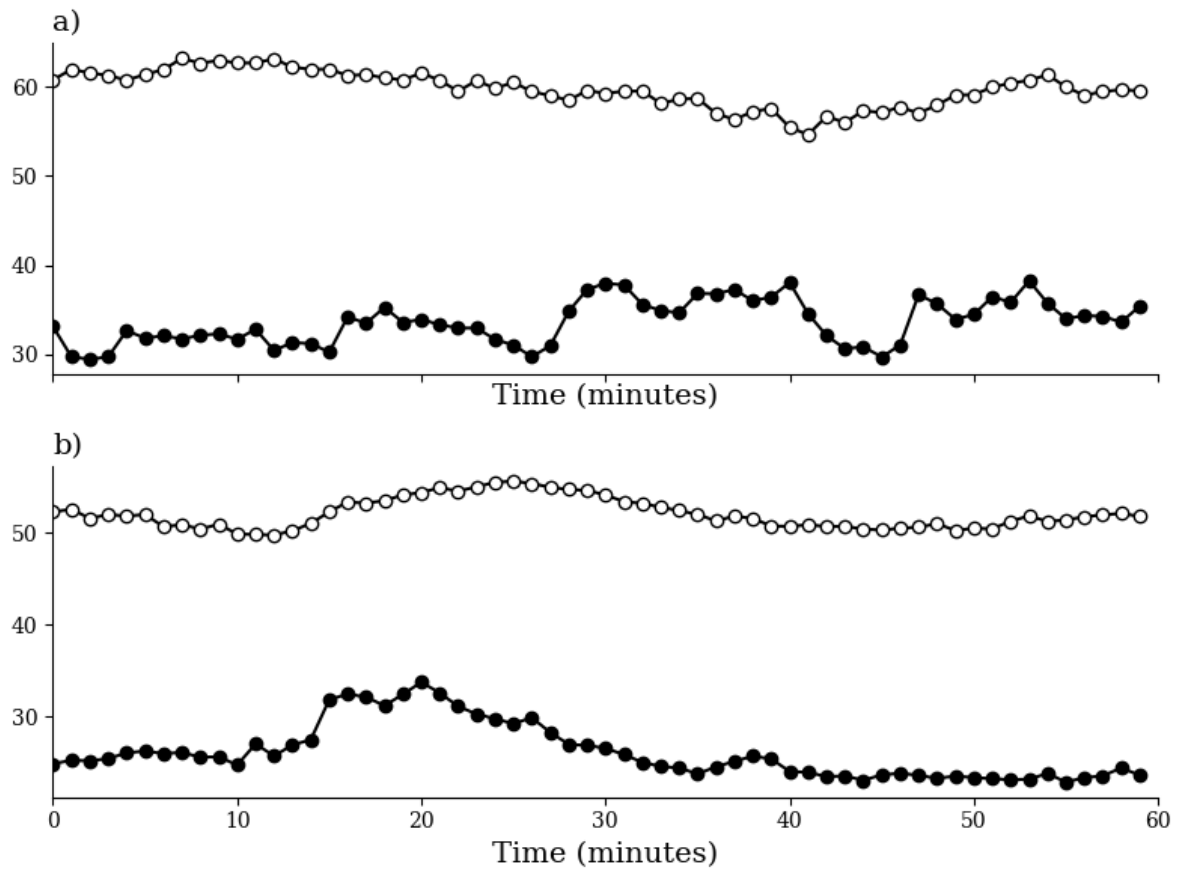
Supplementary table S1 Relationship between cerebral autoregulation features and gestational age and birthweight. Model coefficients (standard error) from univariate linear regression are presented, with statistical significance denoted by *p<0.05. BW: birthweight, CC: cross-correlation.

	Gestational age	p	BW z scores	p
coh_{VLF}	0.00081 (0.0098)	0.935	-0.023 (0.022)	0.309
coh_{LF}	-0.0073 (0.0033)	0.037*	-0.0061 (0.0082)	0.463
COx	0.011 (0.014)	0.428	-0.061 (0.03)	0.05*
Reg	-0.00052 (0.012)	0.965	-0.028 (0.027)	0.296
CC c	-0.0058 (0.007)	0.414	0.015 (0.016)	0.373
CC τ	0.028 (0.077)	0.717	-0.14 (0.18)	0.439
ρ_1	0.0065 (0.0097)	0.508	-0.049 (0.021)	0.023*
ρ_2	0.009 (0.015)	0.542	-0.048 (0.033)	0.154
ρ_5	0.022 (0.019)	0.255	-0.061 (0.044)	0.175

Supplementary table S2 Summary of Spearman correlation coefficients between extracted features. Statistical significance denoted by *p<0.05, **p<0.01, ***p<0.001. CC: cross-correlation.

	coh_{LF}	COx	Reg	CC r	CC τ	ρ_1	ρ_2	ρ_5
coh_{VLF}	0.158	0.394*	0.422*	-0.176	-0.014	0.3	0.352	0.392*
coh_{LF}		-0.034	-0.108	-0.078	-0.057	-0.069	0.005	-0.146
COx			0.745***	-0.489**	-0.069	0.802***	0.876***	0.943***
Reg				-0.254	0.041	0.630***	0.719***	0.736***
CC r					-0.01	-0.21	-0.309	-0.562**
CC τ						-0.102	-0.157	0.01
ρ_1							0.896***	0.656***
ρ_2								0.802***

Supplementary figure S1 Scatter plots of linear relationship with gestational age and birthweight for coh_LF (panels a, b), COx (c, d) and ρ_1 (e, f), respectively.
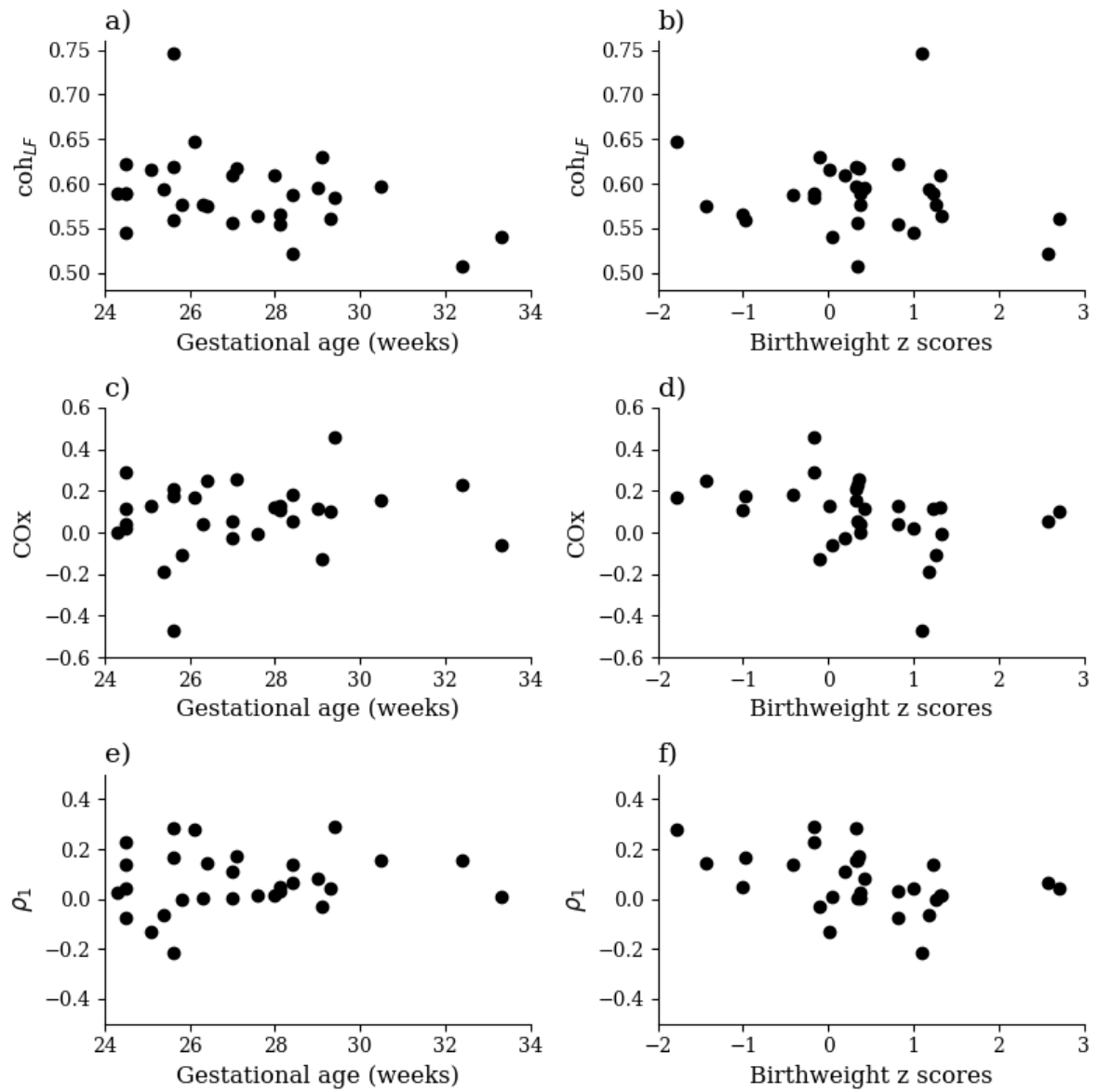
5.2   Concluding remarks

The work presented in this chapter addresses the fourth aim of the dissertation to characterise cerebral autoregulation in preterm infants following caffeine base administration. The thesis applied a range of techniques including detrended cross-correlation and time-domain correlation, finding acute improvements in cerebral autoregulation following caffeine administration. The detrended cross-correlation coefficient was also more discriminative in this context, potentially due to detrending and mitigation of non-stationarities. This work further contributes to our understanding of the CAP trial findings of improved neurocognitive outcomes [12] and later reduced risk of motor impairment [22] associated with caffeine. It also identifies the potential of detrended cross-correlation analysis, a generalisation of the DFA method, to characterise cerebral autoregulation in preterm infants.

The following chapter focuses on model development for detection of malnutrition and body composition monitoring in neonates and infants.

Chapter 6

Malnutrition detection and nutritional status assessment using near-infrared interactance

Previous advanced analysis work in the NICU has focused heavily on the detection of life-threatening outcomes and clinically-important events. In more recent years, there has been growing acknowledgement of nutrition in the management and monitoring of neonates [40]. With the complexity of conditions and procedures faced by patients of the NICU, extrauterine growth restriction remains common, particularly in those who are critically-ill [41]. Malnutrition in early life has both short- and long-term consequences [42]. For example, being born small-for-gestational age (that is, having a birthweight below the 10th percentile at a given gestational age) is associated with later risks of cardiovascular disease and type II diabetes [43, 44]. Barker et al. also find evidence that the pace and pathway of growth in early life influences morbidity risk in later life [15, 16]. Adequate nutrition is necessary for neurocognitive development [13, 14]; work by Richards et al. have demonstrated the association between birthweight and postnatal growth and cognitive development [45].

While growth itself is not a functional outcome, it offers an indication of health, disease and nutrition [40]: measuring and monitoring growth is necessary for identifying infants at risk of malnutrition. Existing approaches for monitoring nutritional status include simple anthropometric metrics such as weight or weight-for-length z scores, as well as air displacement plethysmography (ADP) which is considered one of the criterion methods of body composition assessment in the paediatric population [46–48]. ADP may not be practical in certain circumstances, given cost and portability (for example, in low-middle income countries), as well as the need to place the subject in an enclosed chamber (for example, if

on mechanical ventilation or connected to other monitoring systems).

This chapter addresses the following research questions:

7: What anthropometric features are associated with neonatal malnutrition?

8: What NIR features can be used to model body composition measures of nutritional status?

This chapter applies a range of feature selection techniques to identify potentially predictive features for nutritional assessment, and to develop models for identifying infants at risk of malnutrition and/or later morbidity.

The content presented in this chapter is currently published as:

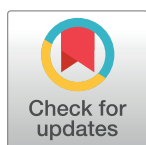- Huvanandana, J., Carberry, A.E., Turner, R.M., Bek, E.J., Raynes-Greenow, C.H., McEwan, A.L., Jeffery, H.E. 2018. An anthropometric approach to characterising neonatal morbidity and body composition, using air displacement plethysmography as a criterion method. *PLOS ONE*, 13, e0195193.

- Huvanandana, J., Jones, P., Jeffery, H.E., Carberry, A.E., Norris, S. and McEwan, A.L., 2017, December. A near-infrared interactance model for the estimation of infant body composition. In *Life Sciences Conference (LSC), 2017 IEEE* (pp. 149-152). IEEE.

6.1 An anthropometric approach to characterising neonatal morbidity and body composition, using air displacement plethysmography as a criterion method

Statement of contributions of joint authorship

- Jacqueline Huvanandana (Candidate): corresponding author, completed the analysis, writing, reviewing and editing of the manuscript

- Angela Carberry: data design and data collection supervision, reviewing and editing of the manuscript

- Robin Turner: provided technical and statistical guidance, reviewing and editing of the manuscript

- Emily Bek: contributed to data analysis and interpretation, reviewing and editing of the manuscript

- Camille Raynes-Greenow: contributed to the study design, reviewing and editing of the manuscript

- Alistair McEwan (Principal Supervisor): provided technical advice on the analysis, reviewing and editing the manuscript

- Heather Jeffery: contributed to the study design, reviewing and editing of the manuscript

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Alistair McEwan **Date:** 24 May 2018

An anthropometric approach to characterising neonatal morbidity and body composition, using air displacement plethysmography as a criterion method

Jacqueline Huvanandana[1]*, Angela E. Carberry[1], Robin M. Turner[2], Emily J. Bek[3], Camille H. Raynes-Greenow[4], Alistair L. McEwan[1], Heather E. Jeffery[1,3,4]

1 School of Electrical and Information Engineering, University of Sydney, Sydney, Australia, **2** School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia, **3** Sydney Medical School, University of Sydney, Sydney, Australia, **4** Sydney School of Public Health, University of Sydney, Sydney, Australia

* j.huvanandana@gmail.com

Abstract

Background

With the greatest burden of infant undernutrition and morbidity in low and middle income countries (LMICs), there is a need for suitable approaches to monitor infants in a simple, low-cost and effective manner. Anthropometry continues to play a major role in characterising growth and nutritional status.

Methods

We developed a range of models to aid in identifying neonates at risk of malnutrition. We first adopted a logistic regression approach to screen for a composite neonatal morbidity, low and high body fat (BF%) infants. We then developed linear regression models for the estimation of neonatal fat mass as an assessment of body composition and nutritional status.

Results

We fitted logistic regression models combining up to four anthropometric variables to predict composite morbidity and low and high BF% neonates. The greatest area under receiver-operator characteristic curves (AUC with 95% confidence intervals (CI)) for identifying composite morbidity was 0.740 (0.63, 0.85), resulting from the combination of birthweight, length, chest and mid-thigh circumferences. The AUCs (95% CI) for identifying low and high BF% were 0.827 (0.78, 0.88) and 0.834 (0.79, 0.88), respectively.

For identifying composite morbidity, BF% as measured via air displacement plethysmography showed strong predictive ability (AUC 0.786 (0.70, 0.88)), while birthweight percentiles had a lower AUC (0.695 (0.57, 0.82)). Birthweight percentiles could also identify low and high BF% neonates with AUCs of 0.792 (0.74, 0.85) and 0.834 (0.79, 0.88). We applied

94

a sex-specific approach to anthropometric estimation of neonatal fat mass, demonstrating the influence of the testing sample size on the final model performance.

Conclusions

These models display potential for further development and evaluation in LMICs to detect infants in need of further nutritional management, especially where traditional methods of risk management such as birthweight for gestational age percentiles may be variable or non-existent, or unable to detect appropriately grown, low fat newborns.

Introduction

Neonatal body composition assessment plays an important role in characterising the nutritional and dietary status of newborn infants. Those with limited body fat face risks of increased mortality and morbidity, with undernutrition linked to inhibited long-term growth and cognitive development [1, 2]. A 2010 report from the World Health Organization (WHO) attributed undernutrition as a contributing factor in one third of child deaths under five years of age [3]. The majority of these deaths occur within the first few days of life and in low and middle income countries (LMICs) [4].

Current validated methods for measuring body composition such as air displacement plethysmography (ADP), dual x-ray absorptiometry and hydrometric methods are often impractical in LMICs, given stipulations of portability, cost and operational expertise. Anthropometric measures such as mid-upper arm circumferences (MUAC), birthweight for gestational age percentiles (henceforth birthweight percentiles) and weight-for-length Z scores are commonly used in place of these more complex techniques to gauge undernutrition [5].

Simple cut-offs have been defined for MUAC to screen for moderate and severe acute malnutrition and although they have been evaluated in older infants (aged 6–60 months) with respect to risk of mortality [6], there is a lack of similar data on its reliability and association with these risks in younger infants (under 6 months). MUAC and abdominal circumference also reflect adiposity [7]. Head circumference reflects brain volume and thus intrauterine brain development [8, 9] while chest circumference has been shown to be a significant predictor of low birthweight [10, 11], commonly used to identify infants at risk from undernutrition. Though these circumferences have not been extensively evaluated in relation to malnutrition risk, their simplicity and scalability may render them suitable candidates for screening use in LMICs. In the newborn period, birthweight percentiles and less often, weight-for-length Z scores are traditionally used to identify malnutrition, though as with all anthropometry, they may be susceptible to measurement inaccuracies. Birthweight percentiles are limited by unknown or inaccurate gestational age in LMIC settings and cannot detect the appropriately grown (10-90th percentile) low fat newborn at risk of significant morbidity [12].

ADP has often been used as the reference method in infants, and has been previously validated for this population [13–15]. Carberry et al. have reported that body fat % (BF%) as measured by ADP offers a better composite measure of poor neonatal outcome than conventional birthweight measurements [12].

Anthropometric equations for the estimation of neonatal body fat have been developed against a range of reference methods. These include total body water as measured via total body electrical conductivity [16, 17], ADP [18, 19] and dual x-ray absorptiometry [20]. A recent validation of four anthropometric equations using skinfold thickness measurements

demonstrated poor explanation of variance (R-squared ranging from 0.55–0.63) between the developed equations and against ADP [21]. High inter-individual variability in the first few days of life may have contributed to the poor agreement observed and there is thus a need for caution in interpretation of the results from predictive equations.

Most models for the estimation of neonatal body fat account for sex of the infants using a single variable in the linear regression model (often 1 = male, 0 = female) [16–19]. This may not allow sufficient adjustments for sex-specific anthropometry [22] and may be biased by the predominance of either sex in the dataset used for model development.

The aim of this work was to develop anthropometric models for various applications within the first few days post-delivery. We sought to develop logistic regression models for identifying infants at risk of malnutrition, first using a composite measure of neonatal morbidity previously developed [12], while the second and third were to screen for low and high BF% neonates measured via the reference method, ADP. We also developed a linear regression model using a sex-specific approach to directly estimate neonatal fat mass (FM) using anthropometric features and thus characterise nutritional status.

Materials and methods

Data collection

Eligible neonates were term (>37 weeks), singletons born at Royal Prince Alfred Hospital, Sydney during September and October 2010. Those with major congenital abnormalities were excluded from the study. Further details of recruitment and study data collection have been previously reported [12]. Briefly, there were 782 eligible neonates born during the study period, 581 of whom were enrolled in the study (75% recruitment rate). Of these, 524 neonates had valid and complete measurements and were included for model development.

Body composition data including BF% and FM was collected via ADP (PEA POD; COSMED, Concord, USA) and anthropometric measurements were collected within 48 hours of birth. ADP applies basic gas laws to determine the body volume from that of the air displaced by the infant in an enclosed chamber, maintained at a constant temperature. Together with the weight measurement from the PEA POD scales, the density of the subject can be determined and, assuming a two-compartment model and constant density for each fat and fat-free mass, the weights of each component can be determined. BF% measurements from ADP was used as the gold standard for subsequent model development.

Anthropometric measurements (length and head, mid-upper arm, mid-thigh, abdominal and chest circumferences) were standardised using skills-based educational methods and competency confirmation [23]. Length was measured to approximately 0.1 cm heel to crown using an Easy-Glide Bearing Infantometer (Perspective Enterprises, Portage, MI). Weight on day of measurement (henceforth, weight) was measured to the nearest gram using the integrated PEA POD scales. To simulate the accuracy of standard scales in LMICs, weight was subsequently rounded to the nearest 5 grams during pre-processing. Circumferences were measured using a paper tape measure. Anthropometry and length were taken by a single researcher except for a subset of approximately 40 infants where duplicate measurements were taken [24].

Ethics. The study was approved by the Human Research Ethics Committees of Royal Prince Alfred Hospital and the University of Sydney (HREC/09/RPAH645, SSA/09/RPAH646, and University of Sydney Ref. No. 12732). Informed parental written consent was obtained, and participation was voluntary.

96

Model development and statistical analyses

Data processing and feature selection was completed in Python (Python Software Foundation, version 2.7.11 https://www.python.org/), with further statistical analysis undertaken in R 3.3.1 [25].

Neonatal morbidity screening. Composite neonatal morbidity was defined on the basis of hypothermia, poor feeding and extended length of stay, as previously described [12]. This composite measure associated with undernutrition was developed using univariate logistic regression to identify the combination of variables that could identify small-for-gestational age neonates based on birthweight percentiles [12]. We completed an exhaustive search of all possible combinations of linear, inverse and square transformations of anthropometric features. Note that for all measures, gestational age was excluded from the feature set for model development as this may be unknown or unreliable in LMICs. We also examined the greatest AUC achieved by a model excluding length as a feature and compared model performance using the Delong method [26], should length boards and appropriate training not be available.

Logistic regression models were constructed using a maximum of four original features, balancing computational efficiency and model performance. Receiver-operator characteristic (ROC) curves were generated for each feature combination, providing an indication of sensitivity and specificity in identifying the class denoting composite morbidity. The final models were selected based on those which maximised the area under the ROC curves (AUC), a measure of predictive ability.

Screening of low and high BF% neonates. A similar approach was adopted for screening of low and high BF% neonates, which were defined respectively as 1 SD below and above the mean, stratified by sex. These cut-offs were consistent with previous work finding the low BF% infants exhibited greater risk of composite neonatal morbidity [12]. Logistic regression models were developed independently using an exhaustive search for combinations of transformed features that yielded the greatest AUC for identifying low and high BF% neonates.

Estimation of neonatal fat. For the estimation of neonatal fat mass (FM) using a linear regression model, relevant features and commonly-used combinations of anthropometric measures were included in the complete set of features. We sought to characterise the underlying drivers behind weight-for-length ratio (W/L) and its higher powers, W/L^2 (body mass index) and W/L^3 (ponderal index) [27]. Inclusion of all three ratios would introduce multiple collinearity effects and thus, factor analysis was applied to determine the ratio for inclusion in the complete feature set. Weight rather than birthweight was used in these ratios given the weight loss observed in the first few postnatal days [28] and the varying ages at measurement.

To mitigate the influence of varying ranges, all continuous variables were standardised (mean = 0, SD = 1). Feature selection was then undertaken using recursive feature elimination, ranking features by their linear model coefficients, repeatedly removing them from the model and determining the optimal set of features. This was determined through minimisation of the root mean squared error score (RMSE) based on 10-fold cross-validation of the dataset. Once the set of features was determined, model fitting was completed on the sex-specific subgroups with the non-standardised features.

The performance of the sex-specific models was compared against combined sex models fitted to the determined feature set combined with a binary variable denoting sex (male = 1, female = 0). For the combined sex model, we also included sex-anthropometry interaction terms and examined the effect on the final model.

97

Model evaluation

Logistic regression models. We compared the developed logistic regression models against common anthropometric indices using the Delong method for comparison of correlated ROC curves [26]. Logistic regression models were further evaluated using leave-one-out cross-validation, with models rejected if the AUC from this was less than or equal to 0.5. This involved using all samples except one in fitting the logistic regression model and subsequently evaluating the predicted probability of the omitted sample. The process was repeated for all samples and the probabilities used to construct a ROC curve for evaluation of the leave-one-out cross validation AUC.

Linear regression models for neonatal fat mass estimation. To investigate the motivation for sex-specific model fitting, we compared anthropometric and other characteristic differences between male and female neonates. We applied a student's t-test for continuous variables such as birthweight, length and head circumference and a chi-squared test for categorical variables. We then compared the performance of sex-specific models for a range of test sample sizes by first dividing the dataset into sex-stratified halves, a training and a testing set. The male and female portions of training set were used to fit sex-specific linear estimation models while an equal-sized subset containing an even distribution of sexes was used to fit the combined sex model. We continually and randomly restricted the testing set, determining RMSE of fat-free mass, FM and BF% estimations for each sample size over 100 iterations. The overall process was repeated for 10 divisions of training and testing sets and the mean RMSE calculated for a given sample size.

Results

The characteristics of the population are summarised in Table 1, with continuous variables expressed as mean and standard deviation (SD) and categorical variables as percentages (%).

Logistic regression models

We developed logistic regression models to screen for a composite measure of neonatal morbidity as well as low and high BF% neonates.

Neonatal morbidity screening. From the exhaustive search of all possible 4-feature models to screen for neonatal morbidity, the following features were frequently included in high scoring models: weight or birthweight, chest or abdominal circumference, mid-thigh circumference and length. The greatest AUC achieved was 0.740 (0.63, 0.85) by the combination of birthweight, length, chest and mid-thigh circumferences (Fig 1). The composite feature is defined in Eq 1, where circumference is denoted by *circ*.

$$CF_{morbidity} = \frac{birthweight \times chest_{circ}}{length \times thigh_{circ}} \tag{1}$$

The greatest AUC for a model without length as a feature was 0.736 (0.62, 0.85) and combined the product of birthweight and abdominal circumferences, divided by that of head and mid-thigh circumferences. The AUC score difference between the reported model (Eq 1) and this length-free model was minimal and not significant.

We compared the ROC curves of the developed models with those of commonly used metrics. Table 2 reports the AUC, standard error and p-values from the Delong method for comparing correlated ROC curves [26]. The developed model (AUC 0.740 (0.63, 0.85)) displayed a high degree of overlap with the BF% ROC curve (AUC 0.786 (0.70, 0.88)) which exhibited the

Table 1. Comparisons of anthropometry and body composition measures for male and female neonates. An independent t-test (two-tailed) was applied to compare continuous variables and a chi-squared test for categorical variables (neonatal composite morbidity and proportions in each fat range). Statistical significance is denoted by *p<0.05, ***p<0.001.

Characteristics	Male	Female	p
n	272	252	
Birthweight (g)	3533 ± 475	3366 ± 411	<0.001***
Weight (g)	3359 ± 448	3193 ± 398	<0.001***
Length (cm)	50.4 ± 1.9	49.2 ± 1.7	<0.001***
Gestational age (weeks)	39.6 ± 1.1	39.5 ± 1.2	0.120
Age at measurement (days)	1.17 ± 0.6	1.19 ± 0.6	0.695
Mid-upper arm circumference (cm)	11.0 ± 1.0	10.9 ± 0.9	0.0897
Head circumference (cm)	35.0 ± 1.1	34.1 ± 1.1	<0.001***
Mid-thigh circumference (cm)	15.1 ± 1.3	15.0 ± 1.2	0.568
Abdominal circumference (cm)	30.6 ± 2.2	30.5 ± 2.0	0.511
Chest circumference (cm)	32.7 ± 1.8	32.4 ± 1.6	0.023*
Neonatal composite morbidity[a] (%)	3.7	3.2	0.176
Proportion low fat (%)	12.5	14.7	0.100
Proportion moderate fat (%)	71.7	71	0.199
Proportion high fat (%)	15.8	14.3	0.143
Body fat %	8.89 ± 4.0	10.09 ± 3.9	<0.001***
Fat mass (g)	310 ± 167	332 ± 155	0.119

[a]Composite neonatal morbidity defined as a composite of hypothermia, poor feeding and extended length of stay. Previously described in [12]

https://doi.org/10.1371/journal.pone.0195193.t001

next highest AUC. MUAC had a significantly poorer AUC of 0.655 (0.51, 0.80) (p = 0.046) than that of BF%.

Screening of low and high BF% neonates. The logistic regression models for low and high BF% exhibited AUCs of 0.827 (0.78, 0.88) and 0.834 (0.79, 0.88) respectively (Fig 1). The features used to construct the corresponding composite features are summarised in Eqs 2 and 3.

$$CF_{low-fat} = \frac{weight^2}{length \times chest_{circ}} \qquad (2)$$

$$CF_{high-fat} = \frac{head_{circ} \times length^2}{birthweight \times weight} \qquad (3)$$

Linear regression models

Feature selection. Factor analysis was applied to the W/L ratio and its 2 higher powers to determine which of the ratios to be included in the set for subsequent feature selection and to avoid multiple collinearity effects. Results showed that 89.5% of the variance could be explained by a single underlying factor, driven mostly by W/L^2 (R-squared > 0.99). This ratio exhibited a R-squared of 0.874 with W/L and 0.809 with W/L^3. R-squared between W/L and W/L^3 was 0.472.

Recursive feature elimination identified the combination of weight, head circumference, mid-thigh circumference and W/L^2 as the optimal set of features for neonatal FM estimation.
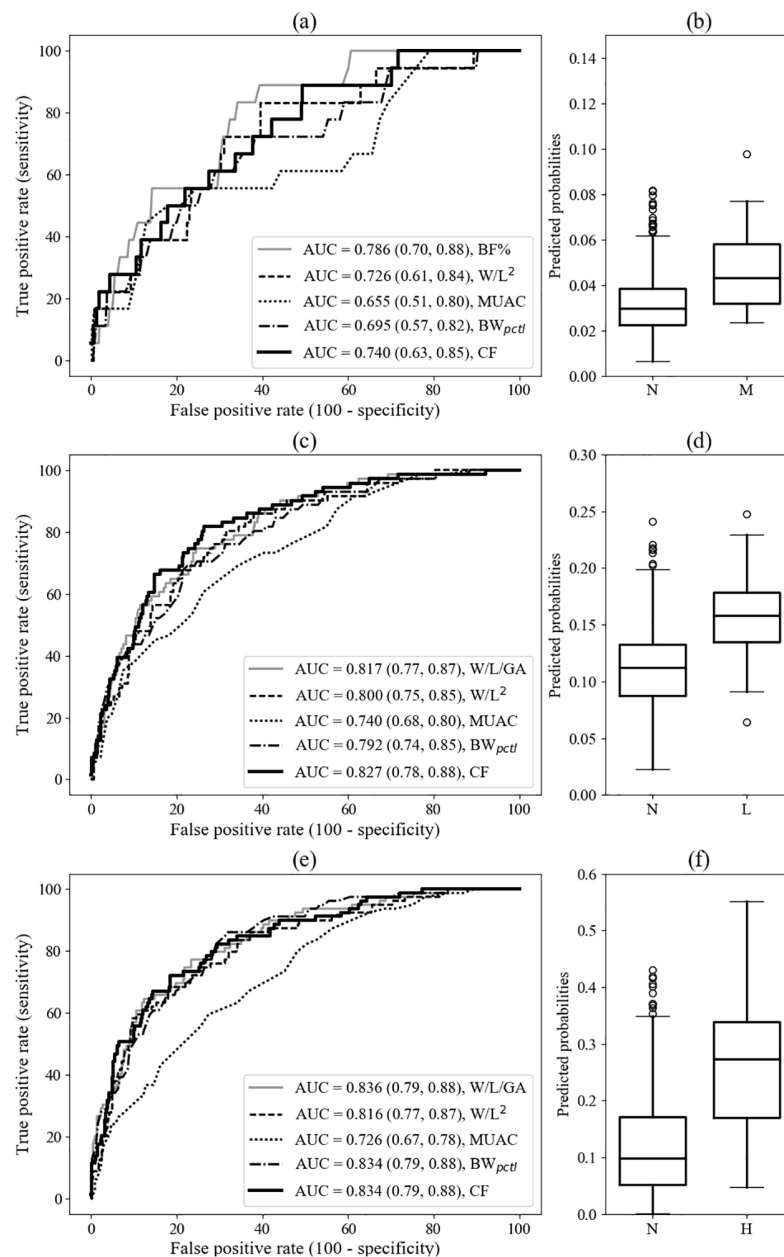
Fig 1. Receiver-operator characteristic curves and predicted probabilities for developed logistic regression models. Panels (a), (c) and (e) display the ROC curves for each of the developed (CF) models and other comparative models for the identification of composite neonatal morbidity [12], low BF% and high BF%, respectively. Comparative models include those fitted using body fat percentage (BF%), weight for length (W/L), mid-upper arm circumference (MUAC) and birthweight percentile (BW$_{pctl}$). Corresponding boxplots in (b), (d) and (f) show the predicted probabilities from the corresponding CF logistic regression models for each of the two classes: negative (N) and positive (M: composite neonatal morbidity, L: low BF% and H: high BF%).

Linear estimation of neonatal fat mass. The sex-specific and combined sex linear regression model coefficients are detailed in Table 3. Weight and W/L^2 were significant predictors of male and female neonatal FM, whereas head circumference was significant (p = 0.018) in the male population only. All three models exhibited similar R-squared statistics of approximately

100

Table 2. Comparison of receiver-operator characteristic curves for the prediction of composite neonatal morbidity, low and high fat BF% using the Delong method [26]. For each pair of logistic regression models, the standard error and p-value from the Delong method for ROC curve comparison are reported [26]. Comparisons include BF% from ADP, weight-for-length-for-gestational age (W/L/GA), weight-for-length-squared (W/L^2), mid-upper arm circumference (MUAC), birthweight percentiles (BW$_{pctl}$) and developed composite feature (CF). Statistical significance is denoted by *p<0.05, **p<0.01 ***p<0.001.

Model	AUC (95% CI)	W/L^2	MUAC	BW$_{pctl}$	CF
Composite neonatal morbidity					
BF%	0.786 (0.70, 0.88)	0.055, 0.24	0.066, 0.046*	0.062, 0.141	0.061, 0.453
W/L^2	0.726 (0.61, 0.84)		0.061, 0.239	0.047, 0.510	0.04, 0.729
MUAC	0.655 (0.51, 0.80)			0.067, 0.548	0.07, 0.227
BW$_{pctl}$	0.695 (0.57, 0.82)				0.051, 0.376
CF	0.740 (0.63, 0.85)				
Low BF%					
W/L/GA	0.817 (0.77, 0.87)	0.012, 0.174	0.028, 0.006**	0.011, 0.031*	0.014, 0.455
W/L^2	0.800 (0.75, 0.85)		0.029, 0.035*	0.021, 0.712	0.019, 0.141
MUAC	0.740 (0.68, 0.80)			0.030, 0.076	0.028, 0.002*
BW$_{pctl}$	0.792 (0.74, 0.85)				0.018, 0.056
CF	0.827 (0.78, 0.88)				
High BF%					
W/L/GA	0.836 (0.79, 0.88)	0.011, 0.06	0.026, ***	0.010, 0.832	0.047, 0.961
W/L^2	0.816 (0.77, 0.87)		0.028, 0.001**	0.177, 0.309	0.049, 0.715
MUAC	0.726 (0.67, 0.78)			0.026, ***	0.047, 0.02*
BW$_{pctl}$	0.834 (0.79, 0.88)				0.044, 0.998
CF	0.834 (0.79, 0.88)				

https://doi.org/10.1371/journal.pone.0195193.t002

0.59. For the combined sex model, all variables including sex were significant predictors of neonatal fat mass. There were no significant interactions between sex and anthropometric features (p > 0.3).

Model evaluation. To characterise model performance, we evaluated RMSE and R-squared statistics for both sex-specific and combined sex models for a range of test sample

Table 3. Linear regression model coefficients for estimation of neonatal fat mass in grams.

Variable	Intercept	Weight (g)	circ$_{head}$ (cm)	circ$_{thigh}$ (cm)	W/L^2 (g/cm^2)	Sex
Male						
Coefficient	-309.54	0.226	-19.93	15.74	243.53	-
SE	261.74	0.035	8.35	8.42	105.11	-
p	0.238	<0.001***	0.018*	0.063	0.021*	-
Female						
Coefficient	-677.90	0.190	-8.280	11.47	390.141	-
SE	270.16	0.038	8.705	7.95	105.060	-
p	0.013*	<0.001***	0.342	0.150	<0.001***	-
Combined sex						
Coefficient	-445.45	0.212	-14.857	13.191	312.795	-47.21
SE	186.35	0.0255	6.008	5.780	74.00	10.14
p	0.017*	<0.001***	0.014*	0.023*	<0.001***	<0.001***

R-squared statistics for the male, female and combined sex regression models were 0.589, 0.591 and 0.590, respectively. Units for each variable are shown in parentheses, with coefficients, standard error (SE) and p value from linear regression model fitting shown for the sex-specific and combined sex models. The variable denoting sex comprises 1 = male and 0 = female. Statistical significance is denoted by
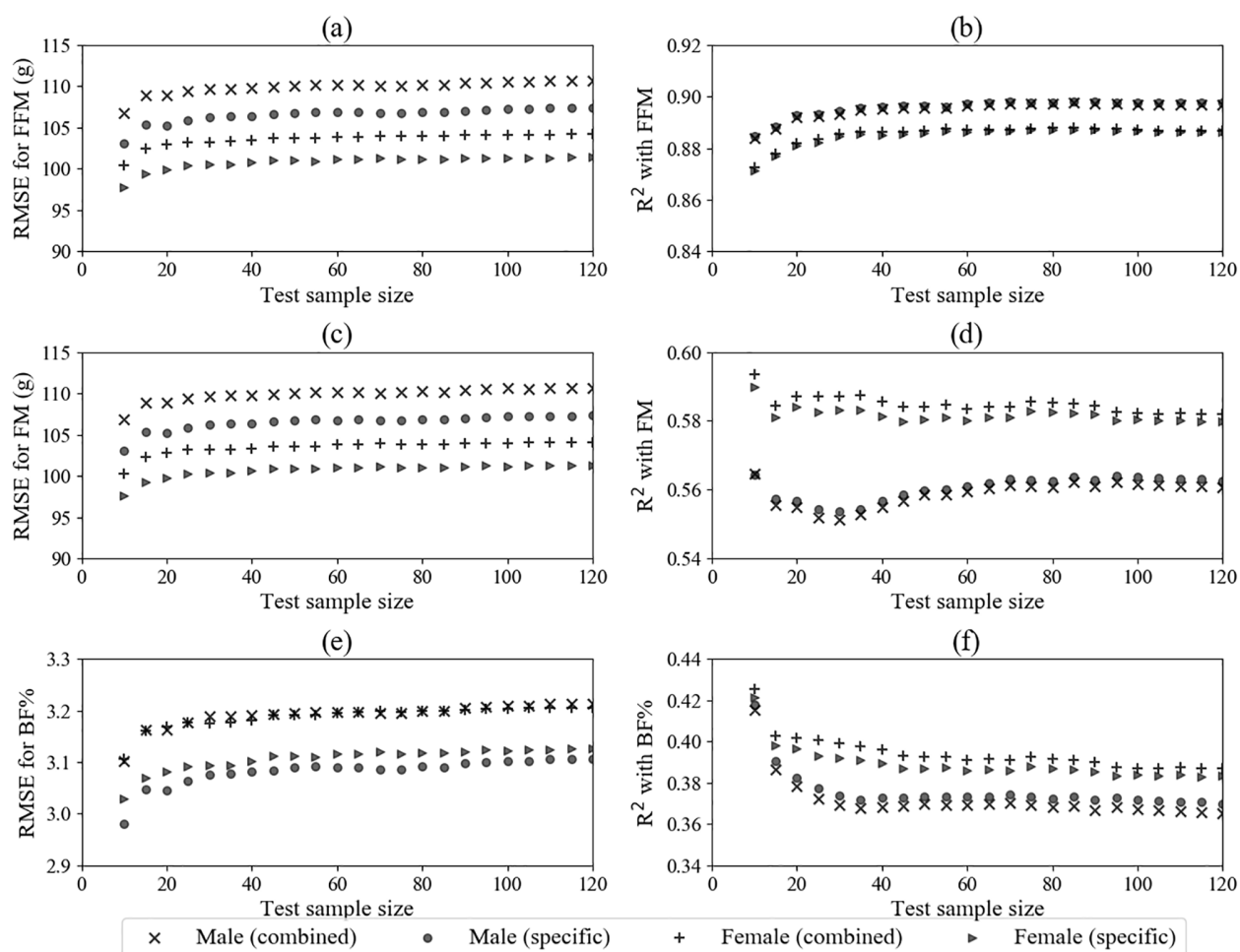
*p<0.05,

***p<0.001.

https://doi.org/10.1371/journal.pone.0195193.t003

101

Fig 2. Mean model RMSE and R-squared statistics for estimations of body composition parameters at a testing sample size. Panels (a)-(b) fat free mass (FFM), (c)-(d) fat mass (FM) and (e)-(f) body fat percentage (BF%) measured via air displacement plethysmography. Population was divided into two sex-stratified halves, with the first half used to fit male and female-specific linear estimation models and an equally-sized subset containing an even distribution of sexes used to fit the combined sex model. The second half or test set was then randomly and repeatedly restricted with root mean squared error (RMSE) and R-squared determined for each iteration.

https://doi.org/10.1371/journal.pone.0195193.g002

sizes, displayed in Fig 2. The R-squared statistics for combined sex and sex-specific models were similar, as also reflected by model fit to the complete dataset in Table 3 (R-squared combined: 0.590, male: 0.589, female: 0.591), though sex-specific models tended to exhibit a lesser RMSE in the estimation of FM, FFM and BF%.

Discussion

Summary of findings

In this study, we developed and evaluated a range of models for characterising neonatal nutritional status. Using a composite neonatal morbidity, we developed a model to detect undernourished newborns which exhibited an AUC of 0.740 (0.63, 0.85). We also examined the greatest AUC achieved by a model excluding length as a feature and found that a combination of birthweight, abdominal, head and mid-thigh circumferences yielded a AUC of 0.736 (0.62, 0.85).

102

Models for identifying low and high BF% neonates exhibited AUCs of 0.827 (0.78, 0.88), and 0.834 (0.79, 0.88), respectively. This suggests potential for application in LMICs, offering a low-cost and scalable approach for screening at birth, though this may depend on measurement accuracy and reproducibility, availability of appropriate equipment, training and evaluation of competency [24]. These factors considered, the models could nevertheless motivate the routine collection of anthropometric measurements, especially considering the socio-economic transition that many LMICs are undergoing, with both under and overnutrition present at birth.

Neonatal morbidity screening

The combination of birthweight, length, chest and mid-thigh circumferences exhibited the greatest AUC of 0.740 (0.63, 0.85) to identify neonatal morbidity. AUC from leave-one-out cross-validation was 0.698. It was interesting to note the presence of the birthweight-to-length ratio, possibly corrected for chest and mid-thigh circumferences as a potential marker for composite neonatal morbidity. Neither of these circumferences have been routinely used as a marker of adiposity, though chest circumference has been identified as a strong predictor of low birthweight [10] and there have been similar correlations reported between mid-thigh circumference, antenatal nutrition [29] and birthweight [30].

The model excluding length as a feature exhibited a AUC of 0.736 (0.62, 0.85) and did not exhibit a significantly poorer performance than that where length was included. Given the more expensive and bulkier nature of length boards compared with paper tape measures for circumference measurements, this length-free composite measure may be preferred for use in LMICs.

MUAC is a simple and fast measurement widely-used to detect undernutrition in LMICs. Though MUAC in infants under 6 months may have predictive value for infant death [31], the difference between the ROC curves for MUAC and BF% would suggest that prediction of morbidity as defined by our composite measure in this population may be improved by accurate measurement of BF% if available, or by using our anthropometric model, subject to further evaluation and validation in an independent dataset. Our model exhibited an AUC of 0.740 (0.63, 0.85) which was the next highest to BF%, among other comparisons including W/L^2 (0.726 (0.61, 0.84)), birthweight percentiles (0.695 (0.57, 0.82)) and MUAC (0.655 (0.51, 0.80)) (Fig 1, Table 2).

Screening of low and high BF% neonates

Both models yielding the greatest AUC for screening low and high BF% contained a ratio between weight and length in some form, with the low BF% neonates consisting of weight, length and chest circumference and the high BF% containing birthweight, weight, length and head circumference.

The developed model for screening low BF% neonates exhibited a AUC of 0.827 (0.78, 0.88), greater than that of the W/L for gestational age model, though the difference was not statistically-significant. In contrast, the W/L for gestational age exhibited the greatest AUC of 0.836 (0.79, 0.88) for identifying high BF% neonates, consistent with previous reports of increasing BF% with increasing gestational age [32], due to rapid fat gain late in gestation [33]. This is followed by both the composite feature and the birthweight percentile models with AUCs of 0.834 (0.79, 0.88), suggesting that for this cohort, accounting for multiple anthropometric measures performs no better than considering the birthweight percentile alone, despite the latter being considered a limited predictor of morbidity and mortality. Such percentiles are nevertheless problematic as gestational age is frequently unreliable in LMICs.

103

Linear estimation of neonatal fat

Anthropometric models for the estimation of neonatal fat mass developed by Catalano et al. [16], Schmelzle et al. [20] and Deierlein et al. [18] exhibit an R-squared of 0.84, 0.94 and 0.81, respectively. A direct comparison between our models and those developed previously is difficult given the lack of corresponding skinfold thickness measurements, differences in the criterion method for FM estimation and demographic variations. Our developed models exhibited an R-squared of 0.59 for both male and female populations, accounting for a lower variance in FM than previous reported models. This highlights the important contribution and correlation of skinfold measurements to FM estimation, though without considerable practice, these measures may have poor reproducibility amongst multiple users [34].

Comparisons of sex-specific anthropometry revealed that males tended to be longer and heavier than females, with larger head and chest circumferences on day of measurement (Table 1). The male neonates in our cohort also had lower BF%, which also remains consistent with previous studies [32]. In the anthropometric models previously developed for estimation of body composition [16–20], sex is often not included or adjusted for using a binary variable in the linear regression model. This adjustment for sex differences in anthropometry and body composition may well be inadequate, especially for robust model development. Though variability in body composition measures were similar across both sex-specific and combined sex models, we observed a greater estimation error for the combined sex models (Fig 2) and a different combination of variables that were significantly predictive of fat mass between male and female-specific models (Table 3).

The population-specific nature of body composition also extends beyond sex; there are also differences between infants of different ethnicities [35, 36], that may be genetic, biological, environmental or composites of these. Body composition is also influenced by perinatal characteristics, infant feeding methods [37] and days after birth, with infants undergoing an initial weight loss particularly during the first four postnatal days [28]. In this period, energy intake is limited until breastfeeding is established by day 5 when fat stores are no longer needed as alternative energy stores and energy is expended to the requirements of extrauterine life including thermoregulation, fluid balance and respiration [38, 39]. These factors contribute to high variability observed in the neonatal period and thus need to be considered in robust model development and application.

The changes in RMSE and R-squared with testing set size in Fig 2 demonstrate the influences of both training and testing sets on the performance and robustness of the final model. These results support sex-specific model fitting, highlighting the potentially inflated error from combining both male and female subjects in the same model with a single variable adjusting for sex. It also aids in establishing a context for model evaluation, with different target variables (fat-free mass, FM and BF%) tending towards different degrees of correlation with anthropometric variables. An understanding of model fit and robustness extends beyond R-squared which characterise the relationship between two variables, rather than agreement or differences between them. Despite similar R-squared statistics across both combined sex and specific model estimations of neonatal body composition, the RMSE tended to be lower for sex-specific models.

Strengths and limitations

The strengths of this study include the large sample size of neonates and the use of ADP as the criterion method which has been specifically validated in this population [13–15]. The model development for detecting neonatal morbidity was limited by the low representation (3.4%) of neonatal morbidity [12]. Although we approached this by using leave-one-out cross-validation

for model evaluation, there is a need for further evaluation on a population with higher neonatal morbidity. The optimal measurement of morbidity may change in a different population, as the composite score used for this analysis was based on a logistic regression analysis of potentially significant factors in the given population [12]. This dataset would ideally be sourced from LMICs where demographic characteristics align with the intended area of application for the model.

Due to the accurate measurement of weight using the integrated PEA POD scale, further validation of accurately measuring weight for the models may be required. Anthropometry including length and circumference measurements were also not obtained in duplicate except for a small subset of approximately 40 infants [24].

Body composition estimation in the neonatal population is especially difficult given the varying criterion methods, the growing evidence to suggest poor agreement between gold standards in this population [40] and the variation in body composition during the first few days of life. The model development and potentially, predictive ability, may be improved by adjusting for additional features such as skinfold measurements, which were not collected in this dataset. Evaluation in an independent population to gauge estimation error and robustness of predictive ability for neonatal morbidity, low and high BF% infants is also needed.

Conclusions

The greatest burden of neonatal and infant undernutrition and morbidity lies in LMICs, where there is an urgent need for suitable, simple approaches to monitor and manage infants. Using combinations of anthropometric features, we fitted models for application in these settings that could detect composite morbidity with an AUC of 0.740 (0.63, 0.85) in neonates in the first few days of life. Composite features involving simple, accurate, easily-measured anthropometric features could also identify low BF% infants with an AUC of 0.827 (0.78, 0.88). These models have demonstrated potential for further development and evaluation in LMICs for identifying infants in need of further nutritional management.

Supporting information

S1 Dataset. Neonatal anthropometric data.
(CSV)

Acknowledgments

Author Contributions

Conceptualization: Camille H. Raynes-Greenow, Heather E. Jeffery.

Data curation: Angela E. Carberry, Heather E. Jeffery.

Formal analysis: Jacqueline Huvanandana, Robin M. Turner, Emily J. Bek, Alistair L. McEwan.

Investigation: Jacqueline Huvanandana.

Methodology: Jacqueline Huvanandana, Robin M. Turner.

105

Project administration: Angela E. Carberry.

Supervision: Robin M. Turner, Camille H. Raynes-Greenow, Alistair L. McEwan, Heather E. Jeffery.

Writing – original draft: Jacqueline Huvanandana.

Writing – review & editing: Jacqueline Huvanandana, Angela E. Carberry, Robin M. Turner, Emily J. Bek, Camille H. Raynes-Greenow, Alistair L. McEwan, Heather E. Jeffery.

References

1. Levitsky DA, Strupp BJ. Malnutrition and the brain: changing concepts, changing concerns. The Journal of Nutrition. 1995; 125(8):2212S.

2. Prado EL, Dewey KG. Nutrition and brain development in early life. Nutrition Reviews. 2014; 72(4):267–84. https://doi.org/10.1111/nure.12102 PMID: 24684384

3. UNICEF. Facts for life: UNICEF; 2010.

4. WHO. World Health Statistics 2010: World Health Organization; 2010.

5. WHO. WHO child growth standards and the identification of severe acute malnutrition in infants and children: a Joint Statement by the World Health Organization and the United Nations Children's Fund. Geneva: World Health Organization. 2009.

6. Pelletier DL, Frongillo EA. Changes in child survival are strongly associated with changes in malnutrition in developing countries. The Journal of Nutrition. 2003; 133(1):107–19. PMID: 12514277

7. Yajnik C, Fall C, Coyaji K, Hirve S, Rao S, Barker D, et al. Neonatal anthropometry: the thin–fat Indian baby. The Pune maternal nutrition study. International Journal of Obesity. 2003; 27(2):173–80. https://doi.org/10.1038/sj.ijo.802219 PMID: 12586996

8. Cooke R, Lucas A, Yudkin P, Pryse-Davies J. Head circumference as an index of brain weight in the fetus and newborn. Early human development. 1977; 1(2):145–9. PMID: 617306

9. O'Connell EJ, Feldt RH, Stickler GB. Head circumference, mental retardation, and growth failure. Pediatrics. 1965; 36(1):62–6.

10. Thi HN, Khanh DKT, Thu HLT, Thomas EG, Lee KJ, Russell FM. Foot length, chest circumference, and mid upper arm circumference are good predictors of low birth weight and prematurity in ethnic minority newborns in Vietnam: A hospital-based observational study. PloS one. 2015; 10(11):e0142420. https://doi.org/10.1371/journal.pone.0142420 PMID: 26555356

11. Goto E. Meta-analysis: identification of low birthweight by other anthropometric measurements at birth in developing countries. Journal of epidemiology. 2011; 21(5):354–62. https://doi.org/10.2188/jea.JE20100182 PMID: 21768738

12. Carberry AE, Raynes-Greenow CH, Turner RM, Askie LM, Jeffery HE. Is body fat percentage a better measure of undernutrition in newborns than birth weight percentiles? Pediatric Research. 2013; 74 (6):730–6. https://doi.org/10.1038/pr.2013.156 PMID: 24002331

13. Ellis KJ, editor Evaluation of body composition in neonates and infants. Seminars in Fetal and Neonatal Medicine; 2007: Elsevier.

14. Ellis KJ, Yao M, Shypailo RJ, Urlando A, Wong WW, Heird WC. Body-composition assessment in infancy: air-displacement plethysmography compared with a reference 4-compartment model. The American Journal of Clinical Nutrition. 2007; 85(1):90–5. PMID: 17209182

15. Ma G, Yao M, Liu Y, Lin A, Zou H, Urlando A, et al. Validation of a new pediatric air-displacement plethysmograph for assessing body composition in infants. The American Journal of Clinical Nutrition. 2004; 79(4):653–60. PMID: 15051611

16. Catalano PM, Thomas AJ, Avallone DA, Amini SB. Anthropometric estimation of neonatal body composition. American Journal of Obstetrics and Gynecology. 1995; 173(4):1176–81. PMID: 7485315

17. Lingwood BE, van Leeuwen A-MS, Carberry AE, Fitzgerald EC, Callaway LK, Colditz PB, et al. Prediction of fat-free mass and percentage of body fat in neonates using bioelectrical impedance analysis and anthropometric measures: validation against the PEA POD. British Journal of Nutrition. 2012; 107 (10):1545–52. https://doi.org/10.1017/S0007114511004624 PMID: 21917194

18. Deierlein AL, Thornton J, Hull H, Paley C, Gallagher D. An anthropometric model to estimate neonatal fat mass using air displacement plethysmography. Nutrition & Metabolism. 2012; 9(1):1.

106

19. Aris I, Soh S, Tint M, Liang S, Chinnadurai A, Saw S, et al. Body fat in Singaporean infants: development of body fat prediction equations in Asian newborns. European Journal of Clinical Nutrition. 2013; 67(9):922–7. https://doi.org/10.1038/ejcn.2013.69 PMID: 23549200

20. Schmelzle HR, Fusch C. Body fat in neonates and young infants: validation of skinfold thickness versus dual-energy X-ray absorptiometry. The American Journal of Clinical Nutrition. 2002; 76(5):1096–100. PMID: 12399284

21. Cauble JS, Dewi M, Hull HR. Validity of anthropometric equations to estimate infant fat mass at birth and in early infancy. BMC Pediatrics. 2017; 17(1):88. https://doi.org/10.1186/s12887-017-0844-6 PMID: 28347278

22. Rodríguez G, Samper MP, Ventura P, Moreno LA, Olivares JL, Pérez-González JM. Gender differences in newborn subcutaneous fat distribution. European Journal of Pediatrics. 2004; 163(8):457–61. https://doi.org/10.1007/s00431-004-1468-z PMID: 15168110

23. Jeffery H, Henderson-Smart D, Hill D. Competency-based learning in neonatology. Medical Education. 1996; 30(6):440–4. PMID: 9217907

24. Wood AJ, Raynes-Greenow CH, Carberry AE, Jeffery HE. Neonatal length inaccuracies in clinical practice and related percentile discrepancies detected by a simple length-board. Journal of Paediatrics and Child Health. 2013; 49(3):199–203. https://doi.org/10.1111/jpc.12119 PMID: 23432733

25. R Core Team. A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837–45. PMID: 3203132

27. Koo WW, Walters JC, Hockman EM. Body composition in neonates: relationship between measured and derived anthropometry with dual-energy X-ray absorptiometry measurements. Pediatric research. 2004; 56(5):694–700. https://doi.org/10.1203/01.PDR.0000142587.59238.BD PMID: 15371563

28. Roggero P, Giannì ML, Orsi A, Piemontese P, Amato O, Moioli C, et al. Neonatal period: body composition changes in breast-fed full-term newborns. Neonatology. 2010; 97(2):139–43. https://doi.org/10.1159/000239767 PMID: 19776647

29. Donnelly JM, Walsh JM, Byrne J, Molloy E, McAuliffe F. Impact of maternal diet on neonatal anthropometry: a randomized controlled trial. Pediatric Obesity. 2015; 10(1):52–6. https://doi.org/10.1111/j.2047-6310.2013.00216.x PMID: 24443392

30. Lawoyin T. Validation and use of a simple device to identify low birth weight babies at birth. African Journal of Medicine and Medical Sciences. 1997; 27(3–4):143–5.

31. Mwangome MK, Fegan G, Fulford T, Prentice AM, Berkley JA. Mid-upper arm circumference at age of routine infant vaccination to identify infants at elevated risk of death: a retrospective cohort study in the Gambia. Bulletin of the World Health Organization. 2012; 90(12):887–94. https://doi.org/10.2471/BLT.12.109009 PMID: 23284194

32. Hawkes CP, Hourihane JOB, Kenny LC, Irvine AD, Kiely M, Murray DM. Gender-and gestational age–specific body fat percentage at birth. Pediatrics. 2011; 128(3):e645–e51. https://doi.org/10.1542/peds.2010-3856 PMID: 21824882

33. Flynn M, Goldberg G, Prentice A, Cole T. Aetiology of obesity III: critical periods for the development of obesity. Obesity: The report of the British Nutrition Foundation Task Force. 1999:45–60.

34. West J, Manchester B, Wright J, Lawlor DA, Waiblinger D. Reliability of routine clinical measurements of neonatal circumferences and research measurements of neonatal skinfold thicknesses: findings from the Born in Bradford study. Paediatric and perinatal epidemiology. 2011; 25(2):164–71. https://doi.org/10.1111/j.1365-3016.2010.01181.x PMID: 21281329

35. Singh KA, Huston-Presley LP. Birth weight and body composition of neonates born to Caucasian compared with African-American mothers. Obstetrics and Gynecology. 2010; 115(5):998. PMID: 20410774

36. Paley C, Hull H, Ji Y, Toro-Ramos T, Thornton J, Bauer J, et al. Body fat differences by self-reported race/ethnicity in healthy term newborns. Pediatric obesity. 2015.

37. Sauder KA, Kaar JL, Starling AP, Ringham BM, Glueck DH, Dabelea D. Predictors of Infant Body Composition at 5 Months of Age: The Healthy Start Study. The Journal of Pediatrics. 2017; 183:94–9. e1. PMID: 28161200

38. Regnault N, Botton J, Blanc L, Hankard R, Forhan A, Goua V, et al. Determinants of neonatal weight loss in term-infants: specific association with pre-pregnancy maternal body mass index and infant feeding mode. Archives of Disease in Childhood—Fetal and Neonatal Edition. 2011;(96):F217–F22.

39. Fonseca M, Severo M, Santos A. A new approach to estimating weight change and its reference intervals during the first 96 hours of life. Acta Paediatrica. 2015; 104(10):1028–34. https://doi.org/10.1111/apa.12894 PMID: 25488548

107

40. Wrottesley S, Pisa P, Micklesfield L, Pettifor J, Norris S. A comparison of body composition estimates using dual-energy X-ray absorptiometry and air-displacement plethysmography in South African neonates. European Journal of Clinical Nutrition. 2016; 70(11):1254–8. https://doi.org/10.1038/ejcn.2016.91 PMID: 27245207

108

6.2 A near-infrared interactance model for the estimation of infant body composition

Statement of contributions of joint authorship

- Jacqueline Huvanandana (Candidate): corresponding author, completed the analysis, writing, reviewing and editing of the manuscript

- Peter Jones: development of hardware and data collection software, training, reviewing and editing of the manuscript

- Heather Jeffery: contributed to the study design, reviewing and editing of the manuscript

- Angela Carberry: contributed to the study design, reviewing and editing of the manuscript

- Shane Norris: supervised data collection, reviewing and editing of the manuscript

- Alistair McEwan (Principal Supervisor): contributed to the study design, provided technical advice on the analysis, reviewing and editing the manuscript

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Alistair McEwan   **Date:** 24 May 2018

A near-infrared interactance model for the estimation of infant body composition

Jacqueline Huvanandana[1], Peter Jones[1], Heather E Jeffery[1,2], Angela E Carberry[1,2], Shane Norris[3]
Alistair L McEwan[1]

Abstract— In low and middle income settings, there is a pressing need for simple, low-cost and robust ways of determining infant body composition. Measuring and monitoring body composition plays an important role in addressing the burden of child deaths under five, where depleted fat stores may put infants at risk of mortality, morbidity in later life and delayed cognitive development. In a subset of 41 infants aged between 2 and 8 months, we develop and evaluate a potential model for using near-infrared interactance (NIR) features for the estimation of infant fat mass and the deuterium dilution method as the reference method. Model evaluation demonstrates the potential of NIR to capture variability in body composition, with a correlation of R = 0.90 and Bland-Altman agreement (mean bias with 95% confidence intervals) of -8 (-553, 536) g with fat mass estimated by a four-compartment (4C) model. With continued development in the larger dataset, this model may enable simple and low-cast body composition assessment of infants in low and middle incomes settings where the burden of infant deaths is greatest.

I. INTRODUCTION

A 2010 UNICEF/WHO report identified undernutrition as a contributing cause in over one third of child deaths under five [1]. The vast majority of these deaths occur in low and middle income countries [2] where access to criterion methods such as air-displacement plethysmography (ADP), dual x-ray absorptiometry and deuterium dilution are not practical given cost, portability, possible risks and operational expertise. Simple and scalable measures such as anthropometry may be more suited to these contexts, though they have their own limitations: mid-upper arm circumferences are widely used for screening moderate and severe acute malnutrition cases, though there is limited data on its reliability and predictive value for mortality in infants under 6 months. Weight-for-length or body mass index Z scores may be susceptible to inaccuracies in the measurement of length [3]. Skinfold thickness measurements may play an important role in anthropometric estimations of body fat, though its reproducibility may be limited by training and between multiple users [4].

The motivation underpinning the use of near-infrared (NIR) interactance in body composition assessment stems from early work by Conway et al. [5] who reported observed differences in the NIR spectra from samples of pork fat and two subjects of different body composition. They reported peaks in the spectrum corresponding to pure fat and water at wavelengths of 930nm and 970nm, respectively. More recent work in this area suggested the potential predictive value of optical density ratios in detecting changes in fat thickness using NIR [6], with a pilot study using NIR to estimate neonatal body fat measured from air-displacement plethysmography reporting a model fit of R > 0.8 in a cohort of 26 neonates [7]. This suggests that the variance in body fat percentage (BF%) can be explained in part by changes in the NIR reflection profiles, though validation of this model in an independent set of infants was not undertaken.

There is increasing acknowledgment that body composition and nutritional status has important implications for mortality, long-term growth, neurocognitive development [8][9] and morbidity in later life [10][11]. This is further motivated by the findings that BF% has displayed a greater predictive value than birthweight percentiles for identifying composite morbidity in a neonatal population [12].

In this study, we developed a model using NIR features for the estimation of infant fat mass using deuterium dilution as a reference. We evaluated the model agreement and correlation with fat mass and BF% as measured by a four-compartment (4C) model. We also compared model performance to that of weight-for-length-squared or body mass index (BMI) and the Slaughter equation for infant body composition assessment using sum of skinfold thickness measurements [13]. The former is often used as a proxy for adiposity in infants, with the use BMI Z-scores for monitoring infant growth. The latter equation, though originally developed in infants and youth (8-29 years) [13], has been validate in a number of populations, including infants from birth to four months [14] [15].

II. METHODS

A. Ethics

This study was approved by the Ethics Committee at the University of Sydney (USyd) and the University of Witwatersrand (Wits) (USyd HREC number: Project No.: 2015/595; Wits HREC number: M150774). The study has also been registered on the Australia and New Zealand Clinical trials Registry (ANZCTR) number: ACTRN12615001318572. The infants and children enrolled were recruited from the Soweto, South Africa pregnancy and follow up clinics over a period of 9 months (April to December 2016).

B. Data collection

Data collected included weight, length, body composition data using deuterium dilution, dual x-ray absorptiometry (DXA) and NIR scans. Weight was measured using electronic scales (Seca 376, Hamburg, Germany). Length was measured using an infantometer (Harpenden, Holtain Model 702) which has a fixed headboard and moveable footboard following a two-trained person technique. Each infant was administered a dose of deuterium, a stable non-radioactive isotope of hydrogen and saliva samples were collected and analysed in duplicate at 1.5 and 3 hours following dose administration. This allowed for total body water to be estimated and used to determine fat free mass (FFM), fat mass ($FM = mass - FFM$) and thus, fat mass percentage ($BF\% = 100 \cdot FM/mass$). Bone mineral content was measured via DXA (Hologic DiscoveryA DXA S/N 86254 APEX software version 4.0.2, Hologic Inc., Waltham, MA, USA). Skinfold measurements were conducted in duplicate using handheld callipers at four anatomical locations; triceps, subscapular, mid-thigh and flank.

All NIR measurements were preceded by reference scans against ambient light and a dark material. NIR measurements were acquired using a spectrometer (QEPro, Ocean Optics), with reflection profiles determined by customised software developed in-house using LabView (QEProInterface v3.1). Figure 1 shows the acquisition of a single NIR measurement on the anterior thigh. For this study, scans were acquired at four anatomical locations; sub-scapular, flank, mid-thigh and triceps. Two sets of measurements were acquired at each of these sites.

Fig. 1. Measurement conducted on the anterior thigh using the Ocean Optics QEPro NIR device

Anthropometric and body composition characteristics of the selected population are summarised in Table I.

Pre-processing, model development and statistical analysis were completed using Python (Python Software Foundation, version 2.7 https://www.python.org/) and the Sci-kit Learn package [16]. Pairs of reflection profiles at a given scan location that met the quality criteria (reflection values between 0 and 100%) and agreed within 10% in the 850-1100nm range were averaged prior to feature extraction. Features were constructed by determining exhaustive combinations of reflection at wavelengths in the range of 850-1100 nm, in increments of 10 nm and expressing the ratio between them

TABLE I
SUMMARY OF DATASET CHARACTERISTICS

Variable	Mean ± SD
n	41
Male (%)	48.8%
Age (months)	4.1 ± 1.7
Weight (g)	6835 ± 1477
Length (g)	62.2 ± 4.8
Subscapular skinfold (mm)	9.3 ± 1.9
Triceps skinfold (mm)	9.5 ± 1.9
Flank skinfold (mm)	11.5 ± 2.7
Mid-thigh skinfold (mm)	19.8 ± 3.6
Fat-free mass (g)	4889 ± 979
Fat mass (g)	1946 ± 645
Body fat percentage (%)	28.0 ± 5.2

as a percentage. This resulted in a set of 300 features for a given scan site.

For feature selection and model evaluation, the dataset (n = 41) was initially divided into two sex-stratified subsets, a training (n = 20) and a testing set (n = 21). In light of the small training set size available, we first reduced the feature set to those which exhibited a correlation ($p < 0.05$) with the target variable, fat mass as measured by deuterium dilution. On this reduced feature set, we then completed an exhaustive search of up to two features and the weight-for-length (W/L) ratio. For each combination of features, the mean estimation error from leave-one-out cross-validation was determined. This process involves fitting all except one sample from the training set to the target variable and evaluating the excluded sample, with a repeat of this process for every possible sample. The feature combination selected for model fitting was determined by minimisation of the mean estimation error.

A linear regression model for the esimation of fat mass in grams (g) was fit using the selected features. This model was subsequently evaluated on the testing set by determining correlation and Bland-Altman agreement with fat mass and corresponding BF% as estimated via the 4C model described in equation 1. This 4C model was based on that adopted by Fomon et al. [17] and Butte et al. [18].

$$FFM = \frac{1009.4108 \cdot TBW + 1000 \cdot BMC}{994 - 2.87851 \cdot TBK_{FFM}} \quad (1)$$

where fat-free mass (FFM), total body water (TBW) and bone mineral content (BMC) are in kg. The total body potassium (TBK) is in mEq/kg FFM and estimated for male and female infants using equations 2 and 3, respectively. These were based on polynomial curve fits to the table for TBK (mEQ/kg) from Fomon et al. [17].

$$TBK = 0.0009m^3 - 0.0505m^2 + 1.1047m + 49.108 \quad (2)$$

$$TBK = 0.0011m^3 - 0.0629m^2 + 1.2957m + 49.113 \quad (3)$$

To characterise the utility of the NIR model in this equation, we also evaluated model performance of three other

comparisons; a linear regression model fitted to infant BMI and sex, the Slaughter equation (equations 4 and 5 for male and female infants, respectively) where $\sum SFT$ is based on the sum of subscapular and triceps skinfold thicknesses in mm and DXA. The latter is often considered one of the criterion methods for infant body composition assessment, though its cost and portability considerations may hinder its application in low and middle income settings.

$$BF\%_M = 1.21 \times \sum SFT - 0.008 \times \sum SFT^2 - 1.7 \quad (4)$$

$$BF\%_F = 1.33 \times \sum SFT - 0.013 \times \sum SFT^2 - 2.5 \quad (5)$$

III. RESULTS

The linear regression model for estimation of infant fat mass (g) is expressed in equation 6, with coefficients summarised in Table II. Model coefficients showed both NIR features and W/L to be significant predictors of fat mass. Table III summarises Bland-Altman agreement and correlation for the model estimations of fat mass and BF% with the selected reference values.

$$FM = \beta_1 \frac{R_{920}}{R_{1090}} + \beta_2 \frac{R_{930}}{R_{1090}} + \beta_3 \frac{W}{L} + \varepsilon \quad (6)$$

TABLE II

LINEAR REGRESSION MODEL COEFFICIENTS FOR ESTIMATION OF FAT
MASS

Feature	Estimate	Standard error	p
β_1	-421.82	126.64	0.0042
β_2	539.30	141.79	0.0016
β_3	28.62	3.31	0.0016
ε	-12206.67	3874	0.0062

IV. DISCUSSION

Using a subset of n = 41 infants aged 2-8 months, we evaluated the potential for using NIR in estimation of infant body composition. Models were fitted to a training set (n = 20) and using fat mass and BF% as measured by deuterium dilution as the reference method. They were subsequently evaluated on the testing set (n = 20) against the 4C model estimates. Limits of agreement with the 4C model for in the testing set were -8 (-553, 536) for fat mass (g) and 0.1 (-7.07, 7.24) for BF%. These limits were comparatively narrower than from estimations based on the BMI model (-40 (-736, 656) g FM and -1.1 (-12.4, 10.2) BF%) and the Slaughter equation [13] (706 (10, 1401) g FM and 9.9 (1.1, 18.8) BF%). This suggests that NIR model does not strongly underestimate nor overestimate both fat mass and BF%. The BMI model tended to overestimate fat by 40 g with 95% of measurements falling within 736 g above and 656 g below the fat mass estimates from the 4C model. The Slaughter equation considerably underestimated infant fat mass, with a mean bias of 736 g in the testing set. The NIR model may

TABLE III

CORRELATION AND LIMITS OF AGREEMENT WITH INFANT FAT MASS (FM)
AND BODY FAT PERCENTAGE (BF%) FOR BOTH TRAINING AND TESTING
SETS[1]. THE NIR MODEL IS SHOWN ALONGSIDE OTHER MODELS BASED ON
A COMBINATION OF BODY MASS INDEX AND SEX (BMI), THE SLAUGHTER
EQUATION BASED ON SUM OF SKINFOLD THICKNESSES (SFT) AND DUAL
X-RAY ABSORPTIOMETRY (DXA)

	Model	Agreement$_{train}$	R_{train}	Agreement$_{test}$	R_{test}
FM (g)	NIR	0.0 (-414, 414)	0.95	-8 (-553, 536)	0.90
	BMI	0.0 (-868, 868)	0.76	-40 (-736, 656)	0.81
	SFT*	638 (-110, 1385)	0.88	706 (10, 1401)	0.86
	DXA	-216 (-1018, 587)	0.93	-134 (-891, 623)	0.92
BF%	NIR	-0.1 (-6.8, 6.6)	0.80	0.1 (-7.1, 7.2)	0.70
	BMI	-0.8 (-13.3, 11.7)	0.27	-1.1 (-12.4, 10.2)	0.33
	SFT	9.1 (-0.2, 18.5)	0.54	9.9 (1.1, 18.8)	0.47
	DXA	-2.9 (-13.2, 7.4)	0.83	-1.4 (-11.8, 9)	0.77

1 All training set statistics reported against the deuterium dilution estimations, while testing sets were validated against the 4C model
* Note that no model fitting was undertaken for training set skinfold thickness models. Bland-Altman analysis and correlation reported for the Slaughter prediction equations [13]

have exhibited stronger agreement and lower mean bias in the testing set as it was fit specifically to the given population of South African infants.

The model trained using BMI (weight-for-length-squared) and sex features exhibited the poorest (R = 0.76) model fit to fat mass in training set, compared with 0.95 and 0.88 for NIR and SFT, respectively. This also translated to the poor correlation (R = 0.27) with BF% as predicted by the 4C model. It is possible that infant fat mass could not be sufficiently explained by BMI and sex variables or that these variables were better correlated with an overall measure of BF%. BMI may have also been influenced by inaccuracies in the measurement of length [3].

The large mean bias of 706 g observed for fat mass estimations for the Slaughter equation suggests a tendency for the SFT model to underestimate fat mass. This may have been due the high inter-individual variability observed in childhood and the population-dependence of these anthropometric models. Previous validations of the Slaughter equation on neonates and infants were undertaken on populations of mostly Caucasian subjects [14][15] and there is a general understanding that anthropometry varies between infants of different sexes and ethnicities [19][20][21]. The correlation (R = 0.86) with fat mass suggests that these anthropometric features (sum of triceps and subscapular skinfolds) are still able to explain considerable variance in infant body composition.

DXA is among the criterion methods for infant body composition assessment, though it may not be suited to use in low and middle income settings due to cost, operational expertise required and portability. It consistently exhibited strong correlation in fat mass and BF% predictions, though it displayed a mean bias of -134 g and -1.4 BF% for testing set estimations, with wider limits agreement than the NIR model. This could be explained by the large contribution

of deuterium dilution against which the NIR model was fitted to the total body water component of the 4C model. This stands in contrast to DXA bone mineral content which forms a smaller portion of the compartment model. In the infant population, there also remains uncertainty around the agreement between gold standard methods for body composition assessment [22].

Feature selection was based on the combination of wavelength ratios which minimised the mean estimation error from leave-one-out cross-validation of the training set. The selected features included R_{920}/R_{1090} and R_{930}/R_{1090}, interestingly close to the peak of pure fat at 930nm as reported by Conway et al [5]. This suggests a direct relationship between the attenuation observed at 930nm and the variation in fat mass. The significant contribution ($p < 0.01$) of these features to the estimation of fat mass is in line with previous observations [6][7], though direct comparisons to the reported correlations are difficult given the different dependent variables (fat mass rather than BF%) and reference methods (the 4C model rather than air-displacement plethysmography).

The small dataset size (n = 41) limited the possible number of features and scan locations for inclusion in model development, though we applied leave-one-out cross-validation as a means of mitigating this. Dataset size also hindered further subdivisions of the data which, for example, by age, which may have impacted the model; BF% has been observed to plateau in infants in the 6-9 month age range [18]. A larger sample size of this population will be available for further validation of this model, with n = 651 infants recruited in total. Total body potassium (TBK) was not measured independently via a potassium counter and instead was estimated based on reference equations, which may impact the accuracy of the 4C model.

V. CONCLUSIONS

This study evaluated the potential of near-infrared interactance for use in infant body composition assessment. The developed models exhibited a mean difference of -8 (-553, 536) for fat mass (g) and 0.1 (-7.07, 7.24) for BF% as measured by a 4C model. They exhibited narrower limits of agreement compared with a BMI and sex-based linear regression model, the Slaughter [13] equation using sum of skinfold thickness measurements and the estimations from DXA. Subject to further model development and validation on a larger dataset, this model combining NIR features of reflection ratios and W/L offers a simple, fast and low-cost approach for infant body composition in low and middle income settings where the burden of infant deaths is greatest.

REFERENCES

[1] UNICEF. *Facts for life*. unicef, 2010.

[2] Joy E Lawn, Simon Cousens, Jelka Zupan, Lancet Neonatal Survival Steering Team, et al. 4 million neonatal deaths: when? where? why? *The lancet*, 365(9462):891–900, 2005.

[3] Anna J Wood, Camille H Raynes-Greenow, Angela E Carberry, and Heather E Jeffery. Neonatal length inaccuracies in clinical practice and related percentile discrepancies detected by a simple length-board. *Journal of paediatrics and child health*, 49(3):199–203, 2013.

[4] Jane West, Ben Manchester, John Wright, Debbie A Lawlor, and Dagmar Waiblinger. Reliability of routine clinical measurements of neonatal circumferences and research measurements of neonatal skinfold thicknesses: findings from the born in bradford study. *Paediatric and perinatal epidemiology*, 25(2):164–171, 2011.

[5] Joan M Conway, Karl H Norris, and CE Bodwell. A new approach for the estimation of body composition: infrared interactance. *The American journal of clinical nutrition*, 40(6):1123–1130, 1984.

[6] A McEwan, S Bian, G Gargiulo, R Morhard, P Jones, F Mustafa, BE BeN, and H Jeffery. Low-cost near-infrared measurement of subcutaneous fat for newborn malnutrition. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 90600A–90600A. International Society for Optics and Photonics, 2014.

[7] Fatin Hamimi Mustafa, Emily J Bek, Jacqueline Huvanandana, Peter W Jones, Angela E Carberry, Heather E Jeffery, Craig T Jin, and Alistair L McEwan. Length-free near infrared measurement of newborn malnutrition. *Scientific reports*, 6, 2016.

[8] Elizabeth L Prado and Kathryn G Dewey. Nutrition and brain development in early life. *Nutrition reviews*, 72(4):267–284, 2014.

[9] David A Levitsky and Barbara J Strupp. Malnutrition and the brain: changing concepts, changing concerns. *The Journal of nutrition*, 125(8):2212S, 1995.

[10] David JP Barker, Clive Osmond, Eero Kajantie, and Johan G Eriksson. Growth and chronic disease: findings in the helsinki birth cohort. *Annals of human biology*, 36(5):445–458, 2009.

[11] David JP Barker. In utero programming of chronic disease. *Clinical science*, 95(2):115–128, 1998.

[12] Angela E Carberry, Camille H Raynes-Greenow, Robin M Turner, Lisa M Askie, and Heather E Jeffery. Is body fat percentage a better measure of undernutrition in newborns than birth weight percentiles? *Pediatric research*, 74(6):730–736, 2013.

[13] Mary H Slaughter, TG Lohman, RAet Boileau, CA Horswill, RJ Stillman, MD Van Loan, and DA Bemben. Skinfold equations for estimation of body fatness in children and youth. *Human biology*, pages 709–723, 1988.

[14] Barbara E Lingwood, Anne-Martine Storm van Leeuwen, Angela E Carberry, Erin C Fitzgerald, Leonie K Callaway, Paul B Colditz, and Leigh C Ward. Prediction of fat-free mass and percentage of body fat in neonates using bioelectrical impedance analysis and anthropometric measures: validation against the pea pod. *British Journal of Nutrition*, 107(10):1545–1552, 2012.

[15] Hansjörg Rudolf Schmelzle and Christoph Fusch. Body fat in neonates and young infants: validation of skinfold thickness versus dual-energy x-ray absorptiometry. *The American journal of clinical nutrition*, 76(5):1096–1100, 2002.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] Samuel J Fomon and Steven E Nelson. Body composition of the male and female reference infants. *Annual review of nutrition*, 22(1):1–17, 2002.

[18] Nancy F Butte, Judy M Hopkinson, William W Wong, E O'Brian Smith, and Kenneth J Ellis. Body composition during the first 2 years of life: an updated reference. *Pediatric research*, 47(5):578–585, 2000.

[19] David A Fields, Sowmya Krishnan, and Amy B Wisniewski. Sex differences in body composition early in life. *Gender medicine*, 6(2):369–375, 2009.

[20] Tai-Fai Fok, Kam-Lun Ellis Hon, Pak-Cheung Ng, Ming-Chung Eric Wong, Hung-Kwan So, Tak-Fai Joseph Lau, Chun-Bong Chow, and Wai-Hong Lee. Normative data for triceps and subscapular skinfold thicknesses of chinese infants. *Acta Paediatrica*, 95(12):1614–1619, 2006.

[21] G Rodriguez, MP Samper, JL Olivares, P Ventura, LA Moreno, and JM Perez-Gonzalez. Skinfold measurements at birth: sex and anthropometric influence. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 90(3):F273–FF275, 2005.

[22] SV Wrottesley, PT Pisa, LK Micklesfield, JM Pettifor, and SA Norris. A comparison of body composition estimates using dual-energy x-ray absorptiometry and air-displacement plethysmography in south african neonates. *European journal of clinical nutrition*, 70(11):1254–1258, 2016.

6.3 Concluding remarks

The work presented in this chapter addresses the fifth aim of the thesis to identify infants at risk of malnutrition and later morbidity. We apply feature extraction and selection techniques to develop two sets of models: logistic regression models to identify malnutrition, measured by a composite neonatal morbidity as well as by low and high body fat percentage (BF%), and linear regression models for the estimation of infant fat mass and BF%. In our cohort of (n = 524) term neonates, we identified a composite neonatal morbidity with an area under receiver-operator characteristic curve (AUC with 95% confidence interval) of 0.740 (0.63, 0.85). We also developed a model using near-infrared interactance features that exhibited a correlation of 0.90 and Bland-Altman agreement (mean bias with 95% confidence intervals) of -8 (-4553, 536) g for fat mass with a gold standard method. These results represent improved performance and narrower limits of agreement compared with other accessible methods such as BMI and sex-based linear regression models, skinfold thickness measurements and the estimations from DXA.

Chapter 7

Conclusions and Future Work

This chapter summarises the research work undertaken as part of this PhD which focuses on the application of advanced analyses to physiological signals in neonatal intensive care. These applications range from the early identification of infants at risk for later morbidity including IVH and nutritional deficiency, to the characterisation of cerebral autoregulation and the physiological impact of caffeine therapy.

7.1   Major contributions of the current work

This section offers an overview of the main findings and/or contributions of each chapter.

Chapter 2 Literature Review

This chapter critically appraised the available literature on physiological variability and its relevance in neonatal intensive care. It discussed current of advanced analyses to assist NICU caregivers in decision-making/risk assessment, and to characterise underlying mechanisms affecting physiological variability. This chapter also discussed the limitations of implementation, challenges in clinical application, and potential pathways of development.

Chapter 3 Identification of preterm infants at risk of IVH

This chapter examined the potential for early identification of infants at risk of developing IVH. We applied DFA to arterial blood pressure and air flow data within 1-3 hours of birth, and found a combination of linear and non-linear features (mean diastolic blood pressure and the pulse interval α_2) able to discern infants who later developed IVH with a sensitivity $> 90\%$ at a specificity of 75%.

This chapter highlighted the importance of pre-processing in the application of DFA and the predictive value of combining both linear and non-linear metrics in quantifying variability.

Chapter 4 Characterising changes in cardiovascular dynamics following caffeine therapy

This chapter applied both linear (mean, SD) and non-linear (DFA: α_1, α_2 and Poincare: SD1, SD2 and SD1/SD2) analyses to physiological data from a cohort of preterm infants receiving caffeine therapy. We found altered non-linear dynamics in heart rate and arterial blood pressure control using DFA and Poincare analysis that suggest elevated parasympathetic activity following a loading dose of 10 mg/kg caffeine base. This work has demonstrated increased pulse pressure variability (SD) in the 2 hours post-caffeine, which may have important implications for caffeine administration in preterm infants with impaired cerebral autoregulation. The findings in this chapter also stand in contrast to the only other study examining these techniques [35] which reported no altered cardiovascular activity in a cohort of 21 infants. This may have been due to differences in varying gestational age between the cohorts, and the use of a mixed-modelling approach in our study to compare changes over multiple windows of analyses, rather than a single timepoint before and after caffeine administration.

Chapter 5 Characterising cerebral autoregulation in preterm infants

This chapter evaluated a range of time and frequency-domain techniques to characterise the correlation between two physiological signals, some of which were reviewed in Chapter 2. We observed a reduction in time-domain correlation following a loading dose of 10 mg/kg caffeine base. These observations contribute to our understanding of CAP trial results, where caffeine was associated with improved neurodevelopmental outcomes at 18-21 months follow-up [12] (though not persisting at 5 years [21]), and reduced risk of motor impairment at 11 years follow-up [22]. We also demonstrated the potential of detrended cross-correlation analysis in characterising cerebral autoregulation, as shown by its discriminative value in identifying the effects of caffeine and its correlation with other time domain techniques.

Chapter 6 Malnutrition detection and nutritional status assessment using near-infrared interactance

This chapter focused on the nutritional aspect of neonatal care which plays a role in morbidity and mortality in later life. The developed models were able to identify a composite measure of morbidity with an AUC of 0.75 (0.63, 0.85), and low BF% neonates with an AUC of 0.836 (0.79, 0.88). NIR interactance signals were also shown to have predictive value in estimating fat mass and BF%, with the fitted model exhibiting narrower limits of agreement with a gold standard method compared to other accessible methods for body composition assessment. The models presented in this chapter have shown potential for screening infants in need of further nutritional management in the NICU and in low and middle income settings.

7.2 Future research directions

Chapter 3 Identification of preterm infants at risk of IVH

While we have performed internal validations and sensitivity analyses to evaluate the robustness of the results, these models require validation on an independent and larger dataset as a next step towards their use in the NICU. Continuous measures of systemic blood pressure were obtained via arterial lines. With the shift towards non-invasive monitoring, it may be of interest to explore photoplethysmography as an alternative to extract similar pulse interval and amplitude-based features used in the final model.

The model could also be extended for multi-class classification, allowing for stratification of the IVH classes according to the Papile system (I to IV). Further work in the area may involve the inclusion of cross-correlation features or other variability analysis techniques that may account for the interaction between respiratory and blood pressure signals which may better capture patient-ventilator asynchrony.

Chapter 4 Characterising changes in cardiovascular dynamics following caffeine therapy

There is continued interest in caffeine therapy and its acute and long-term effects on preterm infants. We have shown that non-linear analysis can characterise cardiovascular impact following a loading dose of caffeine in preterm infants. This analysis may be extended to compare the differences in non-linear dynamics

between a standard 20 mg/kg and higher 80 mg/kg loading dose of caffeine citrate, and similarly, for early versus routine (12 hours after birth) administration.

Chapter 5 Characterising cerebral autoregulation in preterm infants

The work in this chapter may be extended understand tolerance effects during maintenance doses for caffeine therapy. Given the suggestion that impaired cerebral autoregulation may be a causal pathway for brain injury such as IVH, it may also be of interest to understand how these metrics change in the lead up to diagnosis, though no specific treatment for IVH has reached consensus. We may nevertheless continue to characterise the risk factors to help guide clinical care for its prevention. Further work in clarifying the physiological meaning of and other factors contributing to the detrended cross-correlation coefficient is also needed, alongside validation against a gold standard measure of autoregulation, as was done for the cerebral oximetry index in an infant animal model [38]. The application of detrended cross-correlation analysis to describe cerebral autoregulation also requires the inter- and intra-subject variability to be evaluated.

Chapter 6 Malnutrition detection and nutritional status assessment using near-infrared interactance

We applied feature selection techniques for development of models in identifying risks of nutritional deficiency and later morbidity. The anthropometric models for detection of a composite neonatal morbidity requires further validation on a dataset with greater representation of malnutrition. Given the small sample size of the training and testing set for development of the near-infrared interactance model, this approach should be extended to the larger/complete dataset when available. Feature selection and dimensionality reduction (such as principal component analysis or factor analysis) may also yield different key features in a larger cohort and thus, external validation on an independent dataset with greater representation of morbidity and/or malnutrition is also required.

7.3   Summary

The monitoring and management of preterm infants neonatal intensive care represents a unique challenge, balancing life-threatening conditions, treatments or procedures with potentially life-long implications, and the demands of extrauterine life. The opportunities to apply advanced analyses in this context are varied,

contributing to a greater understanding of the physiological variability in relation to conditions such as IVH, mechanisms such as cerebral autoregulation, malnutrition as a risk factor for later morbidity, and the effects of prescribed treatments. These analyses may be incorporated into a multi-parameter model for risk stratification of patients in neonatal intensive care. One example from adult intensive care is the continuous individualised multiorgan variability analysis (CIMVA) software which computes features such as those of heart rate and respiratory rate variability from multiple physiological waveforms [50]. These analyses may also be incorporated into monitoring equipment and made available to caregivers, which may facilitate earlier interventions for critical events/conditions and improved outcomes in both the short and long term.

References

[1] Andrew JE Seely and Peter T Macklem. Complex systems and the technology of variability analysis. *Critical care*, 8(6):R367, 2004.

[2] Srinivas Bolisetty, Anjali Dhawan, Mohamed Abdel-Latif, Barbara Bajuk, Jacqueline Stack, Kei Lui, et al. Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants. *Pediatrics*, pages peds–2013, 2013.

[3] Joseph J Volpe. Cerebellum of the premature infant: rapidly developing, vulnerable, clinically important. *Journal of child neurology*, 24(9):1085–1104, 2009.

[4] Joseph J Volpe. Neurology of the newborn. 5th. *Philadelphia: Saunders Elsevier*, 2008.

[5] Avroy A Fanaroff, Barbara J Stoll, Linda L Wright, Waldemar A Carlo, Richard A Ehrenkranz, Ann R Stark, Charles R Bauer, Edward F Donovan, Sheldon B Korones, Abbot R Laptook, et al. Trends in neonatal morbidity and mortality for very low birthweight infants. *American Journal of Obstetrics & Gynecology*, 196(2):147–e1, 2007.

[6] Eva Landmann, Björn Misselwitz, Jens O Steiss, and Ludwig Gortner. Mortality and morbidity of neonates born at¡ 26 weeks of gestation (1998–2003). a population-based study. *Journal of perinatal medicine*, 36(2):168–174, 2008.

[7] Jeffrey M Perlman, Joseph B McMenamin, and Joseph J Volpe. Fluctuating cerebral blood-flow velocity in respiratory-distress syndrome: relation to the development of intraventricular hemorrhage. *New England Journal of Medicine*, 309(4):204–209, 1983.

[8] Jeffrey M Perlman, Steven Goodman, Katherine L Kreusser, and Joseph J Volpe. Reduction in intraventricular hemorrhage by elimination of fluctuating cerebral blood-flow velocity in preterm infants with respiratory distress syndrome. *New England Journal of Medicine*, 312(21):1353–1357, 1985.

[9] Gerhard Pichler, Berndt Urlesberger, and Wilhelm Müller. Impact of brady-cardia on cerebral oxygenation and cerebral blood volume during apnoea in preterm infants. *Physiological measurement*, 24(3):671, 2003.

[10] David J Henderson-Smart and Antonio G De Paoli. Methylxanthine treatment for apnoea in preterm infants. *The Cochrane Library*, 2010.

[11] Barbara Schmidt, Robin S Roberts, Peter Davis, Lex W Doyle, Keith J Barrington, Arne Ohlsson, Alfonso Solimano, and Win Tin. Caffeine therapy for apnea of prematurity. *New England Journal of Medicine*, 354(20):2112–2121, 2006.

[12] Barbara Schmidt, Robin S Roberts, Peter Davis, Lex W Doyle, Keith J Barrington, Arne Ohlsson, Alfonso Solimano, and Win Tin. Long-term effects of caffeine therapy for apnea of prematurity. *New England Journal of Medicine*, 357(19):1893–1902, 2007.

[13] Elizabeth L Prado and Kathryn G Dewey. Nutrition and brain development in early life. *Nutrition reviews*, 72(4):267–284, 2014.

[14] David A Levitsky and Barbara J Strupp. Malnutrition and the brain: changing concepts, changing concerns. *The Journal of nutrition*, 125(8):2212S, 1995.

[15] David JP Barker. In utero programming of chronic disease. *Clinical science*, 95(2):115–128, 1998.

[16] David JP Barker, Clive Osmond, Eero Kajantie, and Johan G Eriksson. Growth and chronic disease: findings in the helsinki birth cohort. *Annals of human biology*, 36(5):445–458, 2009.

[17] Andrea Bravi, André Longtin, and Andrew JE Seely. Review and classification of variability analysis techniques with clinical applications. *Biomedical engineering online*, 10(1):90, 2011.

[18] Conny MA van Ravenswaaij-Arts, Jeroen CW Hopman, Louis AA Kollée, Joop PL van Amen, Gerard BA Stoelinga, and Herman P van Geijn. The influence of respiratory distress syndrome on heart rate variability in very preterm infants. *Early human development*, 27(3):207–221, 1991.

[19] BD Hanna, MN Nelson, Rosemary C White-Traut, JM Silvestri, U Vasan, P Meleedy Rey, MK Patel, and E Comiskey. Heart rate variability in preterm brain-injured and very-low-birth-weight infants. *Neonatology*, 77(3):147–155, 2000.

[20] Volkan Tuzcu, Selman Nas, Umit Ulusar, Ahmet Ugur, and Jeffrey R Kaiser. Altered heart rhythm dynamics in very low birth weight infants with impending intraventricular hemorrhage. *Pediatrics*, 123(3):810–815, 2009.

[21] Barbara Schmidt, Peter J Anderson, Lex W Doyle, Deborah Dewey, Ruth E Grunau, Elizabeth V Asztalos, Peter G Davis, Win Tin, Diane Moddemann, Alfonso Solimano, et al. Survival without disability to age 5 years after neonatal caffeine therapy for apnea of prematurity. *Jama*, 307(3):275–282, 2012.

[22] Barbara Schmidt, Robin S Roberts, Peter J Anderson, Elizabeth V Asztalos, Lorrie Costantini, Peter G Davis, Deborah Dewey, Judy Dilario, Lex W Doyle, Ruth E Grunau, et al. Academic performance, motor function, and behavior 11 years after neonatal caffeine citrate therapy for apnea of prematurity: an 11-year follow-up of the cap randomized clinical trial. *JAMA pediatrics*, 171(6):564–572, 2017.

[23] Christina Hoecker, Mathias Nelle, Johannes Poeschl, Bernd Beedgen, and Otwin Linderkamp. Caffeine impairs cerebral and intestinal blood flow velocity in preterm infants. *Pediatrics*, 109(5):784–787, 2002.

[24] V Soloveychik, A Bin-Nun, A Ionchev, S Sriram, and W Meadow. Acute hemodynamic effects of caffeine administration in premature infants. *Journal of Perinatology*, 29(3):205–208, 2009.

[25] MB Tracy, J Klimek, M Hinder, G Ponnampalam, and SK Tracy. Does caffeine impair cerebral oxygenation and blood flow velocity in preterm infants? *Acta Paediatrica*, 99(9):1319–1323, 2010.

[26] Jacqueline Huvanandana, Chinh Nguyen, Cindy Thamrin, Mark Tracy, Murray Hinder, and Alistair L McEwan. Prediction of intraventricular haemorrhage in preterm infants using time series analysis of blood pressure and respiratory signals. *Scientific Reports*, 7, 2017.

[27] Ying Zhang, Gregory SH Chan, Mark B Tracy, Murray Hinder, Andrey V Savkin, and Nigel H Lovell. Detrended fluctuation analysis of blood pressure in preterm infants with intraventricular hemorrhage. *Medical & biological engineering & computing*, 51(9):1051–1057, 2013.

[28] Jacqueline Huvanandana, Cindy Thamrin, Mark Tracy, Murray Hinder, Chinh Nguyen, and Alistair McEwan. Advanced analyses of physiological signals in the neonatal intensive care unit. *Physiological measurement*, 38(10):R253, 2017.

[29] Paolo Castiglioni, Gianfranco Parati, Marco Di Rienzo, Roberta Carabalona, Andrei Cividjian, and Luc Quintin. Scale exponents of blood pressure and heart rate during autonomic blockade as assessed by detrended fluctuation analysis. *The Journal of physiology*, 589(2):355–369, 2011.

[30] Luiz Eduardo Virgilio Silva, Carlos Alberto Aguiar Silva, Helio Cesar Salgado, and Rubens Fazan. The role of sympathetic and vagal cardiac control on complexity of heart rate dynamics. *American Journal of Physiology-Heart and Circulatory Physiology*, 312(3):H469–H477, 2017.

[31] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE transactions on biomedical engineering*, 48(11):1342–1347, 2001.

[32] Mary A Woo, William G Stevenson, Debra K Moser, Robert B Trelease, and Ronald M Harper. Patterns of beat-to-beat heart rate variability in advanced heart failure. *American heart journal*, 123(3):704–710, 1992.

[33] Reese H Clark, Barry T Bloom, Alan R Spitzer, and Dale R Gerstmann. Reported medication use in the neonatal intensive care unit: data from a large national data set. *Pediatrics*, 117(6):1979–1987, 2006.

[34] Jacqueline Huvanandana, Cindy Thamrin, Alistair L McEwan, Murray Hinder, and Mark B Tracy. Cardiovascular impact of intravenous caffeine in preterm infants. *Acta Paediatrica*, 2018.

[35] I Ulanovsky, NS Haleluya, S Blazer, and A Weissman. The effects of caffeine on heart rate variability in newborns with apnea of prematurity. *Journal of Perinatology*, 34(8):620, 2014.

[36] Rong Zhang, Julie H Zuckerman, Cole A Giller, and Benjamin D Levine. Transfer function analysis of dynamic cerebral autoregulation in humans. *American Journal of Physiology-Heart and Circulatory Physiology*, 274(1):H233–H241, 1998.

[37] Cole Giller and Domenico Gerardo Iacopino. Use of middle cerebral velocity and blood pressure for the analysis of cerebral autoregulation at various frequencies: the coherence index. *Neurological research*, 19(6):634–640, 1997.

[38] Ken M Brady, Jennifer K Lee, Kathleen K Kibler, Piotr Smielewski, Marek Czosnyka, R Blaine Easley, Raymond C Koehler, and Donald H Shaffner.

Continuous time-domain analysis of cerebrovascular autoregulation using near-infrared spectroscopy. *Stroke*, 38(10):2818–2825, 2007.

[39] Joseph J Volpe. Intraventricular hemorrhage in the premature infantcurrent concepts. part i. *Annals of neurology*, 25(1):3–11, 1989.

[40] Nicholas D Embleton, Jemma Cleminson, and Stefan Zalewski. What growth should we aim for in preterm neonates? *Paediatrics and Child Health*, 27(1):18–22, 2017.

[41] Richard A Ehrenkranz, Naji Younes, James A Lemons, Avroy A Fanaroff, Edward F Donovan, Linda L Wright, Vasilis Katsikiotis, Jon E Tyson, William Oh, Seetha Shankaran, et al. Longitudinal growth of hospitalized very low birth weight infants. *Pediatrics*, 104(2):280–289, 1999.

[42] Reese H Clark, Pam Thomas, and Joyce Peabody. Extrauterine growth restriction remains a serious problem in prematurely born neonates. *Pediatrics*, 111(5):986–990, 2003.

[43] Rachel Huxley, Andrew Neil, and Rory Collins. Unravelling the fetal origins hypothesis: is there really an inverse association between birthweight and subsequent blood pressure? *The Lancet*, 360(9334):659–665, 2002.

[44] Johan G Eriksson, Tom Forsen, Jaako Tuomilehto, Clive Osmond, and David JP Barker. Early growth and coronary heart disease in later life: longitudinal study. *Bmj*, 322(7292):949–953, 2001.

[45] Marcus Richards, Rebecca Hardy, Diana Kuh, and Michael EJ Wadsworth. Birthweight, postnatal growth and cognitive function in a national uk birth cohort. *International Journal of Epidemiology*, 31(2):342–348, 2002.

[46] Kenneth J Ellis. Evaluation of body composition in neonates and infants. In *Seminars in Fetal and Neonatal Medicine*, volume 12, pages 87–91. Elsevier, 2007.

[47] Kenneth J Ellis, Manjiang Yao, Roman J Shypailo, Alessandro Urlando, William W Wong, and William C Heird. Body-composition assessment in infancy: air-displacement plethysmography compared with a reference 4-compartment model–. *The American journal of clinical nutrition*, 85(1):90–95, 2007.

[48] Guansheng Ma, Manjiang Yao, Yan Liu, Aiwei Lin, Hui Zou, Alessandro Urlando, William W Wong, Laurie Nommsen-Rivers, and Kathryn G Dewey.

Validation of a new pediatric air-displacement plethysmograph for assessing body composition in infants. *The American journal of clinical nutrition*, 79(4):653–660, 2004.

[49] Mary H Slaughter, TG Lohman, RAet Boileau, CA Horswill, RJ Stillman, MD Van Loan, and DA Bemben. Skinfold equations for estimation of body fatness in children and youth. *Human biology*, pages 709–723, 1988.

[50] A Bravi. Cimva core description manual, 2013.

Appendix

Appendix I

Logistic regression models for predicting intraventricular haemorrhage in preterm infants using respiratory and blood pressure signals

Logistic regression models for predicting intraventricular haemorrhage in preterm infants using respiratory and blood pressure signals

Jacqueline Huvanandana *, Cindy Thamrin†, Chinh Nguyen†, Mark Tracy‡§ and Murray Hinder*‡ Alistair McEwan*

*School of Electrical Engineering, University of Sydney
†Woolcock Institute of Medical Research, University of Sydney
‡Westmead Hospital, Sydney, Australia
§School of Paediatrics and Child Health, University of Sydney

Abstract—**Despite the decline in mortality rates for extremely preterm infants, intraventricular haemorrhage (IVH) remains a threat to their survival. In this study, we sought to explore logistic regression models for predicting IVH as they would be applied in a clinical setting, using features derived from respiratory and blood pressure signals. Calculated predictors included mean (μ) and the short- and long-term scaling exponents (α_1, α_2) from detrended fluctuation analysis. The model fitted with short-term scaling exponent (α_1) of the beat-to-beat diastolic blood pressure (DBP) exhibited an area under receiver-operator characteristic curve (AUC) of 0.788 (0.62, 0.96), with a sensitivity of approximately 0.875 at a specificity of 0.75. Of the multivariable models explored, the highest AUC was 0.831 (0.66, 1.00), combining μ_{DBP} with α_1 of the beat-to-beat systolic blood pressure (SBP).**

I. INTRODUCTION

The past decades have seen a decline in mortality rates for extremely preterm infants, though intraventricular haemorrhage (IVH) remains common in survivors, often causing permanent brain injury and being associated with poor neurodevelopmental outcomes [1]. Prenatal factors leading to preterm birth inter-relate to the need for resuscitation and cardiorespiratory management within the first 24 hours of life. These play an important role in the development and timing of IVH [2]. The potential to identify infants at high risk of developing IVH may inform management at this critical stage.

Motivated by retrospective studies that have observed altered autonomic functions in cases of IVH [3][4], recent studies have explored detrended fluctuation analysis (DFA) as a means of quantifying this. Their findings suggest the relevance of blood pressure and heart rate variability in distinguishing infants with IVH and those without, characterised by the short-term scaling exponent from DFA [5][6]. This is also in alignment with the reported association of IVH with cerebral blood pressure passivity, that is, where linear changes in cerebral blood flow result in linear changes to blood pressure [7]. Early studies have demonstrated associations between of patient-ventilator asynchrony and increased incidence of IVH [8][9]. This study aimed to explore the potential of features derived from blood pressure and respiratory signals in dis-

cerning preterm infants with IVH from those without, in both univariate and multivariable logistic regression models. This entailed an evaluation of the robustness of these predictors as applied in a clinical context.

II. METHODS

A. Data Collection

Data collection was undertaken at a large tertiary neonatal intensive care unit in Sydney, Australia, as part of a prospective clinical investigation, following informed parental consent. Physiological data, including intra-arterial blood pressure measurements and ventilator-derived air flow was collected from the infants in the 1-3 h period following birth. The cohort consisted of low birth-weight (< 1500 g), a gestational age of < 30 weeks and an absence of significant congenital anomalies.

The study was approved by the Sydney West Area Health Service Human Research and Ethics and conducted in accordance with the World Medical Association Declaration of Helsinki.

B. Signal Processing and Data Analysis

Of the 46 infants enrolled, 28 had sufficiently long and artefact-free arterial blood pressure and airflow recordings. The former was measured using an umbilical or peripheral arterial catheter, collected using a bedside patient monitor (Philips Agilent Systems, Philip Healthcare, North Ryde, Australia), and the raw airflow wave from a ventilator (Babylog 8000, Drägerwerk, Lübeck, Germany). Both signals were sampled at 1 kHz and recorded by a data acquisition system (ADInstruments, Sydney, Australia). Of these 28 infants, 8 subsequently developed IVH and all except 1 were mechanically-ventilated.

Signal processing and feature extraction was completed in Python (Python Software Foundation, version 2.7. https://www.python.org/). Prior to analysis, both arterial blood pressure and airflow signals were down-sampled to 125 Hz for computational efficiency. This was sufficient for peak detection of the respective signals. A sliding 10 min window, shifted

in 30 sec increments, was used across the total recording length. This was done to simulate the manner in which these predictors could be calculated in a clinical setting. Windows which satisfied the quality criteria were included for analysis. This quality criteria involved defining bounds for the allowable number of detected beats ($40 - 250$ beats per minute) and breaths (> 20 breaths per minute) in a given window. The windows were also excluded if no pulse was detected for a 15 sec interval or large spikes in the systolic blood pressure were identified. For each 10 min window that qualified, the extracted features were as follows; the mean (μ) and DFA-derived short- and long-term scaling exponents (α_1, α_2, respectively) of the beat-to-beat mean-arterial (MAP, mmHg), systolic (SBP, mmHg) and diastolic (DBP, mmHg) blood pressure, as well as the pulse interval (PI, ms) between consecutive diastoles, delineating a complete cycle. From the corresponding airflow window, μ, α_1 and α_2 were also extracted using the inter-breath (IBI, ms) intervals, delineated by consecutive maxima.

DFA is a non-linear, time domain-based method for quantifying long-range power law correlations. Developed by Peng and co-workers [10], it involves the removal of the local linear trend prior to determining the root-mean squared fluctuation, thus accounting for non-stationarities which often confound analyses of real-world signals. This fluctuation is defined as:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n]^2} \qquad (1)$$

where $y(k)$ is any given time series, $y_n(k)$ the local linear trend for a given segment, and N the number of data points in the series for a given round of analysis. The scaling exponents represent the gradients of the log-log relationship of n and fluctuation $F(n)$ over a particular window size. In this case, the window sizes defined for the short- and long-term scaling exponents were 4-15 and 15-50 beats, respectively [5]. The application of DFA is further explained by Thamrin et al [11].

The means of each feature across all selected running windows were used to fit both univariate and multivariable logistic regression models. Model performance was evaluated based on the area under the receiver-operator characteristic curve (AUC).

III. RESULTS

A. Univariate Predictors of IVH

A Students T-test comparison of the individual predictors between the two groups is shown in Table I. Note that the features represent the mean across all windows selected for analysis of a given recording.

Table II displays the performance of all fitted univariate logistic regression models. The p-values reported for the ROC are based on a comparison with a chance classifier (AUC = 0.5). The 95% confidence intervals (CIs) for the AUC are also provided [12]. On the basis of confidence intervals that did not span across AUC = 0.5 and the p_{ROC} values reported, the statistically significant ROC curves were found to be the

TABLE I
COMPARISON OF FEATURES BETWEEN INFANTS WHO LATER DEVELOPED INTRAVENTRICULAR HAEMORRHAGE (IVH) AND THOSE WHO DID NOT (NON-IVH)

Feature	IVH	Non-IVH	p
MAP			
μ	32.2±5.7 mmHg	35.2±4.7 mmHg	0.187
α_1	0.95±0.16	0.78±0.19	0.038
α_2	1.10±0.06	1.00±0.18	0.152
SBP			
μ	41.1±9.3 mmHg	42.7±5.4 mmHg	0.608
α_1	0.85±0.12	0.69±0.15	0.015
α_2	1.06±0.09	0.97±0.16	0.155
DBP			
μ	25.0±3.7 mmHg	29.0±4.6 mmHg	0.049
α_1	0.85±0.11	0.68±0.20	0.037
α_2	1.02±0.08	0.93±0.16	0.126
PI			
μ	421±26 ms	419±33 ms	0.908
α_1	0.62±0.09	0.62±0.08	0.905
α_2	0.76±0.20	0.65±0.11	0.152
IBI			
μ	950±71 ms	950±71 ms	0.864
α_1	0.58±0.05	0.57±0.06	0.763
α_2	0.54±0.06	0.52±0.04	0.252

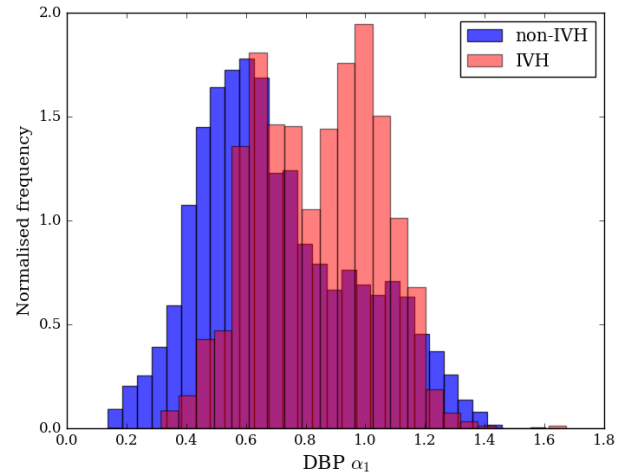Values are reported as mean±SD

Fig. 1. Histograms of DBP α_1 for IVH and non-IVH groups over all analysed windows

short-term scaling exponents (α_1) for the beat-to-beat MAP, SBP and DBP. The long-term scaling exponent α_2 for DBP exhibited an AUC of 0.706 ($p < 0.05$), despite not being significantly different ($p > 0.05$) across the two groups from a t-test comparison (Table I). The only predictive linear feature was mean DBP, yielding an AUC of 0.750 (0.55, 0.95). Given space constraints, we have included the histograms of only the best performing feature (DBP α_1) exhibiting an AUC of 0.788 (0.62, 0.96) in Fig. 1, and one of the respiratory features (IBI α_2) with an AUC of 0.644 (0.37, 0.92) for reference in Fig. 2.

TABLE II
UNIVARIATE LOGISTIC REGRESSION MODELS

Model	AUC	p_{ROC}	Odds Ratio	p_{OR}
MAP				
μ	0.638 (0.37, 0.91)	0.318	3.00 (0.54,16.69)	0.406
α_1	0.781 (0.61, 0.96)	0.002	9.00 (1.35,59.79)	0.043
α_2	0.688 (0.49, 0.88)	0.06	5.00 (0.87,28.86)	0.151
SBP				
μ	0.613 (0.30, 0.93)	0.483	5.00 (0.87,28.86)	0.151
α_1	0.787 (0.61, 0.97)	0.002	9.00 (1.35,59.79)	0.043
α_2	0.675 (0.44, 0.91)	0.137	3.00 (0.54,16.69)	0.406
DBP				
μ	0.750 (0.55, 0.95)	0.013	1.80 (0.31,10.39)	0.843
α_1	0.788 (0.62, 0.96)	0.001	21.00 (2.05,215.19)	0.009
α_2	0.706 (0.51, 0.90)	0.042	3.00 (0.54,16.69)	0.406
PI				
μ	0.519 (0.27, 0.77)	0.884	1.00 (0.15,6.64)	1.000
α_1	0.525 (0.27, 0.78)	0.845	1.00 (0.15,6.64)	1.000
α_2	0.631 (0.36, 0.91)	0.352	1.80 (0.31,10.39)	0.843
IBI				
μ	0.594 (0.35, 0.84)	0.456	1.80 (0.31,10.39)	0.843
α_1	0.550 (0.30, 0.80)	0.698	1.00 (0.15,6.64)	1.000
α_2	0.644 (0.37, 0.92)	0.31	1.80 (0.31,10.39)	0.843

95% CI reported for AUC [12] and odds ratios
Odds ratios reported at a specificity of 0.75

TABLE III
MULTIVARIABLE LOGISTIC REGRESSION MODELS

Model	AUC	p_{ROC}	Odds Ratio	p_{OR}
μ_{MAP}, μ_{DBP}	0.800 (0.61, 0.99)	0.377	9.00 (1.35,59.79)	0.043
$\alpha_{1,MAP}$, μ_{DBP}	0.806 (0.64, 0.97)	0.062	5.00 (0.87,28.86)	0.151
$\alpha_{1,MAP}$, $\alpha_{1,SBP}$	0.812 (0.65, 0.97)	0.298	21.00 (2.05,215.19)	0.009
$\alpha_{1,DBP}$, μ_{DBP}	0.813 (0.65, 0.97)	0.059	5.00 (0.87,28.86)	0.151
$\alpha_{2,PI}$, μ_{DBP}	0.825 (0.67, 0.98)	0.121	5.00 (0.87,28.86)	0.151
$\alpha_{1,SBP}$, μ_{DBP}	0.831 (0.66, 1.00)	0.003	5.00 (0.87,28.86)	0.151

95% CIs reported for AUC [12] and odds ratios
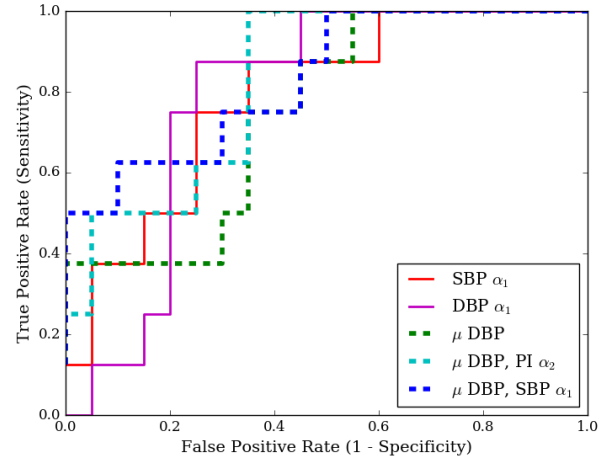Odds ratios reported at a specificity of 0.75

Fig. 3. Receiver-Operator Characteristic Curves

Fig. 2. Histograms of IBI α_2 for IVH and non-IVH groups over all analysed windows

B. Multivariable Predictors of IVH

The possible improvements to IVH prediction using multivariable logistic regression models are shown in Table III, while the receiver-operator characteristic (ROC) curves for two models with the highest AUC are displayed in Fig. 3. The p_{ROC} values reported are those from the Delong method comparison with one of the univariate linear feature models (mean MAP, AUC = 0.638) [12]. Though improvements in the AUC were achieved, the only model to perform statistically significantly better than this reference was the combined $SBP\alpha_1$ and mean DBP model ($p < 0.01$). It is necessary to note however, that these were not statistically significant when compared with the best performing model (DBP α_1, Table II) which is also shown in Fig. 3.

IV. DISCUSSION

The objective of this study centred on evaluating features for their capacity to discern infants who later developed IVH from those who did not. Of the univariate models, the highest AUC achieved was 0.788 (0.62, 0.96) and 0.787 (0.61, 0.97) for the short-term scaling exponents of DBP and SBP, respectively. The model that yielded the overall highest AUC of 0.831 (0.66, 1.00) was the combination of mean DBP and SBP α_1. This was also the sole model of those evaluated to exhibit statistically significant improvement on the reference model (mean MAP, AUC = 0.638).

The comparison of AUC improvements, though not statistically significant compared to the best performing univariate model (DBP α_1, AUC = 0.788), would suggest a potential to improve classification performance with additional variables.

The odds ratios however, need to be taken into consideration for model evaluation. In the ROC sensitivity range of 0.75-0.9, the univariate models exhibit equal if not higher specificity to the multivariable models. The improvement in AUC for the latter models is sourced from the increased sensitivity in the higher range of specificities (0.9-1.0). This would suggest that, depending on the intended application of the model, whether to exclude infants at low-risk of developing IVH, or to identify those at a high-risk, different markers may be used.

Despite the relevance of respiration mechanics in the development of IVH, this study found that individually, the respiratory features explored (μ_{IBI}, IBI α_1 and IBI α_2) were not effective predictors of IVH. This is particularly evident in the significant overlap Fig. 2 between both groups. All of the subjects had synchrony between ventilator inflations and patient breathing with a sensitive flow trigger system captured with a hot wire pneumotach integral to the mechanical ventilator. It is also possible these three features were not suited to capturing significant, adverse patient-ventilator interaction leading to IVH that was alluded to in earlier studies [8][9]. An avenue of further exploration may involve cross-correlation or means other than linear combination of these features for multivariable evaluation.

The predictive capacity of the short-term scaling exponents (α_1) for beat-to-beat MAP, SBP and DBP are in alignment with previously reported findings [5], and their high correlation would suggest that general fluctuations in blood pressure can be used as markers for identifying infants at high-risk of developing IVH. These fluctuations may be influenced by factors including high ventilator pressure, possibly reducing venous return to the pulmonary circulation, and patency of the ductus arteriosus [13]. Recent studies have also used blood-pressure derived features as proxies for cerebral perfusion, where features explored include the diastolic closing margin [14], and significant deviations from a subject-specific optimal value for MAP [15]. These findings are also in alignment with the altered vagal nerve activity reported in infants with IVH [16], where studies have observed that vagal blockade by atropine could lead to increased short-term scaling exponents of blood pressure, although the exact mechanism of influence is not clear. Of all models evaluated, the highest AUC achieved was 0.831 (0.66, 1.00). This model was fitted using mean DBP and SBP α_1, which could be a reflection of a widened pulse pressure as seen in symptomatic patent ductus arteriosus.

The short-term scaling exponents of the beat-to-beat fluctuations in blood pressure were shown to be relatively robust markers in identifying IVH. The considerable overlap in the histograms in Figs. 1 and 2 would caution the use of a single frame of reference for comparison of the two groups, irrespective of the feature's predictive capacity. In accounting for all qualifying windows however, a degree of separation is possible, as shown in Table I, comparing the mean of these markers.

Limitations of the study include the size of the dataset ($n = 28$), the representation of IVH of 29% and signal quality. The latter of these was addressed through the quality controls applied to exclude certain segments windows from the analysis.

V. CONCLUSION

Overall, this study found the short-term scaling exponents of beat-to-beat MAP, DBP and SBP, as well as mean DBP to be relatively robust markers in identifying infants at high-risk of developing IVH. The prediction performance improved marginally through linear combination of these features. In a clinical context, such an approach to predictive monitoring could be adopted. That is, a running 10 min window could continuously evaluate the blood pressure short-term scaling exponents for sufficiently high-quality windows and identify high-risk infants using a defined mean threshold.

REFERENCES

[1] Srinivas Bolisetty, Anjali Dhawan, Mohamed Abdel-Latif, Barbara Bajuk, Jacqueline Stack, Kei Lui, et al. Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants. *Pediatrics*, pages peds–2013, 2013.

[2] Eva Landmann, Björn Misselwitz, Jens O Steiss, and Ludwig Gortner. Mortality and morbidity of neonates born at¡ 26 weeks of gestation (1998–2003). a population-based study. *Journal of perinatal medicine*, 36(2):168–174, 2008.

[3] Conny MA van Ravenswaaij-Arts, Jeroen CW Hopman, Louis AA Kollée, Joop PL van Amen, Gerard BA Stoelinga, and Herman P van Geijn. The influence of respiratory distress syndrome on heart rate variability in very preterm infants. *Early human development*, 27(3):207–221, 1991.

[4] BD Hanna, MN Nelson, RC White-Traut, JM Silvestri, U Vasan, P Meleedy Rey, MK Patel, and E Comiskey. Heart rate variability in preterm brain-injured and very-low-birth-weight infants. *Neonatology*, 77(3):147–155, 2000.

[5] Ying Zhang, Gregory SH Chan, Mark B Tracy, Murray Hinder, Andrey V Savkin, and Nigel H Lovell. Detrended fluctuation analysis of blood pressure in preterm infants with intraventricular hemorrhage. *Medical & biological engineering & computing*, 51(9):1051–1057, 2013.

[6] Volkan Tuzcu, Selman Nas, Umit Ulusar, Ahmet Ugur, and Jeffrey R Kaiser. Altered heart rhythm dynamics in very low birth weight infants with impending intraventricular hemorrhage. *Pediatrics*, 123(3):810–815, 2009.

[7] Heather O'Leary, Matthew C Gregas, Catherine Limperopoulos, Irina Zaretskaya, Haim Bassan, Janet S Soul, Donald N Di Salvo, and Adré J du Plessis. Elevated cerebral pressure passivity is associated with prematurity-related intracranial hemorrhage. *Pediatrics*, 124(1):302–309, 2009.

[8] Jeffrey M Perlman, Steven Goodman, Katherine L Kreusser, and Joseph J Volpe. Reduction in intraventricular hemorrhage by elimination of fluctuating cerebral blood-flow velocity in preterm infants with respiratory distress syndrome. *New England Journal of Medicine*, 312(21):1353–1357, 1985.

[9] Jeffrey M Perlman, Joseph B McMenamin, and Joseph J Volpe. Fluctuating cerebral blood-flow velocity in respiratory-distress syndrome: relation to the development of intraventricular hemorrhage. *New England Journal of Medicine*, 309(4):204–209, 1983.

[10] C-K Peng, Shlomo Havlin, H Eugene Stanley, and Ary L Goldberger. Quantification of scaling exponents and crossover phenomena in non-stationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):82–87, 1995.

[11] C Thamrin and G Stern. New methods: what do they tell us? fluctuation analysis of lung function. *Eur Respir Mon*, 47:310–324, 2010.

[12] Elizabeth R DeLong, David M DeLong, and Daniel L Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845, 1988.

[13] Praveen Ballabh. Intraventricular hemorrhage in premature infants: mechanism of disease. *Pediatric research*, 67(1):1–8, 2010.

[14] Christopher J Rhee, Jeffrey R Kaiser, Danielle R Rios, Kathleen K Kibler, R Blaine Easley, Dean B Andropoulos, Marek Czosnyka, Peter Smielewski, Georgios V Varsos, Craig G Rusin, et al. Elevated diastolic closing margin is associated with intraventricular hemorrhage in premature infants. *The Journal of pediatrics*, 2016.

[15] Cristine Sortica da Costa, Marek Czosnyka, Peter Smielewski, Subhabrata Mitra, Gordon N Stevenson, and Topun Austin. Monitoring of cerebrovascular reactivity for determination of optimal blood pressure in preterm infants. *The Journal of pediatrics*, 167(1):86–91, 2015.

[16] Naoki Koketsu, Michael A Moskowitz, Hermes A Kontos, Masayuki Yokota, and Takeo Shimizu. Chronic parasympathetic sectioning decreases regional cerebral blood flow during hemorrhagic hypotension and increases infarct size after middle cerebral artery occlusion in spontaneously hypertensive rats. *Journal of Cerebral Blood Flow & Metabolism*, 12(4):613–620, 1992.

Appendix II

Reducing false arrhythmia alarms in the ICU using multimodal signals and robust QRS detection

I contributed to the data analysis and algorithm development for false arrhythmia alarm detection as well as revising the corresponding sections of the research paper.

Reducing false arrhythmia alarms in the ICU using multimodal signals and robust QRS detection

Nadi Sadr[1]**, Jacqueline Huvanandana**[1]**, Doan Trang Nguyen**[1]**, Chandan Kalra**[1]**, Alistair McEwan**[1] **and Philip de Chazal**[1,2]

[1] School of Electrical and Information Engineering, University of Sydney, NSW 2006, Sydney, Australia
[2] Charles Perkins Centre, University of Sydney, NSW 2006, Sydney, Australia

E-mail: nadi.sadr@sydney.edu.au

CrossMark

Abstract

This study developed algorithms to decrease the arrhythmia false alarms in the ICU by processing multimodal signals of photoplethysmography (PPG), arterial blood pressure (ABP), and two ECG signals. The goal was to detect the five critical arrhythmias comprising asystole (ASY), extreme bradycardia (EBR), extreme tachycardia (ETC), ventricular tachycardia (VTA), and ventricular flutter or fibrillation (VFB). The different characteristics of the arrhythmias suggested the application of individual signal processing for each alarm and the combination of the algorithms to enhance false alarm detection. Thus, different features and signal processing techniques were used for each arrhythmia type. The ECG signals were first processed to reduce the signal interference. Then, a Hilbert-transform based QRS detector algorithm was utilized to identify the QRS complexes, which were then processed to determine the instantaneous heart rate. The pulsatile signals (PPG and ABP) were processed to discover the pulse onset of beats which were then employed to measure the heart rate. The signal quality index (SQI) of the signals was implemented to verify the integrity of the heart rate information. The overall score obtained by our algorithms in the 2015 Computing in Cardiology Challenge was a score of 74.03% for retrospective and 69.92% for real-time analysis.

Keywords: intensive care unit, photoplethysmography, arterial blood pressure, electrocardiogram, asystole, arrhythmia false alarm, signal quality index

(Some figures may appear in colour only in the online journal)

1. Introduction

An intensive care unit (ICU) involves a large number of medical devices, background noise and alerting signals of the devices with a number of attending medical staff (Donchin and Seagull 2002). An ICU aims to monitor the biological signals of patients in critical conditions. The monitoring systems often incorporate alarms to attract staff attention (Donchin and Seagull 2002). The alarms from the physiological monitors can be classified into technically correct or false groups (Lawless 1994). False alarms are alerting signals of monitoring equipment with no associated clinical cause (Chambrin *et al* 1999) or relation with life-threatening conditions of the patient (Donchin and Seagull 2002). Studies show that over 85% of the ICU alarms are false (Lawless 1994, Chambrin 2001, Sendelbach and Funk 2012) and that they have a number of negative effects. The extra noise generated by the false alarms negatively impacts the patient's sleep and increases the stressors in ICU which can reduce the recovery rate (Novaes *et al* 1997). The unwanted alarms can also lead to vital monitoring equipment being switched off (Lawless 1994). Also, the medical staff lose sensitivity to frequent false alarms which in turn increases the likelihood of missing true alarms (Clifford *et al* 2006). Alarm fatigue is a destructive outcome of the large number of false alarms in ICU in which the clinical staff ignore the alerting signals or change the settings to a level of deactivation which has been identified as a critical health and safety problem (Sendelbach and Funk 2012).

The disruptive consequences of false alarms can be alleviated in two ways. The first solution is automatic detection of false alarms and the source of the alarm. Secondly, medical staff can help solve the issue by persistent observation of the systems and signals and diagnosing the false alarms (Imhoff and Kuhls 2006).

In this paper, we developed a signal processing system for automatic detection of the following arrhythmias: ventricular tachycardia (VTA), ventricular fibrillation or flutter (VFB), extreme tachycardia (ETC), extreme bradycardia (EBR) and asystole (ASY) by processing one or more of the following signals: the photoplethysmography (PPG), the arterial blood pressure (ABP), and two electrocardiogram (ECG) signals. To train and test our system, we have used the signals of the *2015 Computing in Cardiology Challenge Dataset*. The detections are then used to assess the validity of alarms generated by ICU equipment with goal of identifying false alarms. The signal processing algorithms we describe here are the basis of our entry in the *PhysioNet/Computing in Cardiology Challenge 2015*. An early description of these algorithms were published in *Computing in Cardiology Challenge 2015* (Sadr *et al* 2015).

2. Input data

The input dataset was provided by the *PhysioNet/Computing in Cardiology Challenge 2015* (Clifford *et al* 2015). The dataset incorporates 1250 arrhythmia alarms which were selected randomly from four hospitals in US and Europe. Less than three alarms from the five arrhythmia types were selected from an individual patient and they are often more than five minutes apart. The dataset was divided into 750 open-access recordings used as a learning set and 500 recordings for test set which were hidden. Train and test set were comprised of signals recorded from different patients. Five hundred and ninety recordings in the training data included four signals comprising of two ECG signals and two pulsatile signals (the photoplethysmogram (PPG) and arterial blood pressure (ABP)). While the first ECG signal was mostly lead II and the second ECG signal was mostly lead aVr, there were a number of recordings where different leads were recorded. One hundred and sixty recordings in the training data
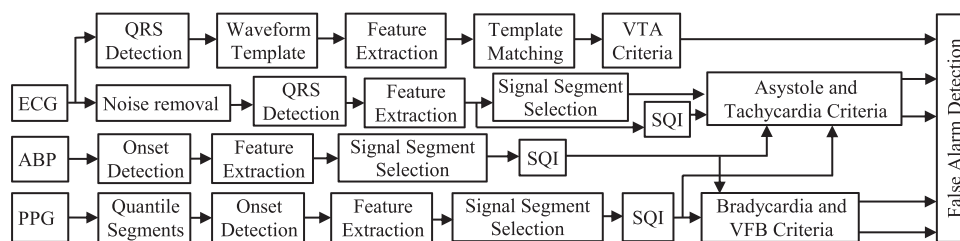
1341

134

Figure 1. Block diagram of the proposed system for arrhythmia false alarm detection in the ICU. Abbreviations: arterial blood pressure (ABP), electrocardiogram (ECG), signal quality index (SQI), ventricular flutter or fibrillation (VFB), ventricular tachycardia (VTA).

comprised of three of the four signals listed above (i.e. two ECG and one pulsatile or one ECG and two pulsatile signals). Each patient had a maximum of two recordings of separate alarms in the dataset. The chosen recordings had been annotated by three or more experts and the alarm outcome was determined by agreement of at least two-thirds of the experts. Recordings that did not have a two-thirds agreement were excluded. Each alarm was annotated as 'true', 'false', or 'impossible to tell'. Each record includes an alarm at the fifth minute from the start of the record and the corresponding arrhythmia event happens within ten seconds of the alarm. If further arrhythmias occurred before the fifth minute of a record, they were not annotated. The repeated alarms and information from alarms prior to the annotated one are not employed to reduce the probability of transferring errors from one alarm to the next one. Resampling was applied to the four sensor signals at the rate of 250 Hz, 16 bit. Band pass filtering in the range of 0.05–40 Hz was implemented with an FIR filter. Also, common notch filters were utilized for noise removal of powerline noise. Pacemaker and other noise artefacts still existed in the ECG signals. In some cases, movement artefacts, failure in sensor connection, line flush, coagulation and other interferences also influenced the pulsatile signals.

3. Signal analysis

The block diagram of the proposed system for false arrhythmia alarm detection is shown in figure 1. A high level description of system is provided here and more detail is given in sections 3.1–3.3. The top-down order of the signal blocks represents the priority for signal selection in the analysis process.

VTA detection relied solely on features extracted from the ECG signals which were processed without noise removal and using template matching. QRS detection was applied to the ECG for identification of QRS complexes. A reference waveform template was generated from the first QRS and subsequent QRSs were compared to the reference to determine if they were irregular beats. If five or more beats were deemed irregular in the alarm segment then the VTA alarm confirmation was set to true, otherwise it was set to false.

The alarm confirmation procedure of asystole and tachycardia were similar. First, signal interference was removed from the ECG signals. QRS beats were detected from the clean ECG signals and discriminating features were extracted. A segment of the ECG signal containing the alarm was identified and SQI measurements determined. A similar process was applied to the ABP and PPG signals resulting in SQI and feature values. Finally, for the alarm segment, the features of the ECG signals, the ECG SQI measures, the features and SQI measures of the available pulsatile signals served as inputs to assess the validity of the tachycardia

alarms. Processing for the asystole alarm evaluation was similar except that we did not use the SQI measures of the ECG signals.

The process of detecting bradycardia and VFB false alarms were identical. The pulsatile signals were employed to diagnose these two arrhythmias. After distinguishing the onset beats of ABP signals, features were measured and the segment of the alarm was identified for criteria assessment and SQI evaluation. The available pulsatile signals, ABP and PPG, were processed similarly, with the exception of the PPG processing which included a quantile segmentation step prior to onset detection. This was followed by feature extraction and identifying the alarm segment and the features in that segment. Finally, the SQI of the alarm segment was measured and the features from available ABP and PPG with their SQI measures were used to confirm the bradycardia or VFB alarm status.

In the following sections, we first describe the signal processing methods for interference removal, heartbeat identification, SQI measurement and feature extraction for the ECG, PPG and ABP signals. We then describe our hierarchical processing of the ECG and pulsatile signals to determine the final alarm status.

3.1. ECG signals

In order to diagnose the high risk arrhythmias, the ECG signals are functional and informative. The recordings in the Challenge dataset are comprised of lead II and/or lead aVr and/or other leads. The block diagram of the proposed system to detect false arrhythmia alarms in the ICU containing the ECG signal processing algorithm is shown in figure 1. The ECG signals are mostly corrupted by movement artefact, pacemaker, and fibrillator signals. A first step was to detect and remove these artefacts. Filtration, described below, was applied to the raw ECG signals to remove the unwanted interference. The filtered signals were then processed to find the QRS complexes. After calculation of RR-intervals from the QRS detection points and applying the above signal processing steps, there were still QRS complexes in some recordings that were not detected successfully. To attempt to recover these missed signal beats, we detected heart beats in the other sensor channels and then used the beat detections across all channels to obtain an enhanced recognition. The final step was feature extraction for arrhythmia detection.

3.1.1. Interference removal. The ECG signals of the challenge database were distorted by motion artefact, powerline interference, baseline drift, displacement of sensors and instrumentation noise produced by pacemakers. Baseline drifts lead to deformation of the ST segment which plays an important role in arrhythmia detection, results in failure in false alarm recognition. Thus, elimination of baseline wander is an essential part of interference removal for detecting arrhythmias and false alarms.

Interference removal was performed by applying filters for noise reduction. The ECG signals of the database were distorted by baseline wander noise which originated from movement, respiration and perspiration affecting the electrode impedance (Tinati and Mozaffary 2006). Baseline wander noise affects the low frequency component of the ECG signals (Jain and Shakya 2014) and can influence the clinical interpretation of ECG signal. In this study, baseline wander noise was removed by two median filters (de Chazal *et al* 2003). The first median filter with 200 ms width is applied to remove the QRS complexes and P waves. Then, the resulting PQRS-free signal is used to apply the second median filter. The width of the second median filter was 600 ms to eliminate T waves. Thus, the output of the second median filter did not include the information from the ECG waves and contained only the baseline wander. By subtracting the output of the second median filter from the raw input ECG signals, the resulting signal contained the P-QRS-T complexes minus baseline wander. This method
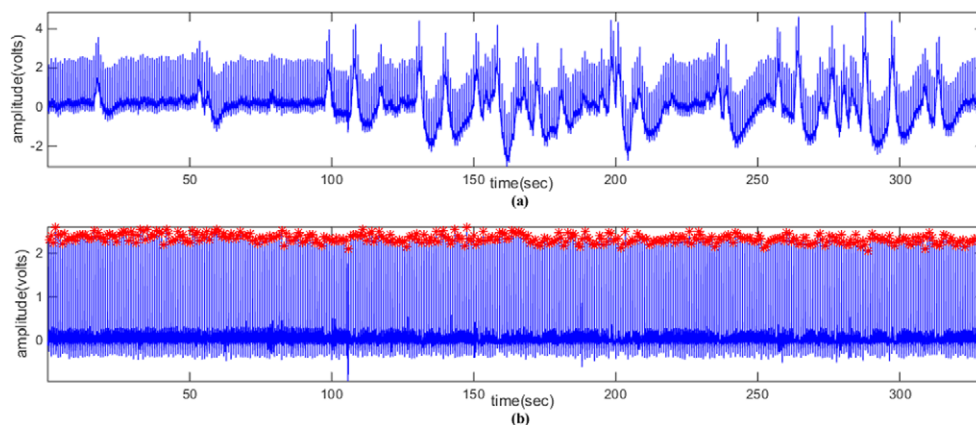
Figure 2. Sample result of applying interference removal to ECG signal (a64) and noise removal. (a) Raw ECG lead II with Asystole as a false alarm. (b) The result of applying interference removal on the input ECG. Stars are the R peaks detected by Hilbert QRS algorithm.
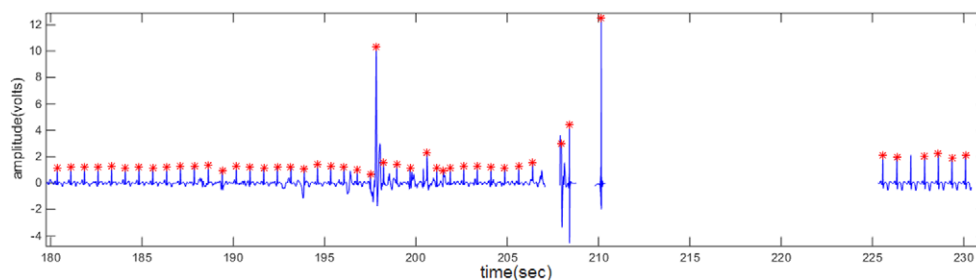

Figure 3. Result of the application of Hilbert QRS detector on ECG II signal after interference removal.

was applied to both available ECG signals of each recording. The result of implementing this algorithm on an ECG signal of the Challenge training set is shown in figure 2, where baseline wander noise is easily seen in figure 2(a) of the ECG recording (a64). The resulting signal after this interference removal is shown in figure 2(b) which reveals that interferences including the baseline wander were appropriately eliminated. Subsequent to interference removal and denoising the ECG signals, QRS detection was applied to the signals. Reliable identification of QRS complexes is difficult due to the changing nature of their morphology and the influence of unwanted interference on the ECG signal (Thakor *et al* 1984). By removing the unwanted interference from the ECG signals, we can improve the likelihood of successful arrhythmia detection. The other important factor is the changing morphology of the QRS complexes which were taken into account in the utilized QRS detection algorithm explained in the following section.

3.1.2. RR interval and signal segment selection. The first step toward feature extraction for arrhythmia recognition by ECG signal is QRS detection. There are various QRS detection algorithms but selecting a reliable method is highly significant for false arrhythmia alarm recognition. In this work, the QRS complexes were identified by a Hilbert transform based

1344

137

algorithm (Benitez *et al* 2001). The result of applying Hilbert QRS detector on an ECG signal of the Challenge training set which contains missing values is shown in figure 3. The Hilbert QRS detector employed in this study (Shouldice *et al* 2004, Redmond and Heneghan 2006) was tested on various ECG signals and leads. It was reported to reliably detect QRS complexes of all common leads with a satisfactory noise tolerance (Shouldice *et al* 2004). The algorithm used in this paper was previously tested on the MIT-BIH Arrhythmia database and obtained 98.4% positive predictive accuracy and sensitivity of 98.5% (Hickey *et al* 2004).

Finally, the QRS detections were used for feature extraction. The RR-intervals were determined by calculating the time difference between two adjacent QRS detections. Then, the information closer to the alarm is utilized for false alarm identification. Thus, the segment comprising the alarm is selected for arrhythmia recognition. In this study, the alarm segment begins 16 s prior to the alarm and is ended by the alarm which occurs at the fifth minute of the signal.

3.1.3. ECG SQI. Visual observation of QRS detection points and corresponding RR intervals revealed that some of the heart beats were missed or falsely detected. Missing value intervals and noisy alarm segments can produce issues in signal processing and suspect QRS detection points. Also, ECG artefact was reported as a reason for false arrhythmia alarms (Aboukhalil *et al* 2008). Therefore, before further processing, the quality of the ECG signal was assessed.

In order to assess the quality of the ECG signal, signal quality index (SQI) was exploited to determine if it possessed reliable information for false alarm detection. The signal evaluation index has been widely studied (Silva *et al* 2011, Clifford and Moody 2012).

In this paper, four tests were applied to determine the ECG SQI. If the alarm segment satisfied the tests, it was allowed to proceed for further processing. The first test determined if the segment was empty. No heart beat in the segment indicated a failure of the heart beat identification algorithms and was indicative of the presence of significant signal interference. In the second test, the number of the detected QRS detection complexes or the available beats of each ECG signal was measured in the segment. This test allowed recognition of the signals with a high proportion of missing heart beats and the inspection of the proportion of motion artefact, failure in sensor attachment and other noises in the segment. The minimum number of beats was set to ten beats. If an ECG signal segment contained less than this minimum, it was not considered further in arrhythmia detection. The third test was the maximum RR-interval or minimum heart rate. This test examines the physiological reliability of the heart rate and indicates the noisy alarm segments and missed QRS complexes. The maximum measure of the third test was set to six seconds. If all RR-interval in a segment were less than six seconds, the test was passed. The fourth ECG SQI test was the standard deviation of the ECG in the segment containing the alarm. This test helps identify the segments with a high percentage of noise and artefact. The optimum standard deviation was adopted as 0.05 over the whole segment. These tests address most of the observed corruptions on the ECG signals comprising the level of noise and the percentage of missed or spurious QRS detections in the segment. The output of the ECG SQI algorithm determines whether the ECG signal is satisfactory for next processing.

It should also be noted that not all steps of the ECG SQI were evaluated for all of the arrhythmia detections. Studies identified that SQI evaluation diminished the accuracy of arrhythmia diagnosis due to their noisy manifestation (Behar *et al* 2013). Since the behaviour of some of the arrhythmias such as VTA is homogenous to noise structure, ECG SQI reduced the performance and was removed from the false alarm detection of those arrhythmias. Further details will be described in section 3.3.

3.2. Pulsatile signals

It is reported that implementation of pulsatile signals which contain cardiac cycles improves the diagnosis of false arrhythmia alarms when combined with cardiac cycle information from ECG signals (Aboukhalil *et al* 2008). In the *PhysioNet/CinC Challenge 2015* dataset, one or both arterial blood pressure (ABP) and photoplethysmogram (PPG) signals were available and we utilized them to reduce the false alarm rate. The signal processing algorithms of the sample submission provided by the *PhysioNet/CinC Challenge 2015* were utilized for determining the beat onset points from the pulsatile signals and arrhythmia identification in this study. We provide a short description of these algorithms in the next sections.

3.2.1. ABP signal.

Arterial blood pressure (ABP) was another signal used to verify the false arrhythmia alarms. As it is recorded separately to the ECG leads, it rarely contains identical interference to the ECG signal (Aboukhalil *et al* 2008). Also, ABP is regarded as the pressure signal with the least noise and artefact (Clifford *et al* 2006).

There were three PhysioNet open-source algorithms employed to process the ABP. In order to find the onset of the ABP pulses, the 'wabp' algorithm was executed on ABP signal (Goldberger *et al* 2000). The Length transform is exploited in this technique (Zong *et al* 2003) and noise removal and feature enhancement was applied through the algorithm. Following this, ABP features were calculated with the 'abpfeature' algorithm. The features include systolic pressure, diastolic pressure, systolic area, and mean pressure on the onset beats of the ABP pulses. Next, ABP quality index (SQI) was estimated by the 'jSQI' algorithm at each ABP detected beats (Sun *et al* 2004). The ABP SQI algorithm explores if the features are physiologically plausible. The features that were not clinically reliable were eliminated. Lastly, the time between the pulse onsets in the ABP signal was measured to generate pulse intervals of ABP signal to be used for further signal processing.

3.2.2. PPG signal.

The other pulsatile signal used for false alarm detection was the photoplethysmogram (PPG) which was available in many of the recordings in the learning set of PhysioNet/CinC Challenge 2015. The open-source PhysioNet algorithms were employed to process the PPG signal. Firstly, the signal was divided into three partitions by the open-source 'quantile' algorithm. The three quantiles used were 5%, 50% and 95%. Next, the subtraction of third quantile and first quantile was measured and employed to detect pulse onsets. The onset beats of the PPG waveform was verified with 'wabp' algorithm (Goldberger *et al* 2000). Then, the pulse intervals were measured by the difference of the adjacent onset beats and used to calculate the heart rate. Finally, the PPG signal quality was evaluated with 'ppgSQI' algorithm through a beat template correlation technique.

3.3. Alarm detection

The segment containing the alarm from the available signals of each recording was selected from the 16 s prior to the alarm ending. The heart rates and intervals corresponding to the alarm segment were used for further signal processing. This study aimed to recognize the false arrhythmia alarms in real-time and avoid using the data following the alarm occurrence. Our proposed algorithm can also be implemented in a retrospective manner which uses the information after the alarm. In the *PhysioNet/CinC Challenge 2015* dataset, the alarms were set to appear at five minutes after the beginning of the signal. To guarantee alarm inclusion in the segment, the segment started 16 s before the alarm time. Next, the beats in the alarm segment were identified for the available signals of each recording and the corresponding RR-intervals,

pulse intervals and heart beats were chosen. Finally, the features were calculated from the heart rates, RR-intervals or pulse intervals of the pulsatile waveforms. It should be noted that signal processing for each arrhythmia alarm condition were separately executed with different models and features (Behar *et al* 2013).

3.3.1. Multimodal signal. Interference in the ECG signal could be a source of the false arrhythmia alarms. Using other leads of the ECG signal simultaneously with other signals to combine the information could improve the diagnosis (Aboukhalil *et al* 2008). Thus, exploiting information from multimodal signals enhances the arrhythmia detection. It was reported that false arrhythmia alarms were better recognized by multimodal signal fusion which was widely discussed in the PhysioNet/Computing in Cardiology Challenge 2014 (Moody *et al* 2014). Various algorithms studied robust detection of the heart beats for multimodal recordings and signal fusion purposes (Silva *et al* 2015).

The highest score of the challenge was achieved by a SQI based method (Johnson *et al* 2015). They noted that the onset of the ABP pulses appear with a delay after the heart pumps blood out of left ventricle. The delay between detected onset beats of the blood pressure waveform and R peaks of the ECG signals were collected and the R peaks were matched according to the delay (Silva *et al* 2015). Then, the SQI measures of the signals were used to identify the high quality signal (ECG or ABP) which was then used for heart beat detection (Johnson *et al* 2015). A major focus was on employing different peak detectors for blood pressure and ECG signal to compare and evaluate their detection outcome.

In our study, the Hilbert QRS detector identified the QRS complexes with a decent accuracy (Hickey *et al* 2004, Shouldice *et al* 2004). So the QRS detections of the ECG signals were reliable measures with which to proceed the signal processing and advance to the pulse onsets of pulsatile signals provided by 'wabp' algorithm. Also the accuracy of the Hilbert QRS detector implied that application and comparison of other peak detectors for ECG signals is not essential.

Considering the available signals in the majority of the training set, the signals were prioritized from ECG signals to the pulsatile signals for feature extraction. However, unwanted interferences can corrupt the signal properties which are significant measures for arrhythmia detection. As previously mentioned, the noisy signals are known as a major source of false arrhythmia alarms (Aboukhalil *et al* 2008). Thus, the SQI of the signals were evaluated to identify noisy signals. If the signal passed the SQI tests, then features of the signal were extracted. The signal SQI measures were processed in the following order of first available ECG, second ECG, ABP and PPG. The SQI of the ECG signals was evaluated as explained in section 3.1.3. For all alarms with the exception of bradycardia and VFB, the result was decided based on the highest priority signal with the best quality of those considered and the alarm suppression or trigger was determined by that signal. For bradycardia and VFB, all considered signals satisfying the quality criteria were used in the decision-making process. The algorithm has been shown in the block diagram of the system in figure 1.

Combining the features and information of multimodal signals and evaluating the signal quality addressed the intervals with missing values or noises such as failure in sensor attachment and motion artefact as an observed issue in the challenge training set. This approach with multimodal signals benefits false arrhythmia alarm assessment in dealing with signals recorded in a real environment. This paper analyzed each arrhythmia through a different approach which will be explained in the following section. The contribution of each signal of training data in arrhythmia alarm detection is shown in table 1. It was identified after implementing and running the algorithm with the training set. It could be seen that multimodal signals benefit the arrhythmia recognition differently with various usage distributions. For

Table 1. The use of each signal in the decision criteria for each arrhythmia in the training set.

Signals	Asystole (%)	Bradycardia	Tachycardia	VFB	VTA
First ECG	75.4	Not used	100%	Not used	88.6%
Second ECG	1.6	Not used	Not used	Not used	3.2%
ABP	3.4	10.1%	Not used	22.4%	Not used
PPG	13.9	13.5%	Not used	8.6%	Not used
Not suppressed[a]	5.7	76.4%	0%	69.0%	8.2%

[a] Algorithm was not able to suppress the alarm using any of the four input signals.

instance, the table demonstrates that 75.4% of the asystole alarms were detected by the first ECG signals and 13.9% of the asystole alarms were detected by PPG signals. Only 5.7% of the asystole alarms could not be suppressed using any of the four input signals.

3.3.2. Asystole detection.   Asystole (ASY) was defined as an absence of a heart beat for at least four seconds (Clifford *et al* 2015). Thus, the minimum threshold for asystole detection was set to four seconds with a tolerance of 0.5 s. The priorities of signals used for asystole detection were defined as the first ECG, followed by the second ECG followed by the available pulsatile signals. The criteria for using the ECG signals encompassed successful beat detection, a maximum RR-interval of less than the defined threshold for signal quality and a standard deviation (SD) within the defined threshold. This was less restrictive for use of the lower priority pulsatile signals where an availability of detected beats was sufficient. As is shown in the block diagram of the system in figure 1, the ECG SQI measures were not employed in asystole detection. This algorithm design decision was made as we found that implementing SQI measures tended to knock out heart beats, which increased the likelihood of false asystole alarm detections. The first signal to satisfy the aforementioned selection criteria was then used to determine the alarm result. The feature used for asystole was the maximum RR-interval of the segment which, if above the specified threshold and tolerance, triggered an alarm and otherwise, suppressed it. An alarm on the selected sensor resulted in the final decision being set to true alarm.

3.3.3. Extreme bradycardia detection.   Extreme bradycardia (EBR) was defined as five continuous beat intervals greater than 1.5 s (Clifford *et al* 2015). We detected extreme bradycardia by processing the estimated minimum heart beats and identifying five or more consecutive beats with intervals exceeding 1.5 s. The pulsatile signals were exploited for EBR alarm recognition and the minimum heart rate of the available ABP and PPG signals were measured. Five or more consecutive beats with intervals exceeding 1.5 s were identified. The average heart rate of the beats were calculated and the minimum of these average heart rates in the alarm segment was recognized as the features called 'Low HR'. Firstly, the SQI of the pulsatile signal was assessed. The SQI threshold for pulsatile signals set to 0.9. If the SQI of the signal satisfied the threshold, then the alarm segment was checked if it contained beats meeting the above criteria. If the feature was over the threshold with the tolerance for either pulsatile signals of ABP or PPG, the alarm was set to true. Otherwise, the alarm was assigned to false.

3.3.4. Extreme tachycardia detection.   Extreme tachycardia (ETC) was defined as a heart rate elevation of more than 140 beats per minute for 17 consecutive beats (Clifford *et al* 2015). The algorithm begins with processing the first ECG signal, followed by the second ECG and the pulsatile signals, in the same order as that for asystole detection. The last feature of the ECG SQI which was the standard deviation of the segment was omitted in the SQI evaluation for tachycardia alarm detection. Instead, a minimum number of beats defined for tachycardia served

141

as an additional criterion for signal selection. This threshold was set to 11, which was chosen based on an iterative process, varying the threshold and adjusting according to the tachycardia detection performance on the training set. If ECG SQI reported a high quality signal and there was a sufficient number of beats detected in the segment, the signal was utilized for the next phase of tachycardia detection. Tachycardia minimum threshold was set to 110 bpm with a tolerance of 10 bpm. For the ABP signal, the SQI was compared to the threshold of 0.9 and the number of identified beats in the alarm segment was compared to the minimum acceptable beats for tachycardia detection. The heart rates from the selected signal were then able to trigger an alarm in two ways; the first was if the number of beats exceeded 30 for the acquired segment or secondly, if the number of beats above the tachycardia threshold and tolerance exceeded the minimum acceptable beats. If neither of these criteria were fulfilled, the alarm was suppressed.

3.3.5. Ventricular tachycardia detection. Ventricular tachycardia (VTA) was defined as five or more ventricular beats with heart rate higher than 100 beats per minute (Clifford *et al* 2015). Diagnosis of VTA was obtained by a template subtraction process using the raw ECG signals only. We did not use the pulsatile signals for VTA. Also, by removing the ECG SQI measures used in the other alarms from the process, the performance of false alarm detection was enhanced as VTA signals generally had poor SQI.

The first QRS complex in the series was taken as the reference template against which the subsequent waveforms were compared. A beat-to-beat sliding window was applied to the alarm segment to detect each QRS complex. The standard deviation (SD) and mean value of each QRS waveform were subsequently calculated and compared with the peak of the waveform. The complexes with peaks that did not lie within 1 SD of the mean were chosen for evaluation. Then, the mean value of each complex as well as the mean of the template waveform was removed. The waveforms with a SD that did not lie within 0.6 of the overall SD of the segment were labeled as 'irregular' waveforms. This feature was called 'filter vector'. If there were four or more irregular waveforms in the alarm segment, that is, the minimum threshold of 5 beats for VTA with a tolerance of 1 beat, the VTA alarm was set to true. Otherwise, it was labeled as a false alarm.

3.3.6. Ventricular flutter or fibrillation detection. Ventricular flutter or fibrillation (VFB) was assumed to be fibrillatory, flutter, or oscillatory waveform for at least 4 s (Clifford *et al* 2015). It is recognized as a difficult condition to detect using ECG signal (Clayton *et al* 1993, Jekova 2000). Different methods were applied in the studies such as threshold crossing intervals (TCIs) (Thakor *et al* 1990), autocorrelation function (ACF) (Chen *et al* 1987), and complexity measure (Zhang *et al* 1999) which the results were compared in studies (Clayton *et al* 1993, Jekova 2000). The comparisons showed the importance of threshold tuning and choosing the appropriate criteria.

To detect VFB in this study, the ABP SQI was evaluated and compared to the threshold of 0.9. If SQI was above threshold, then the maximum heart rate in the alarm segment was compared to the VF threshold. The VFB threshold was set to 250 bpm with a tolerance of 10 bpm. If the maximum heart rate of alarm segment was greater than the VFB threshold with the tolerance, the VFB alarm was set to true. A similar algorithm was repeated for PPG signal. Either of the pulsatile signals satisfying the criteria resulted in an alarm being triggered.

4. Results and discussion

The results of train and test set are shown in tables 2 and 3 respectively. The best alarm detection was achieved for the tachycardia alarm which obtained a score of 96% for train and 99% for test set. The average score of train set was 79%. While for the test set, the real-time score

Table 2. The results of true positive rates, true negative rates, and scores of training set.

	TP	FP	FN	TN	TPR (%)	TNR (%)	Score (%)
Asystole	0.164	0.057	0.016	0.762	91.11	93.04	87.11
Bradycardia	0.517	0.247	0	0.236	100	48.86	75.3
Tachycardia	0.936	0.043	0	0.021	100	32.81	95.7
VFB	0.103	0.19	0	0.707	100	78.82	81.0
VTA	0.246	0.361	0.015	0.378	94.25	51.15	58.87
Average	0.393	0.18	0.006	0.421	98.50	70.05	79.49
Gross	0.383	0.225	0.009	0.383	97.70	62.99	73.94

Table 3. Results of final submission from test set.

	TPR (%)	TNR (%)	Score (%)
Asystole	78	93	82.46
Bradycardia	100	52	71.13
Tachycardia	100	80	99.10
VFB	100	59	65.52
VTA	91	55	58.07
Real-time	95	65	69.92
Retrospective	98	66	74.03

achieved 69.9% and retrospective score reached 74% which was placed among the top ten scores of the *PhysioNet/Computing in Cardiology Challenge 2015*.

While ECG signals assisted in arrhythmia detection, the noisy characteristics of ECG signals were found to trigger false alarms. Hence, multiple steps were taken to minimize the corruption and negative effect of artefact. Interference removal of ECG signal, signal quality index, and utilizing information from multimodal signals was investigated to differentiate the inferences and improve false alarm detection. On the other hand, in some arrhythmias, interference removal, noise reduction and examining the signal quality did not help in false alarm reduction. We did not apply interference removal to ECG for the VTA arrhythmias. This was because the VTA signals exhibited behavior similar to the noise which our noise removal algorithms were designed to remove. For further improvement, a redesign of our noise removal algorithms so they did not knock out the VTA signal could enhance the results. The VTA alarm identification was reported as the most difficult alarm for detection among the entries of the *PhysioNet/Computing in Cardiology Challenge 2015* (Clifford *et al* 2015). An analysis of algorithms of the top scored entries revealed that better VTA and VFB alarm detection was achieved through algorithms that included descriptive statistics and QRS detection by amplitude envelopes using Fourier and Hilbert transform (Plesinger *et al* 2015), statistical analysis and hand-selected transform (Plesinger *et al* 2015), phase wrapping and machine learning (Ansari *et al* 2015) and adaptive frequency tracking and adaptive mathematical morphology approach (Fallet *et al* 2015). The top entries utilized all of the signals comprising the ECG signals for VFB detection. In our approach, ECG signals were not utilized in the VFB and extreme bradycardia detection algorithms, and hence incorporating ECG information may improve the false alarm detection of these arrhythmias.

Table 4. The features that were selected in the evaluation process for each alarm type.

Features[a]	Asystole	Bradycardia	Tachycardia	VFB	VTA
Number of first ECG beats	√		√		√
Max RR of first ECG	√		√		
First ECG SD	√				
Number of second ECG beats	√		√		√
Max RR of second ECG	√		√		
Second ECG SD	√				
Number of ABP beats	√	√	√		
Max pulse intervals of ABP	√				
ABP SQI		√	√	√	
Low HR of ABP		√			
Max HR of ABP				√	
Number of PPG beats	√	√	√		
Max of RR PPG	√				
PPG SQI		√	√	√	
Low HR of PPG		√			
Max HR of PPG				√	
Length of filter vector					√
Number over threshold			√		

[a] The features were measured over the alarm segment and the detected beats and peaks in the segment.

Table 4 shows a list of the features that were incorporated and contributed in each arrhythmia alarm detection. It demonstrates which features were responsible in making the decision of each arrhythmia alarm identification. The last feature of the table, 'Number over threshold', refers to the number of beats above the tachycardia threshold and tolerance which exceeded the minimum acceptable beats. The results of arrhythmia detection and their employed features of table 4 suggest that the drawback of the proposed algorithm is the variety of features used. The more features employed, the better detection is achieved. We obtained our best result with extreme tachycardia alarm detection which the results of this exploration at table 4 showed to have the largest variety of features.

The various characteristics of the arrhythmias led to different implementation processes for individual arrhythmia alarm detection. This meant that some required noise reduction, while others needed the application of raw data with minimal noise removal. SQI evaluation improved the processing performance in some of the arrhythmia detection. Thus, a fixed method of signal quality evaluation is not suitable for analysis of a variety of arrhythmias with different properties. The evaluation techniques should be adapted to each arrhythmia.

The proportion of each arrhythmia alarm detected by each signal of training set is shown in table 1. Since we had the hierarchical or priority-based approach to selecting the signals to use (i.e. ECG was used firstly, then ABP followed by PPG), we did not necessarily use all of the signals to make the final decision. The results are comparable with the obtained scores from train and test set. It can be observed that asystole alarm identification was mainly detected by the first ECG signal and the distributions match the order of selection criteria. For instance, the

ECG signals were firstly investigated for asystole detection and then the algorithm progressed to pulsatile signals for processing. The results from investigating the usage of the signals (see table 1) validated our assumption of giving the highest priority to ECG signals for asystole detection. It revealed that only 5.7% of the asystole alarms were not suppressed. In contrast, 76.4% of the bradycardia alarms and 69.0% of the VFB alarms were not suppressed by input signals. As we did not utilise ECG signals for detection of either of these alarms, the results suggest that utilising the ECG signals may improve suppression of these false alarms. The VTA alarm detection algorithm relied heavily on the first ECG signal for heart beat detection. A small number of cases (3.2%) were detected with second ECG signal and 8.2% were not successfully suppressed.

We found that varying the threshold setting significantly affected false alarm detection. Thus, implementation of parameter optimization methods such as SVM (support vector machine) as a threshold tuning and model selection algorithm could enhance the scores.

As a final comment, we describe one signal processing step we trialed and abandoned as it did not result in improvement in the scores of either the train or test set. The signal processing step attempted to boost the heart rate identification by multimodal signal integration. The algorithm examined the detected beats of the ECG signals. In case of low quality ECG signals or missing beats, it switched to pulsatile signals. In order to match the R peaks of ECG with pulse onsets of pulsatile signals, the delay between R peaks of ECG signals and pulse onset of pulsatile signals was measured. We adapted a fusion method proposed in the *PhysioNet/ Computing in Cardiology Challenge 2014* by Johnson *et al* (2015). The peaks of the available ECG signals and pulsatile signals in the alarm segment were checked. In the case that more than 90% of the R peaks were followed by the pulse onsets of the pulsatile signal, the delays between the peaks were measured. The average of the delays in the alarm segment was set to the delay for the whole segment containing the alarm. A default delay value of 200 ms was used for the segments which did not satisfy the criteria. Then, the R peaks and the corresponding pulse onset beats of the pulsatile signal were compared in a one second window through the whole alarm segment. The percentage of the R peaks matching the pulsatile onset beats in an interval of the corresponding delay between them was calculated. If the matching rate was above 90% then the signal quality was deemed acceptable. As our algorithm was not successful, further work is needed to improve the integration technique. Finding an optimum matching rate and adjusting the delay for the available signals could enhance the performance of the algorithm.

5. Conclusion

Our result placed us among the top ten scores of the *PhysioNet/Computing in Cardiology Challenge 2015*. Our proposed system achieved the highest score in detecting tachycardia false alarms. Our best performing algorithm used multimodal signals, combined the information from ECG and pulsatile signals, extracted and evaluated a number of features of the signals for alarm identification. Modification of the signal quality measures for different arrhythmias rather than employing a fixed SQI for every arrhythmia, setting the threshold in an iterative performance evaluation, and considering various possible effects of each arrhythmia on the features of the signals enhanced the arrhythmia identification performance. For future alarm management systems, a modified noise removal algorithm, adaptive SQI measurements for each arrhythmia, multimodal signal integration with optimum matching rate and adjusted delay could improve the performance.

Acknowledgments

References

Aboukhalil A, Nielsen L, Saeed M, Mark R G and Clifford G D 2008 Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform *J. Biomed. Inform.* **41** 442–51

Ansari S, Belle A and Najarian K 2015 Multi-modal integrated approach towards reducing false arrhythmia alarms during continuous patient monitoring: The PhysioNet Challenge 2015 *Comput. Cardiol.* **42** 1181–4

Behar J, Oster J, Li Q and Clifford G D 2013 ECG signal quality during arrhythmia and its application to false alarm reduction *IEEE Trans Biomedical Eng.* **60** 1660–6

Benitez D, Gaydecki P A, Zaidi A and Fitzpatrick A P 2001 The use of the Hilbert transform in ECG signal analysis *Comput. Biol. Med.* **31** 399–406

Chambrin M C, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C and Boniface B 1999 Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis *Intensive Care Med.* **25** 1360–6

Chambrin M C 2001 Alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit. Care* **5** 184–8

Chen S, Thakor N V and Mower M M 1987 Ventricular fibrillation detection by a regression test on the autocorrelation function *Med. Biol. Eng. Comput.* **25** 241–9

Clayton R H, Murray A and Campbell R W F 1993 Comparison of 4 techniques for recognition of ventricular-fibrillation from the surface ECG *Med. Biol. Eng. Comput.* **31** 111–7

Clifford G D, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D and Mark R G 2015 The PhysioNet/computing in cardiology challenge 2015: reducing false arrhythmia alarms in the ICU *Comput. Cardiol.* **42** 273–6

Clifford G D, Aboukhalil A, Sun J X, Zong W, Janz B A, Moody G B and Mark R G 2006 Using the blood pressure waveform to reduce critical false ECG alarms *Comput. Cardiol.* **33** 829–32

Clifford G D and Moody G B 2012 Signal quality in cardiorespiratory monitoring *Physiol. Meas.* **33** E01

de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P and O'Malley M 2003 Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea *IEEE Trans. Biomed. Eng.* **50** 686–96

Donchin Y and Seagull F J 2002 The hostile environment of the intensive care unit *Curr. Opin. Crit. Care* **8** 316–20

Fallet S, Yazdani S and Vesin J 2015 A multimodal approach to reduce false arrhythmia alarms in the intensive care unit *Comput. Cardiol.* **42** 277–80

Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P Ch, Mark R G, Mietus J E, Moody G B, Peng C-K and Stanley H E 2000 PhysioBank, PhysioToolkit and PhysioNet: Components of a new research resource for complex physiologic signals *Circulation* **101** e215–e220

Hickey B, Heneghan C and de Chazal P 2004 Non-episode-dependent assessment of paroxysmal Atrial Fibrillation through measurement of RR interval dynamics and atrial premature contractions *Ann. Biomed. Eng.* **32** 677–87

Imhoff M and Kuhls S 2006 Alarm algorithms in critical monitoring *Anesth. Analg.* **102** 1525–37

Jain N and Shakya D K 2014 Denoising baseline wander noise from electrocardiogram signal using fast ICA with multiple adjustments *Int. J. Comput. Appl.* **99** 34–9

Jekova I 2000 Comparison of five algorithms for the detection of ventricular fibrillation from the surface ECG *Physiol. Meas.* **21** 429–39

Johnson A E W, Behar J, Andreotti F, Clifford G D and Oster J 2015 Multimodal heart beat detection using signal quality indices *Physiol. Meas.* **36** 1665–77

Lawless S 1994 Crying wolf: False alarms in a pediatric intensive care unit *Crit. Care. Med.* **22** 981–5

Moody G, Moody B and Silva I 2014 Robust detection of heart beats in multimodal data: The PhysioNet/ Computing in Cardiology Challenge *Comput. Cardiol.* **41** 549–52

Novaes M A F P, Aronovich A, Ferraz M B, Knobel E 1997 Stressors in ICU: patients' evaluation *Intensive Care Med.* **23** 1282–5

146

Plesinger F, Klimes P, Halamek J and Jurak P 2015 False alarms in Intensive Care Unit monitors: detection of life-threatening arrhythmias using elementary algebra, descriptive statistics and fuzzy logic *Comput. Cardiol.* **42** 281–4

Redmond S J and Heneghan C 2006 Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea *IEEE Trans. Biomed. Eng.* **53** 485–96

Sadr N, Huvanandana J, Nguyen D T, Kalra C, McEwan A and de Chazal P 2015 Reducing false arrhythmia alarms in the ICU by Hilbert QRS detection *Comput. Cardiol.* **42** 1173–6

Sendelbach S and Funk M 2012 Alarm fatigue, a patient saftey concern *Nurs. Clin. North Am.* **47** 375–82

Shouldice R B, O'Brien L M, O'Brien C, de Chazal P, Gozal D and Heneghan C 2004 Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features *Sleep* **27** 784–92

Silva I, Moody B, Behar J, Johnson A, Oster J, Clifford G D and Moody G B 2015 Robust detection of heart beats in multimodal data *Physiol. Meas.* **36** 1629–44

Silva I, Moody G B and Celi L 2011 Improving the quality of ECGs collected using mobile phones: The PhysioNet/Computing in Cardiology Challenge 2011 *Comput. Cardiol.* **38** 273–6

Sun J X, Reisner A T and Mark R G 2004 A signal abnormality index for arterial blood pressure waveforms *Med. Biol. Eng. Comput.* **42** 698–706

Thakor N V, Webster J G and Tompkins W J 1984 Estimation of QRS complex power spectra for design of a QRS filter *IEEE Trans. Biomed. Eng.* **BME-31** 702–6

Thakor N V, Zhu Y S and Pan K Y 1990 Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm *IEEE Trans. Biomed. Eng.* **37** 837–43

Tinati M A and Mozaffary B 2006 A wavelet packets approach to electrocardiograph baseline drift cancellation *Int. J. Biomed. Imaging* **2006** 1–9

Zhang X S, Zhu Y S, Thakor N V and Wang Z Z 1999 Detecting ventricular tachycardia and fibrillation by complexity measure *IEEE Trans. Biomed. Eng.* **46** 548–55

Zong W, Heldt T, Moody G B and Mark R G 2003 An open-source algorithm to detect onset of arterial blood pressure pulses *Comput. Cardiol.* **30** 259–62

1354

147

Appendix III

Length-free near infrared measurement of new-born malnutrition

I contributed to the data modelling and statistical analysis with biostatistics guidance from Alun Pope. This paper established a foundation for the data modelling approach that was further developed and described in Section 6.

SCIENTIFIC REP⚙RTS

Length-free near infrared measurement of newborn malnutrition

Fatin Hamimi Mustafa[1], Emily J. Bek[1], Jacqueline Huvanandana[1], Peter W. Jones[1], Angela E. Carberry[1], Heather E. Jeffery[1,2,3], Craig T. Jin[1] & Alistair L. McEwan[1]

Under-nutrition in neonates can cause immediate mortality, impaired cognitive development and early onset adult disease. Body fat percentage measured using air-displacement-plethysmography has been found to better indicate under-nutrition than conventional birth weight percentiles. However, air-displacement-plethysmography equipment is expensive and non-portable, so is not suited for use in developing communities where the burden is often the greatest. We proposed a new body fat measurement technique using a length-free model with near-infrared spectroscopy measurements on a single site of the body - the thigh. To remove the need for length measurement, we developed a model with five discrete wavelengths and a sex parameter. The model was developed using air-displacement-plethysmography measurements in 52 neonates within 48 hours of birth. We identified instrumentation required in a low-cost LED-based screening device and incorporated a receptor device that can increase the amount of light collected. This near-infrared method may be suitable as a low cost screening tool for detecting body fat levels and monitoring nutritional interventions for malnutrition in neonates and young children in resource-constrained communities.

Accurate determination of the nutritional status of newborns is a major public health problem because under-nutrition increases the risk of immediate mortality and impacts on growth and cognitive development[1]. According to the World Health Organization (WHO), 44% of all under-five child deaths every year are neonates in their first 28 days of life, with most of these deaths occurring in the first week of life and the greatest incidence of these occurring in low-middle income settings[2]. Accurate assessment of body composition is vital as early identification of under and over nutrition in neonates and children and can guide interventions for nutritional management for clinicians and health-care workers[3]. Conventional approaches for recognising under-nutrition include the use of population-based percentiles (<10th, 5th, or 3rd percentiles), which rely on weight for gestational age and sex[4]. An alternative to the weight percentiles is to measure body composition.

A variety of methods are available to assess body composition in neonates and children. One high-cost technique uses the PEA POD system, a method based on air displacement plethysmography (ADP). The ADP technique (PEA POD; COSMED, Concord, CA) is often considered the criterion method for determining body composition and is accurate, safe, and noninvasive[5]. Carberry *et al.* showed that under-nutrition and risk of neonatal morbidity are more closely associated with body fat percentage (BF%) measured using ADP rather than the conventional birth weight percentiles[4]. Another high cost and highly accurate technique is dual-energy X-ray (DEXA), however its use is limited to one scan per year as it uses low dose ionising radiation[6]. Deuterium dilution for the measurement of total body water may be an ideal method for newborns as it involves less compliance, but requires trained staff for accurate dose delivery and sample collection, and carries a risk of delay due to time requirements for sample processing. Other techniques such as hydrostatic underwater weighing are unsuitable for newborns while less expensive techniques such as a skinfold thickness measurements require a high degree of training and can be inaccurate with poor predictive value due to incorrect lifting of the skin fold during the measurement, especially in lean newborns[7]. In this light, there is an urgent need for low cost and portable devices to assess body composition and nutrition as a point of care tool.

[1]School of Electrical and Information Engineering, Faculty of Engineering and Information Technologies, University of Sydney, New South Wales, Australia. [2]Sydney School of Public Health, University of Sydney, New South Wales, Australia. [3]Former Senior Staff Specialist in Neonatology, Royal Prince Alfred Hospital, Sydney, Australia. Correspondence and requests for materials should be addressed to F.H.M. (email: fatin.mustafa@sydney.edu.au)
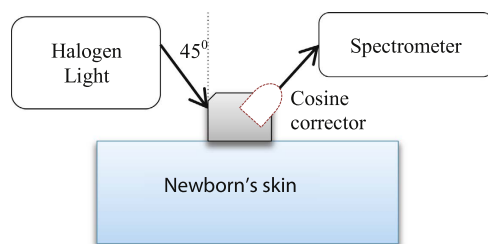
Figure 1. Near-infrared body fat measurement set up. Dotted object is the cosine corrector.

A relatively new technique for measuring body fat levels in neonates and children uses the near-infrared (NIR) interactance method. The NIR interactance method studies the response of light at specific wavelengths to variations in thickness of the subcutaneous fat layer. The principle is based on light interaction with the various tissue types including the skin, muscle, bone and fat. Depending on the wavelength and optical properties of each tissue, the light is absorbed or reflected by different magnitudes before being captured by a photo-detector[8]. NIR is considered safe, rapid and noninvasive if the power of the incident light is low enough not to heat the skin. This method can also be made mobile and affordable, where a NIR device can be directly connected and easily monitored via portable computing devices. However, the NIR method is limited by its sensitivity to hydration and skin color[8,9]. NIR body fat has been studied extensively in the adult population[8,10–12], but there is limited research on its use in neonates and young children[13]. Studies have found that healthy neonates and adults possess different skin structures including thickness of skin layers, size of cells and size of fibres, where the parameters are smaller in neonates[14–16]. Neonatal skin appeared to be more hydrated than adult skin under an electron micrograph[14,16] and, in general, water content in newborns was higher than in adults at 81% compared to 73%[17].

Previously, we found that NIR measurements taken on the newborn's thigh combined with weight and length data can provide a reasonable estimate for BF%[18]. The device was based on inexpensive LEDs and photodiodes. In this paper we consider a reflectance NIR measurement system having two different configurations: with and without a cosine corrector device connected at the collecting side of the probe. The cosine corrector acts as an optical diffuser that allows light to be collected from a wider range of angles compared with capturing the light using an uncorrected sub-miniature A-type (SMA) fibre cable.

We have developed a statistical model to estimate BF% in newborns from NIR reflectance measurements based on BF% from ADP measurements. As hydration is the major concern affecting NIR measurements in newborns, we exploit the NIR spectral absorption peaks of fat and water. Our model utilises three different ratios using five different wavelengths and an additional sex parameter. Our aim is to develop the best model for NIR measurement based on the ADP measurements, not to compare the two measurement methods at this stage.

Methods

Data Collection. The NIR measurements were taken at the tertiary referral hospital, Royal Prince Alfred Hospital (RPAH), Sydney between September 2014 and December 2014 on newborn of various ethnic backgrounds. Maternal conditions during pregnancy, birth details and maternal and paternal demographics including ethnicity, age, height, weight, date of birth, and education background were recorded. The measurements of all subjects were conducted in duplicate or triplicate on the skin surface of both anterior and medial thighs. The thigh was chosen as the measurement site because it is a convenient location that is accessible while breastfeeding. All devices were tested for medical safety by the RPAH Biomedical Engineering department and were found to meet IEC60601 medical safety regulations.

Measurement Set up. A schematic of the NIR measurement setup is illustrated in Fig. 1. The cosine corrector device is shown with a dotted line to indicate that it may or may not be included in the measurement process. A tungsten halogen light (Mikropack HL-2000-FHSA, 6.7 mW, 360 nm to 2400 nm range) was connected to a 3D-printed fibres holder via a SMA fibre (Thorlabs, M28L01, ⌀400 μm, 0.39 NA). The holder was designed with two holes positioned at −45° and +45° angles to the normal surface to locate the transmitter and receptor fibres respectively. At the receptor side, another SMA fibre (Thorlabs, M14L01, ⌀50 μm, 0.22 NA) was used to optically connect the holder to a spectrometer (Ocean Optics QEPRO-FL, 350 nm to 1100 nm range, SNR 1000:1) and then the response signal was recorded for 20s with OceanView 1.4 software (Ocean Optics). The measurement was repeated on different subjects with a cosine corrector (Thorlabs CCSA1, ⌀4 mm) coupled between the holder and the SMA fibre receptor.

BF% and weight were recorded by the PEA POD ADP measurement system. The BF% was determined by placing the naked newborn inside a closed chamber and air displacement was measured using pressure and volume changes. Body density was calculated from measured body mass and the calculated body volume[19]. Gestational age and length were obtained from the PEA POD database for data analysis.

Diffuse Reflectance Model. A simplified diffuse reflectance model of a layered medium, $R(x, y)$ from the Radiative Transfer Equation (RTE) with a corrected diffuse approximation (CDA) was defined[20]:

$$R(x, y) = -2\pi \int_{-1}^{NA} t(\gamma)I'(\gamma, x, y, 0)\gamma d\gamma \tag{1}$$
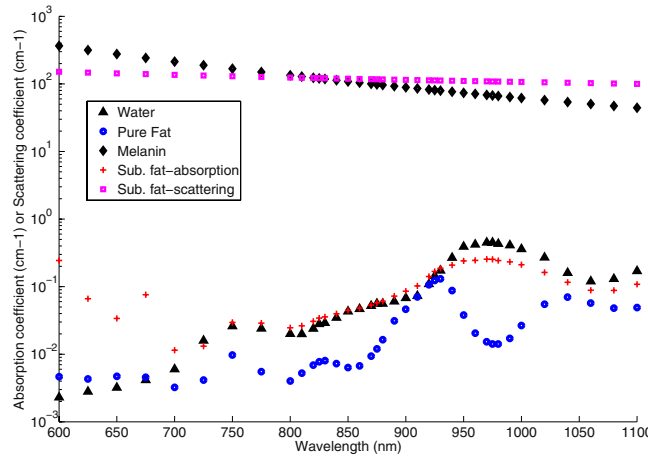
Figure 2. Absorption coefficient spectrum of melanin, absorption coefficient spectrum of pure water, absorption coefficient spectrum of pure fat and absorption coefficient spectrum of subcutaneous fat layer and scattering coefficient spectrum of subcutaneous fat layer.

where NA is defined as the numerical aperture of the detector used that is aligned normally to the boundary plane at $z = 0$, the $t(\gamma)$ is the Fresnel transmission coefficient due to the refractive index mismatch at the boundary, and the quantity I' relates to I as:

$$I'(\gamma, x, y, z) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} I(\gamma, \varphi, x, y, z)\gamma d\gamma \tag{2}$$

where $I(\gamma, \varphi, x, y, z)$ is the radiance over the range of angles ($\gamma = $ cosine ϑ, and φ) exiting the skin collected by the detector at positions indicated by the vector $<x, y, z>$. The angle ϑ is the elevation angle with respect to the *z-axis* in spherical coordinates, while φ is the azimuthal angle of the position vector. Note that the range of $-\pi < \varphi < \pi$ is due to the assumption of uniform scattering. The $I(\gamma, \varphi, x, y, z)$ depends on the optical properties: the absorption coefficient as function of absorption length, $\mu_a (l_a)$, the scattering coefficient as function of scattering length, $\mu_s (l_s)$, and the anisotropy, g. In detail these are given by:

$$\begin{aligned} I(\gamma, \varphi, x, y, z) \ = \ & \frac{H_1(\gamma, z)f(x, y)}{2\pi} \\ & + [1 - H_2(\gamma, z)]\phi(x, y, z) - 3\beta k_1[\gamma - H_3(\gamma, z)]\delta_z\phi(x, y, z) \\ & + O(\beta^2) + O(\alpha) \end{aligned} \tag{3}$$

where $\beta = 1/w(\mu_s)$, $\alpha = \mu_a/\mu_s$, $k_1 = 1/3\mu_s(1 - g)$. The w is the beam width, the $\phi(x, y, z)$ is the solution of the boundary value that can be solved either by Laplace's equation or the diffusion equation. Meanwhile, the $f(x, y)$ is the incident beam profile, which is set as Gaussian beam, and H_n ($n = 1, 2, 3$) denotes a half space Green's function[20]. Whilst the $t(\gamma)$ from equation (1) is given by:

$$t(\gamma) = \frac{2n_1\cos\theta_1}{n_1\cos\theta_1 + n_2\cos\theta_2} \tag{4}$$

where n_1 and n_2 are the refractive indices of the ambient material and skin respectively, θ_1 is the transmission angle of the light source and θ_2 is the reflected angle of the light from the skin. The NA from equation (1) is given by:

$$NA = n_i \ \sin\sigma \tag{5}$$

where n_i denotes the refractive index of material outside the fibre, which in our case is the air ($n = 1$), and σ donates maximum half acceptance angle of the fibre or cosine corrector. The NA of the used cosine corrector is 1.0 compared to 0.22 of the SMA fibre as stated in section 2.2, which result in acceptance angles from 0° to 180° and from 77.3° to 102.7° for the cosine corrector and the SMA fibre cable respectively.

Statistical Model Development. In previous NIR studies, the absorbance (A) ratio at two different wavelengths, $A(\lambda_1)/A(\lambda_2)$ was derived by K. Norris *et al.* in order to remove and normalise the baseline offset[21]. The idea of a ratio at two different wavelengths was used in developing our statistical model. To reduce the influence of water absorption, our preference was given to select ratios that were based on wavelengths highly influenced by fat and water[9,22,23]. Figure 2 shows the scattering spectrum of the subcutaneous fat layer, absorption spectrum of pure fat, absorption spectrum of melanin, absorption spectrum of pure water and also calculated absorption spectrum of subcutaneous fat layer following the Meglinski's equation model. From Fig. 2, the dominant effect of water can be observed between 850 nm and 1050 nm. The spectral curve of the subcutaneous fat layer imitates the curve of pure water even though the peak of pure fat is clearly at 930 nm. This is due to the water content in the
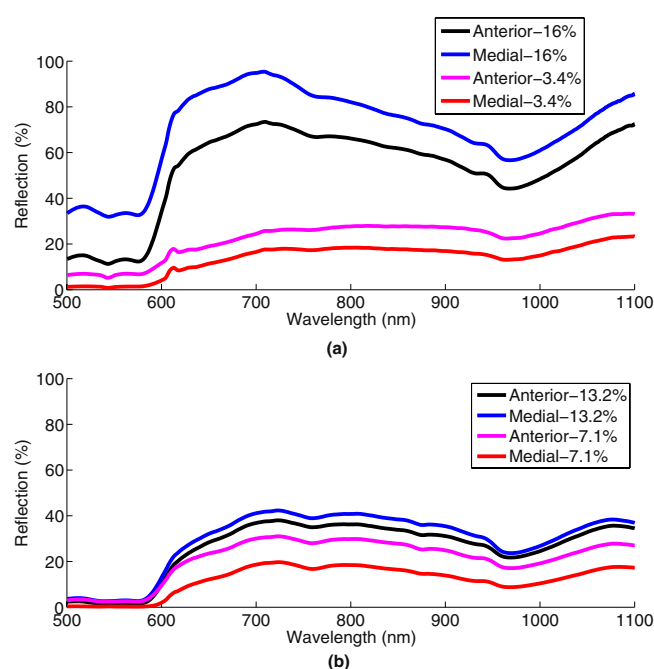
151

Figure 3. Near-infrared reflection from the anterior and medial thighs of the highest and lowest BF% of two subjects in each cohort (**a**) with cosine corrector, (**b**) without cosine corrector.

subcutaneous fat layer[9]. Scattering has a high influence in the subcutaneous fat layer (in Fig. 2), but the unidentified fat and water constituents over the spectrum make them difficult to separate.

We planned to use more than two wavelengths that were highly influenced by fat and water to counter the effect of melanin in the epidermal layer[9], shown in Fig. 2. As a consequence of the high absorption possessed by the melanin spectrum, we only selected subjects from the white skin category in developing our statistical model due to very low number of subjects from the other skin categories.

We used Matlab software (Version R2012b; Mathworks Incorporated, Natick, Massachusetts), SPSS (Version 22), Microsoft Excel (Version 2013) and R (Version 3.3.1) to perform the statistical analysis. Linear piecewise interpolation was first applied on the spectrometer readings to determine the reflection in 10 nm intervals within the range of 850 nm to 1100 nm. The model development process involved evaluating all possible ratios of these readings to determine key combinations of wavelengths that exhibited the highest correlation with BF% measured by ADP. A maximum of three sets of ratios and sex were used as input variables for an ordinary least-squares linear regression model, with NIR BF% as the target variable and BF% from ADP as a reference. Model performance was evaluated on the basis of coefficients between the predicted and actual values (significant level of 0.05, two-tailed test). The prediction of BF% by NIR was evaluated against ADP BF% from regression lines. The distribution of the data was tested using a Shapiro-Wilk test to ensure that the differences were normally distributed (Gaussian). Residuals were also plotted to compare variability of the developed model over the range of ADP measurements.

Ethics. This study was approved by the Human Research Ethics Committee of the Royal Prince Alfred Hospital, Sydney, Australia and all experimental methods were carried out in accordance with relevant guidelines and regulations (Protocol No.; X14-118, HREC/09/RPAH/645). Informed and written consent was obtained from the parents of the newborns.

Results

Sixty subjects were first measured using ADP. They were then split into two cohorts: Cohort 1 was measured using the NIR device with a cosine corrector fitted (the first 30 subjects) and Cohort 2 was measured without a cosine corrector (the next 30 subjects). To mitigate the risk of excessive movement, the newborns were measured while sleeping, immediately after a feed or during feeding. Figure 3(a,b) show NIR reflection spectra obtained from NIR measurements with using the cosine corrector and without using the cosine corrector respectively both for the anterior and medial thighs of two subjects. The subject selection was based on those having the highest and the lowest BF% from the ADP measurements. At this level of detail there appeared to be baseline offsets of the spectra that supported the implementation of ratios at two different wavelengths in the developed model.

Table 1 shows the characteristics of the neonates studied (total n = 60: n = 30 for each cohort). The development of the model in this study only considered white skin subjects (total n = 52: n = 26 for each cohort) due to the high influence of melanin over NIR spectra. The data for dark skin subjects could not be meaningfully modeled due to low numbers (total n = 8: n = 4 for each cohort). We found significant correlation between NIR absorption using the cosine corrector and ADP for both anterior and medial thighs of white skin subjects (correlation coefficient, R = 0.877 and R = 0.839 respectively) as shown in Table 2. Root Mean Squared Error (RMSE)

	White Skin[*]	Dark Skin[*]	Total	Sex	Body Fat % (Mean ± SD)	Gestational Age Weeks (Mean ± SD)	Weight Kg (Mean ± SD)	Length cm (Mean ± SD)
Cohort 1 With cosine corrector	26 (Caucasian-23 (88%)), Asian-3 (12%))	4 (Asian-4 (100%))	30	63% Male	10.10 ± 3.41	39.28 ± 1.60	3.23 ± 0.49	49.57 ± 3.18
Cohort 2 Without cosine corrector	26 (Caucasian-16 (62%), Asian-10 (38%))	4 (Asian-3 (75%), Aboriginal-1(25%))	30	50% Male	11.13 ± 3.72	39.47 ± 1.27	3.31 ± 0.43	49.42 ± 2.19

Table 1. Characteristics of 60 subjects studied. [*]Based on ethnicity information with skin colour recorded.

	Anterior thigh			Medial thigh		
	R	RMSE	P-value	R	RMSE	P-value
Cohort 1 (white skin subset, n = 26)	0.877	1.77	<0.001	0.839	2.0	<0.001
Cohort 2 (white skin subset, n = 26)	0.519	3.38	0.143	0.519	3.38	0.143

Table 2. Results from statistical analysis of NIR absorption for anterior and medial thighs of white skin subjects (total n = 52: n = 26 for each cohort).
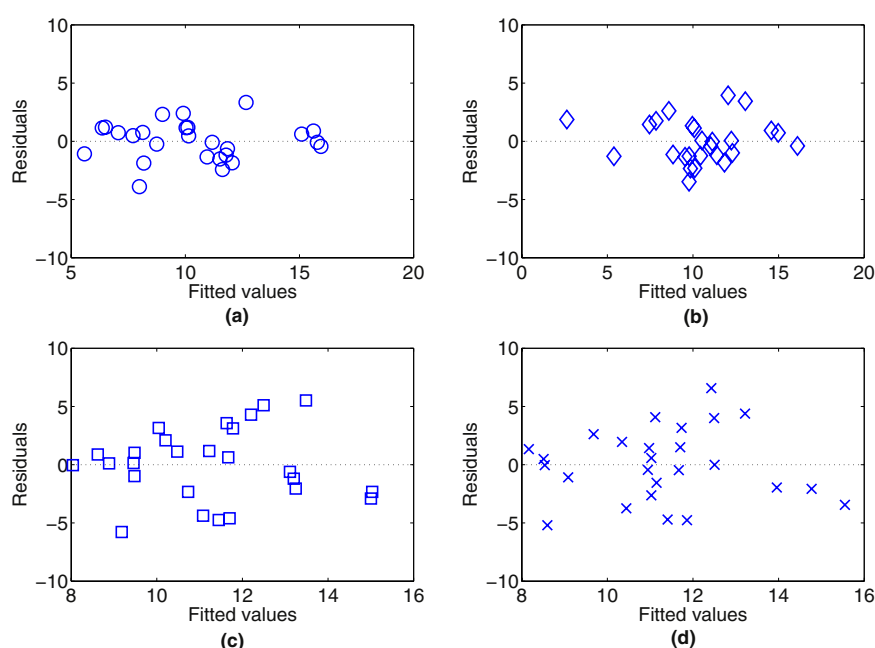
Figure 4. Residual plots of NIR BF% and ADP BF% with white skin subjects (total n = 52: n = 26 for each cohort). Plot (**a**) NIR with cosine corrector on anterior thigh, (**b**) NIR with cosine corrector on medial thigh, (**c**) NIR without cosine corrector on anterior thigh while (**d**) NIR without cosine corrector on medial thigh.

and *p*-values of Cohort 1 (white skin subset) of both sites are generally lower than those of Cohort 2. This is attributed to the greater light detection obtained using the cosine corrector.

The residual plots shown in Fig. 4 with and without the cosine corrector for both anterior and medial thighs of white skin subjects are evenly and are randomly dispersed throughout the x-axis (fitted value), which indicates that the assumptions of linearity and homoscedasticity are valid. The Shapiro-Wilk test also shows that the data sets follow a normal distribution.

The model based on white skin subjects utilised three different reflection ratios at five different wavelengths and a parameter of sex as:

$$BF\% = A_1 + A_2 r_1 + A_3 r_2 + A_4 r_3 + A_5 G \tag{6}$$

where BF% is the body fat percentage measured using ADP, r_1 is the ratio of 890 nm/1020 nm, r_2 is the ratio of 920 nm/1010 nm, r_3 is the ratio of 1010 nm/900 nm, and *G* is the sex which is assigned the value of 1 for male and 0 for female. A_1, A_2, A_3, A_4, and A_5 indicate the constant coefficients of the respective parameter. The values of these constant coefficients were obtained following a statistical model and are shown in Table 3.

Illustrated in Table 4 are the mean and standard deviation of the NIR BF% obtained from equation (6) and the ADP BF% of the two cohorts for the anterior and medial thighs of white skin subjects. Notice that the mean value for ADP BF% and NIR BF% are completely in agreement. The variability of the measurements is less for NIR (2.83%) than ADP (3.62%). The residual plots show a dissimilarity of variability between cohorts, where the

	A₁	A₂	A₃	A₄	A₅
Cohort 1 (white skin subset)					
Anterior thigh	−317.70	255.18	−83.25	193.38	−1.64
Medial thigh	−186.79	225.77	−141.47	115.04	−1.81
Cohort 2 (white skin subset)					
Anterior thigh	−214.29	103.11	−0.64	128.45	2.26
Medial thigh	−225.26	111.74	1.27	127.77	−2.07

Table 3. Values of the constant coefficients used in equation (6).

	ADP BF% (Mean ± SD) %	NIR BF% (Mean ± SD) %	
		Anterior thigh	Medial thigh
Cohort 1 (white skin subset, n = 26)	10.44 ± 3.373	10.44 ± 2.832	10.44 ± 2.830
Cohort 2 (white skin subset, n = 26)	11.26 ± 3.617	11.26 ± 1.875	11.26 ± 1.876

Table 4. Mean and standard deviation of NIR BF% in equation (6) and ADP BF% of two cohorts of white skin subjects (total n = 52: n = 26 for each cohort).

residual of the NIR with the cosine corrector (Fig. 4(a,b) for anterior and medial thighs respectively) was lower than the NIR without using the cosine corrector (Fig. 4(c,d) for anterior and medial thighs respectively) demonstrating less variability with the cosine corrector.

The regression lines of the NIR model BF% against the ADP BF% show that the NIR model over-estimates BF% for lean neonates, with the cosine corrector (Fig. 5(a,b)) reducing the over-estimation of BF% compared to the model without the cosine corrector (Fig. 5(c,d)).

Melanin or skin colour has been found to influence NIR measurements on skin[9]. We extended our analysis to include all subjects of both cohorts (n = 60: n = 30 for each cohort) for anterior and medial thighs using our developed NIR BF% model. The effect of including the mixed skin colours in the analysis is reflected in the statistical results in Table 5, which shows reduced correlation values and higher RMSE and *p*-values than those presented in Table 4. Nevertheless, the statistical results in Table 5 (mixed skin colour) are consistent with the results in Table 4 (white skin colour only) in showing that NIR using the cosine corrector was always better than NIR without using the cosine corrector.

As the choice of using three different wavelength ratios was somewhat arbitrary, we show the effects on R using between one and five ratios in our NIR BF% model for white skin colour subjects in Table 6 (n = 26 of cohort 1 for both anterior and medial thighs). The selection of wavelengths was still referred to absorption peaks of water and fat. The increases of R which resulted from increasing the number of wavelength ratios used agrees with a study of breast imaging by Lo *et al.*, which they found that adding more wavelengths up to eight improved extraction errors[24]. Although the R continues to increase with the number of wavelength ratios used, the increase is leveling off and the inclusion of these additional wavelengths would involve more cost in the final device which is intended for use with discrete diodes for low cost. We included the subject length parameter in Table 6, row 5 for comparison with other independent ratios of NIR where it can be seen that it provides at least as much information as two additional ratios or three additional wavelengths.

Discussion

To our knowledge, this is the first study developing NIR BF% for neonates using BF% from a gold standard body composition ADP technique to develop models. Our findings showed that NIR configuration with the cosine corrector resulted in the highest correlation and the lowest residual. We found that the NIR measurements after adding the cosine corrector showed higher correlation between NIR BF% and ADP BF% (R = 0.877 and R = 0.839 for anterior and medial thighs respectively on white skin subjects). The NIR using the cosine corrector also showed the lowest residuals with the ADP BF% indicating the ability to fit an improved model. The improvement offered by the cosine corrector in the NIR body fat measurement demonstrates the importance of appropriate equipment and device selection in NIR configurations. This was in agreement with a study by Hwang *et al.*, in which they used varied types of LEDs with different view angles (lamp at 20° over miniature chip at 120°) in NIR phantom experiments resulting in higher sensitivity by the miniature chip[25].

Anthropometric parameters (e.g.; length and weight) and age were not included in our developed model as these measurements are often unknown or inaccurate in low resource settings. Our objective was to determine whether this NIR technique could be used independently for easy point of care measurement of nutritional status. Our developed model however depended on an additional parameter, sex, because this factor influences BF% at birth[26]. Nevertheless, the sex parameter is always accessible and does not need any equipment. Measurement of length has been used as one of the primary indicators of foetal, neonatal and child nutrition[27]. However, length measurements are often problematic due to inter- and intra-observer variability unless appropriate equipment including an appropriate length board with extensive training are used[27]. The Futrex device requires anthropometric parameters of age, weight, height and level of exercise in their developed model[28]. Those parameters may often be inaccessible or unreliable, especially in a low-income setting.
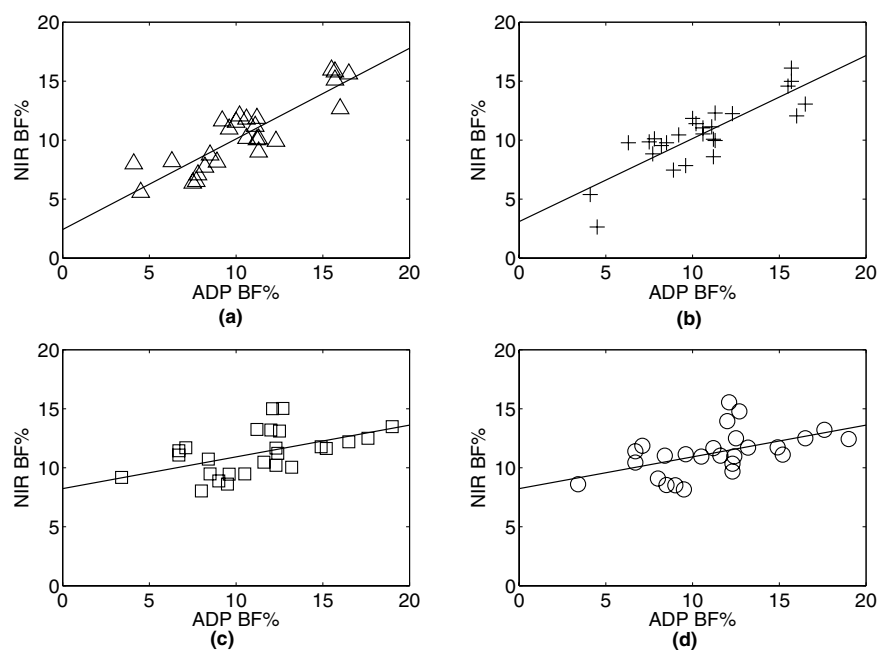
Figure 5. Linear regression lines of NIR BF% and ADP BF% with white skin subjects (total n = 52: n = 26 for each cohort). Plot (**a**) NIR with cosine corrector on anterior thigh, (**b**) NIR with cosine corrector on medial thigh, (**c**) NIR without cosine corrector on anterior thigh while (**d**) NIR without cosine corrector on medial thigh.

	Anterior thigh			Medial thigh		
	R	**RMSE**	***P*-value**	**R**	**RMSE**	***P*-value**
Cohort 1 (all subjects, n = 30)	0.820	2.10	<0.001	0.719	2.53	<0.001
Cohort 2 (all subjects, n = 30)	0.363	3.74	0.45	0.420	3.64	0.28

Table 5. Results from statistical analysis of NIR absorption for anterior and medial thighs of all subjects (total n = 60: n = 30 for each cohort) with (Cohort 1) and without (Cohort 2) a cosine corrector.

No.	Cohort 1 – With Cosine Corrector (n = 26)	R/RMSE (Anterior thigh)	R/RMSE (Medial thigh)
1	$BF\% = r_1 + G$	0.498/3.05	0.462/3.12
2	$BF\% = r_1 + r_2 + G$	0.602/2.87	0.721/2.49
3	$BF\% = r_1 + r_2 + r_3 + G + r_4$	0.877/1.81	0.839/2.05
4	$BF\% = r_1 + r_2 + r_3 + G + r_4 + r_5$	0.879/1.84	0.854/2.01
5	$BF\% = r_1 + r_2 + r_3 + G + L$	0.891/1.71	0.867/1.88

Table 6. BF% estimations, R and RMSE of less and more than three ratios. Ratios of r_1, r_2, and r_3 are the ratios in equation (6) while added ratios of r_4, and r_5 are 930 nm/1050 nm, and 970 nm/930 nm respectively and L is length. Subjects are from white skin colour (n = 26 for Cohort 1).

Past NIR subcutaneous fat studies using simulations and phantom measurement methods have found that light reflection increases logarithmically with the thickness of the fat layer[25,29]. The logarithmic relationship can be attributed to the high influence of fat on light scattering. As the thickness of fat increases, more light is scattered back to the ambient surface until at a critical thickness, the point where the light is saturated and there is no increase in backscattered light with any further increases of the thickness of fat. This critical thickness constrains the NIR method to the lower subcutaneous fat thicknesses associated with undernutrition in neonates and children. For example the subcutaneous fat layer beneath the thigh skin of low birth weight and normal full term neonates were found to be 1.7 mm to 3.0 mm and 3.0 mm to 5.0 mm respectively[12]. To our knowledge, there have been no specific studies to determine NIR maximum thickness detection, which might rely on the type of NIR devices used and acceptance power source emitted[25].

The strengths of our study include that this was the first development of NIR BF% models using BF% from ADP as a reference in a newborn population in order to look at the potential of direct and low cost LED-based

155

NIR device being comparable to the gold standard ADP in screening undernutrition and morbidity in newborns. NIR commercial devices, such as Futrex have only targeted on children above 5 years to adult populations[30–32]. Since this study was only conducted in newborns, other populations across infancy (newborn up to 2 years) with higher number of subjects, different skin colours and environments (temperatures) need to be investigated in order to establish the developed model and to formally test levels of agreement on a different dataset than that used to develop the models. Testing for agreement with several other gold standard body composition techniques using new NIR dataset is also essential to ensure the robustness of the NIR BF% model. For implementation, five LEDs could be allocated around a low-cost cosine corrector that couples to a wavelength-range detector and filters.

In conclusion, we have developed a NIR-based BF% model using gold standard ADP measurements. The developed NIR BF% models have utilised three ratios at five different wavelengths with the introduced cosine corrector to determine newborn body composition. The results showed significant correlation and agreement with ADP. We have shown that our device may have the potential to identify undernourished newborns who are at significant risk of associated morbidity such as hypothermia, hypoglycaemia and mortality in settings where a gold standard device would not normally be available. This is particularly useful for low resource settings where equipment to screen for hypoglycemia (glucometers, blood glucose analysis) and hypothermia (low reading thermometers) are limited or lacking. In such settings, a point of care, accurate, robust and low-cost device is needed to distinguish between the pathologically versus constitutionally small for gestational age neonates. The future plan is to test the device in a randomised controlled trial measuring relevant health outcomes.

References

1. Levitsky, D. A. & Strupp, B. J. Malnutrition and the brain: changing concepts, changing concerns. *The Journal of nutrition* **125,** 2212S–2220S (1995).
2. Organization, W. H. *World health statistics 2010.* (World Health Organization, 2010).
3. Wells, J. & Fewtrell, M. Measuring body composition. *Archives of disease in childhood* **91,** 612–617 (2006).
4. Carberry, A. E., Raynes-Greenow, C. H., Turner, R. M., Askie, L. M. & Jeffery, H. E. Is body fat percentage a better measure of undernutrition in newborns than birth weight percentiles? *Pediatric research* **74,** 730–736 (2013).
5. Christensen, K. & Kushner, R. Measuring body fat in the clinical setting. *Obesity Management* **3,** 93–95 (2007).
6. Damilakis, J., Adams, J. E., Guglielmi, G. & Link, T. M. Radiation exposure in X-ray-based imaging techniques used in osteoporosis. *European radiology* **20,** 2707–2714 (2010).
7. Olhager, E. & Forsum, E. Assessment of total body fat using the skinfold technique in full-term and preterm infants. *Acta Paediatrica* **95,** 21–28 (2006).
8. Conway, J. M., Norris, K. H. & Bodwell, C. A new approach for the estimation of body composition: infrared interactance. *The American journal of clinical nutrition* **40,** 1123–1130 (1984).
9. Jacques, S. L. Optical properties of biological tissues: a review. *Physics in medicine and biology* **58,** R37 (2013).
10. Möller, R. *et al.* Estimating percentage total body fat and determining subcutaneous adipose tissue distribution with a new noninvasive optical device LIPOMETER. *American Journal of Human Biology* **12,** 221–230 (2000).
11. Hong, H. K., Jo, Y. C., Choi, Y. S., Park, H. D. & Kim, B. J. An optical system to measure the thickness of the subcutaneous adipose tissue layer. Paper presented at *The 8th Annual IEEE Conference on Sensors: IEEE SENSORS 2009, Christchurch, New Zealand.* San Diego, CA: IEEE, (doi: 10.1109/icsens.2009.5398349) (Oct., 2009).
12. Hwang, I. D. & Shin, K. Fat thickness measurement using optical technique with miniaturized chip LEDs: A preliminary human study. Paper presented at *Engineering in Medicine and Biology Society, 2007: 29th Annual International Conference of the IEEE,* IEEE, (doi: 10.1109/iembs.2007.4353351) (2007).
13. Kasa, N. & Heinonen, K. Near-infrared interactance in assessing superficial body fat in exclusively breast-fed, full-term neonates. *Acta Pædiatrica* **82,** 1–5 (1993).
14. Holbrook, K. A histological comparison of infant and adult skin. *Neonatal Skin: Structure and Function. New York: Marcel Dekker,* 3–31 (1982).
15. Barel, A. O., Paye, M. & Maibach, H. I. *Handbook of cosmetic science and technology.* (CRC Press, 2014).
16. Stamatas, G. N., Nikolovski, J., Luedtke, M. A., Kollias, N. & Wiegand, B. C. Infant skin microstructure assessed *in vivo* differs from adult skin in organization and at the cellular level. *Pediatric dermatology* **27,** 125–131 (2010).
17. Wang, Z. *et al.* Hydration of fat-free body mass: new physiological modeling approach. *American Journal of Physiology-Endocrinology and Metabolism* **276,** E995–E1003 (1999).
18. McEwan, A. *et al.* Low-cost near-infrared measurement of subcutaneous fat for newborn malnutrition. Paper presented at *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring,* International Society for Optics and Photonics. (doi: 10.1117/12.2044764) (2014).
19. Ward, L. C., Poston, L., Godfrey, K. M. & Koletzko, B. Assessing Early Growth and Adiposity: Report from an EarlyNutrition Academy Workshop. *Annals of Nutrition and Metabolism* **63,** 120–130, doi: 10.1159/000350702 (2013).
20. Rohde, S. B. Modeling diffuse reflectance measurements of light scattered by layered tissues, PhD thesis, University of California (2014).
21. Workman Jr, J. & Springsteen, A. *Applied spectroscopy: a compact reference for practitioners.* (Academic Press, 1998).
22. Meglinski, I. V. & Matcher, S. J. Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions. *Physiological measurement* **23,** 741 (2002).
23. Vogel, A. J. Noninvasive Optical Imaging Techniques as a Quantitative Analysis of Kaposi's Sarcoma Skin Lesions, PhD thesis, University of Maryland (2007).
24. Lo, J. Y. *et al.* Wavelength Optimization for Quantitative Spectral Imaging of Breast Tumor Margins. *PLoS ONE* **8,** e61767, doi: 10.1371/journal.pone.0061767 (2013).
25. Hwang, I. D., Shin, K., Ho, D.-S. & Kim, B.-M. Evaluation of chip LED sensor module for fat thickness measurement using tissue phantoms. Paper presented at *Engineering in Medicine and Biology Society, 2006: 28th Annual International Conference of the IEEE,* IEEE. (doi: 10.1109/iembs.2006.259965) (2006).
26. Hawkes, C. P. *et al.* Gender-and gestational age–specific body fat percentage at birth. *Pediatrics* **128,** e645–e651 (2011).
27. Wood, A. J., Raynes-Greenow, C. H., Carberry, A. E. & Jeffery, H. E. Neonatal length inaccuracies in clinical practice and related percentile discrepancies detected by a simple length-board, Neonatal length measurement inaccuracies. *Journal of paediatrics and child health* **49,** 199–203, doi: 10.1111/jpc.12119 (2013).
28. Hartmann, S. *et al.* Phantom of Human Adipose Tissue and Studies of Light Propagation and Light Absorption for Parameterization and Evaluation of Noninvasive Optical Fat Measuring Devices. *Optics and Photonics Journal* **5,** 33 (2015).

29. Nilubol, C., Treerattrakoon, K. & Mohammed, W. S. Monte Carlo modeling (MCML) of light propagation in skin layers for detection of fat thickness. Paper presented at *Southeast Asian International Advances in Micro/Nano-technology*, International Society for Optics and Photonics. (doi: 10.1117/12.863536) (2010).
30. Fthenakis, Z. G., Balaska, D. & Zafiropulos, V. Uncovering the FUTREX-6100XL prediction equation for the percentage body fat. *Journal of medical engineering & technology* **36,** 351–357 (2012).
31. Heyward, V. H. & Gibson, A. *Advanced fitness assessment and exercise prescription 7th edition*. (Human kinetics, 2014).
32. Moon, J. R. *et al.* Percent body fat estimations in college women using field and laboratory methods: a three-compartment model approach. *Journal of the International Society of Sports Nutrition* **4,** 16 (2007).

Acknowledgements

Author Contributions

F.H.M. and A.M. provided the main engineering idea. F.H.M. wrote the manuscript. F.H.M., E.B. and P.J. were involved in the clinical measurements on subjects at RPAH. F.H.M. and J.H. contributed to the statistical model development while A.C. and H.J. provided clinical input and advice involving subjects studied. H.J. and A.M. recognised the need for a suitable device to measure body fat in young children at risk of undernutrition in low-income settings. All authors (F.H.M., A.M., E.B., P.J., J.H., A.C., H.J. and C.J.) reviewed the manuscript.

Additional Information

Competing financial interests: This project was funded by the Bill and Melinda Gates Foundation Grand Challenges Scheme OPP1111820. All authors received a salary or scholarship funding from The University of Sydney who have filed a provisional patent. The authors declare that they are bound by confidentiality agreements that prevent them from disclosing their financial interests in this work.

How to cite this article: Mustafa, F. H. *et al.* Length-free near infrared measurement of newborn malnutrition. *Sci. Rep.* **6**, 36052; doi: 10.1038/srep36052 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix IV

How do different brands of size 1 Laryngeal Mask Airway compare to facemask ventilation in a dedicated Larygeal Mask Airway teaching manikin?

I contributed to the signal processing of and feature extraction from the ventilator data.

How do different brands of size 1 laryngeal mask airway compare with face mask ventilation in a dedicated laryngeal mask airway teaching manikin?

Mark Brian Tracy,[1,2] Archana Priyadarshi,[1,2] Dimple Goel,[1,2] Krista Lowe,[1] Jacqueline Huvanandana,[3] Murray Hinder[1,3]

[1]Westmead Hospital, Neonatal Intensive Care Westmead, Westmead, New South Wales, Australia
[2]Department of Paediatrics and Child Health, Sydney University, Sydney, New South Wales, Australia
[3]Faculty of Engineering and Information Technologies, BMET Institute, Sydney University, Sydney, New South Wales, Australia

Correspondence to
Dr Mark Brian Tracy, Paediatrics and Child Health, Sydney University, PO Box 533, Wentworthville, NSW 2145, Australia; mark.tracy@sydney.edu.au

ABSTRACT

Background International neonatal resuscitation guidelines recommend the use of laryngeal mask airway (LMA) with newborn infants (≥34 weeks' gestation or >2 kg weight) when bag-mask ventilation (BMV) or tracheal intubation is unsuccessful. Previous publications do not allow broad LMA device comparison.

Objective To compare delivered ventilation of seven brands of size 1 LMA devices with two brands of face mask using self-inflating bag (SIB).

Design 40 experienced neonatal staff provided inflation cycles using SIB with positive end expiratory pressure (PEEP) (5 cmH$_2$O) to a specialised newborn/infant training manikin randomised for each LMA and face mask. All subjects received prior education in LMA insertion and BMV.

Results 12 415 recorded inflations for LMAs and face masks were analysed. Leak detected was lowest with i-gel brand, with a mean of 5.7% compared with face mask (triangular 42.7, round 35.7) and other LMAs (45.5–65.4) (p<0.001). Peak inspiratory pressure was higher with i-gel, with a mean of 28.9 cmH$_2$O compared with face mask (triangular 22.8, round 25.8) and other LMAs (14.3–22.0) (p<0.001). PEEP was higher with i-gel, with a mean of 5.1 cmH$_2$O compared with face mask (triangular 3.0, round 3.6) and other LMAs (0.6–2.6) (p<0.001). In contrast to other LMAs examined, i-gel had no insertion failures and all users found i-gel easy to use.

Conclusion This study has shown dramatic performance differences in delivered ventilation, mask leak and ease of use among seven different brands of LMA tested in a manikin model. This coupled with no partial or complete insertion failures and ease of use suggests i-gel LMA may have an expanded role with newborn resuscitation as a primary resuscitation device.

What is already known on this topic?

► The International Liaison Committee on Resuscitation guidelines recommend use of LMA at newborn resuscitation if adequate face mask positive pressure ventilation or endotracheal intubation insertion cannot be accomplished.
► Use of non-inflatable LMA in paediatric patients is beneficial due to possible cuff hyperinflation of inflatable cuff types.
► Efficacy of LMA use during neonatal resuscitation remains unclear.

What this study adds?

► Size 1 LMAs tested in this manikin study showed marked performance variation.
► The i-gel brand LMA showed superior performance for provision of peak inspiratory pressure, positive end expiratory pressure and mask leak compared with other LMAs and face masks tested.
► Inexperienced users of LMA showed higher proficiency in attaining adequate lung inflation and ease of insertion using i-gel brand LMA.

INTRODUCTION

Laryngeal mask airways (LMA) are listed in the International Liaison Committee on Resuscitation (ILCOR) neonatal resuscitation guidelines as a means of assisting positive pressure ventilation (PPV) during resuscitation of term and near-term infants where face mask PPV is not adequate and resuscitator skills to effect endotracheal intubation are insufficient.[1] LMA may also be of assistance with newborn infants with rare significant craniofacial abnormalities such as Pierre Robin syndrome where upper airways obstruction is severe and endotracheal intubation even in expert hands may be very difficult. Simply put, the LMA may be a life-saving device when 'you can't ventilate and you can't intubate'.[2 3] Assessment of LMA for use by local birth attendants (physicians and midwives) in developing and developed countries in manikin models compared with face mask PPV showed the feasibility of training and use of LMA in these settings and recommended further evaluation.[4–6]

The 2005 Cochrane review on the use of LMA versus bag-mask ventilation or endotracheal intubation comments that evidence from observational studies suggests LMA can provide rescue airway and achieve effective PPV during newborn resuscitation if both bag-mask ventilation and intubation have been unsuccessful.[7] Two recent prospective, unblinded, randomised, controlled trials comparing LMA to face mask ventilation for newborn ≥34 weeks' gestation found there was less need for endotracheal intubation using LMA and it was more effective than face mask.[8 9]
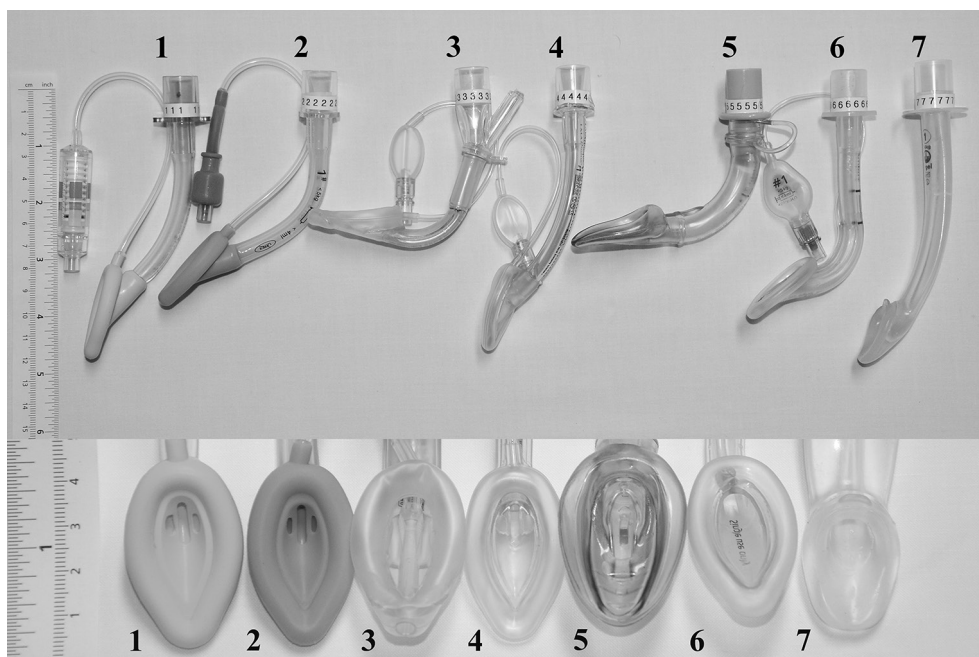
Original article

Figure 1 Laryngeal mask airway tested, physical characteristics: (1) Ultimate, (2) PRO-Breathe, (3) Supreme, (4) Unique, (5) air-Q, (6) AuroOnce and (7) i-gel.

Training for use of LMA in newborn resuscitation requires leak-free manikins with anatomically correct pharyngeal structures to apply LMA cushion to seal. Size 1 LMA brands differ substantially in stiffness of angled tube, the angle of the tube, tube internal diameter, and the size, shape and type of mask cushion (inflatable with a syringe, auto inflate with PPV or non-inflatable contoured gel) (figure 1). This suggests the potential for variation in performance characteristics. Studies to date of comparisons between LMA brands have been limited to up to three LMA model comparisons.[10–15] Clinical studies of LMA use in infancy by anaesthetists have been conducted in more controlled clinical situations of elective surgery with stable patients; brand comparisons have focused on ease of use, time to insert and inflation pressure at which audible leak occurs.[10–12] The neonatal resuscitation focus on sufficient tidal volume (TV) with least peak inflation pressure, in newborn infants with significant lung disease, has dictated the search for devices and methods to reduce face mask leak. There is a need to guide clinicians how a range of different LMA devices perform in dedicated LMA training manikins and clinical practice at birth.

We aimed to examine delivered ventilation and airway leak to an anatomically correct manikin head designed specifically to train in LMA use attached to a neonatal test lung. This crossover study was designed to test seven size 1 LMAs compared with two face masks (round and triangular). Our null hypothesis was that there would be no differences between face mask and brands of LMA in the rate of successful insertion (LMA), delivered ventilation, maximal inflation pressure, mask leak and particularly the ability to achieve targeted positive end expiratory pressures (PEEP). Primary outcomes were the delivered ventilation to the manikin, and the secondary outcomes were the insertion failure tally and subjective assessment of ease of LMA insertion.

METHODS
Forty clinicians in a busy tertiary neonatal intensive care unit (9 consultants/fellows, 13 registrars and 18 nurses) agreed to participate in this study. All were experienced in bag-mask ventilation

and variably less experienced with LMA use. A new AirSim Baby manikin head (JR10001, TruCorp, Belfast, Ireland) was modified by changing the supplied lung bag to a test lung (SmartLung Infant, IMT Medical, Buchs, Switzerland) with a static compliance of 2 mL/cmH$_2$O and resistance of 50 cmH$_2$O/L/s, simulating a near-term infant (figure 2). Seven internationally available size 1 LMA single-use devices were examined: (1) Ultimate (Ultimate Medical, Tianjin Medis, China), (2) PRO-Breathe (Well Lead Medical, China), (3) LMA Supreme (The Laryngeal Mask, Seychelles), (4) Unique LMA (The Laryngeal Mask), (5) air-Qsp (Cookgas, Malaysia), (6) AuroOnce (Ambu A/S, Ballerup, Denmark) and (7) i-gel (Intersurgical, Wokingham, Berkshire, UK).

Two face mask devices were compared with LMA: (1) Ambu (Ambu A/S) triangular size infant (Part No 252 052) and (2) Laerdal (Stavanger, Norway) round size 1 (Part No 851600). PPV was supplied using an Ambu self-inflating bag (SIB) (SPUR II Part No 335 102 000) with Ambu manometer and PEEP valve (0–20 cmH$_2$O, Part No 199 102 001). Two neonatal respiratory function monitors (RFM) (Florian, Acutronics, Switzerland, and Cosmo Novametrix) were used to determine the TV, flow and inflation pressures at two points. At point 1, the pneumotach of the Florian RFM was sited between the SIB and the proximal end of the device under test (LMA or face mask). At point 2, the pneumotach of the Cosmo RFM was sited between the test lung and the manikin head. Both RFMs were calibrated with an external syringe of known volume and pressure/flow via a traceable reference ventilator analyser (PF300, IMT Medical). Participants were blinded to both RFM displays. The manikin head was assessed for system leak and leak points at the base of the head where airway bifurcates and the corner of the lip was sealed with silicone. Test lung, manikin head and measurement system (RFM pneumotach's proximal pressure lines) were pressurised to a static pressure of 50 cmH$_2$O, and over 120 s there was no fall in pressure, indicating the system was leak-free.

A PowerLab data acquisition system (Part No ML880, ADInstruments, Australia) with a sample rate of 200 Hz and
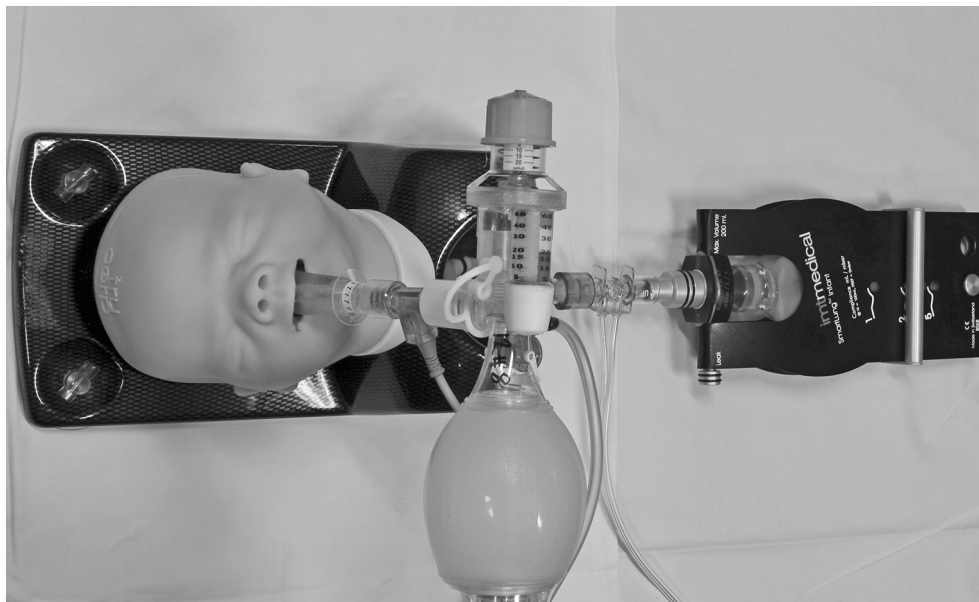
Figure 2 Manikin/lung test set-up.

a laptop computer collected analogue signals for volume, airflow and airway pressure from both RFMs. Respiratory parameters from each RFM (peak inspiratory pressure (PIP), PEEP, TV), for each breath, were determined by a customised program software algorithm (Python Software Foundation).

Leak during PPV was determined as (TV proximal (SIB) − TV distal (test lung))/TV proximal (SIB)×100.

Each subject received extensive instruction and practice with both face mask and LMA insertion/cuff inflation with each brand using the manikin head over several sessions. Competency with each method was assessed in the training phase by the instructor determining adequate test lung inflation and lack of audible leak.

Participants were asked to deliver 2 min of PPV aiming to achieve adequate test lung inflation for each randomised device; a 2 min rest period was provided between each test sequence. If the LMA device required cuff inflation, a volume of 4 mL was used, reflecting recommended maximum inflation volume. If the subject judged test lung inflation inadequate or there was no test lung movement, this was noted as an unsuccessful attempt; a second placement of the device was allowed, and two unsuccessful attempts were noted as a complete failure. Successful PPV following the second insertion was judged a partial failure (table 1). PEEP valve was set at 5 cmH$_2$O and PIP was not specified beyond adequate test lung inflation. The inflation rate (40–60) is as per ILCOR guidelines.[1] At the end of each sequence the subjects were instructed to provide four inflations of maximum SIB compression to determine if delivery of PIP up to the SIB overpressure value of 40 cmH$_2$O (+/−5 cmH$_2$O) could be achieved, simulating the need to increase PIP during PPV at resuscitation.[3 16]

Table 1 Delivered test lung respiratory data, device insertion failures and device insertion rating

Primary outcomes						Secondary outcomes	
	Mask leak,* %	PEEP,* cmH$_2$O	PIP,* cmH$_2$O	Inflation volume,* mL	Max PIP,*‡ cmH$_2$O	LM insertion† Failures, no failure/ partial/complete	LM insertion† Easy/Difficult
	Estimated mean (SE)	Estimated mean (SE)	Estimated mean (SE)	Estimated mean (SE)	Estimated mean (SE)		
Face mask							
M1. Triangular	43.0 (0.44)	3.0 (0.04)	22.7 (0.18)	28.1 (0.19)	29.5 (0.87)	40/0/0	40/0
M2. Round	34.9 (0.39)	3.6 (0.04)	25.5 (0.16)	31.3 (0.17)	31.6 (0.80)	40/0/0	40/0
Laryngeal mask							
1. Ultimate	48.8 (0.38)	2.2 (0.03)	20.6 (0.16)	30.3 (0.17)	25.5 (0.84)	36/2/2	34/6
2. PRO-Breathe	64.7 (0.40)	1.3 (0.04)	16.5 (0.16)	25.4 (0.17)	20.9 (0.81)	33/6/1	26/14
3. Supreme	46.2 (0.37)	2.4 (0.03)	22.0 (0.15)	29.0 (0.16)	29.2 (0.81)	37/3/0	38/2
4. Unique	64.5 (0.40)	0.7 (0.04)	14.3 (0.16)	27.2 (0.18)	18.5 (1.01)	26/9/5	27/13
5. air-Q	58.5 (0.39)	1.0 (0.04)	16.8 (0.16)	28.9 (0.17)	22.4 (0.83)	28/12/0	25/15
6. AuroOnce	62.2 (0.41)	0.7 (0.04)	17.0 (0.17)	29.7 (0.18)	19.5 (0.83)	28/7/5	16/24
7. i-gel	4.0 (0.42)	4.8 (0.04)	29.5 (0.17)	34.5 (0.19)	36.7 (0.95)	40/0/0	40/0

*Parameters significant at p<0.001, analysis of variance for repeated measures, estimated means with SE.
†Frequency counts for LMA only tested, Fisher's exact test, p<0.001.
‡Averaged four maximal inflations to achieve highest possible PIP.
LM, laryngeal mask; PEEP, positive end expiratory pressure; PIP, peak inspiratory pressure.
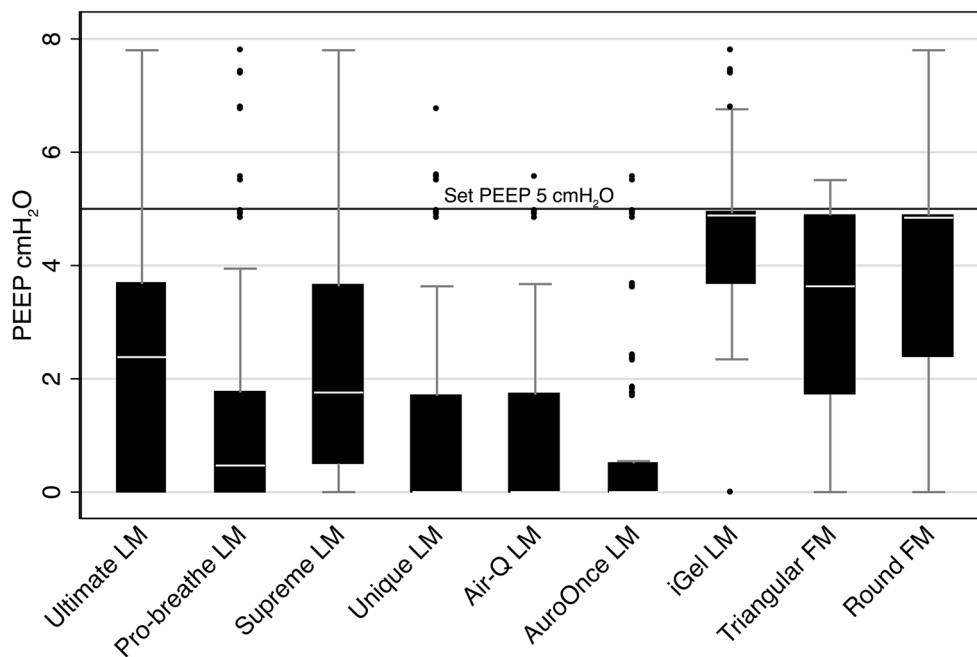
F3

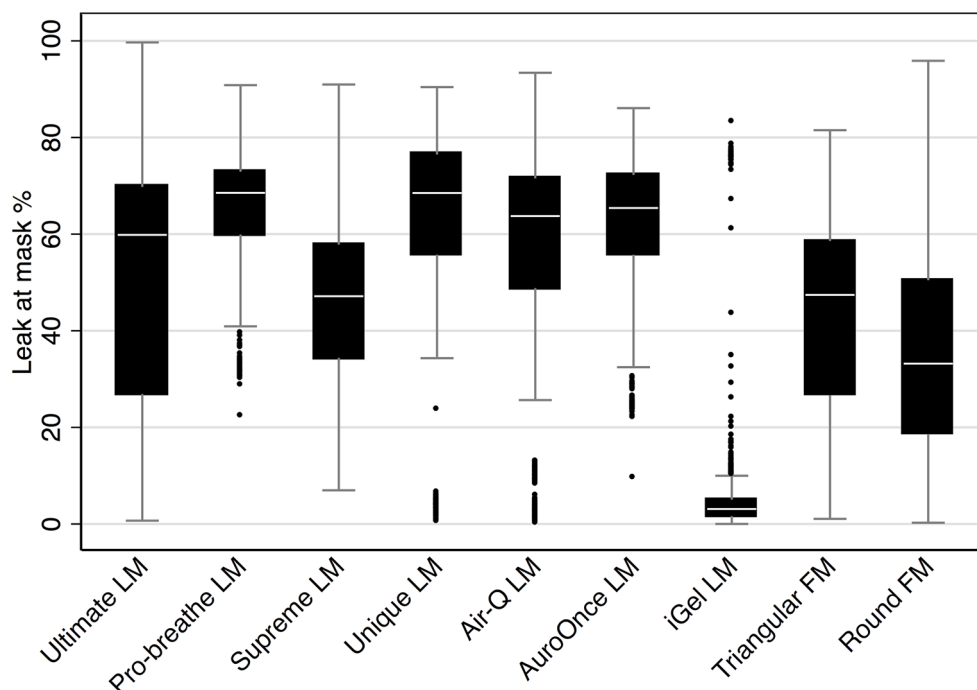# Original article

Figure 3 Box and whisker plot: test lung delivered positive end expiratory pressure. The median of 25–75th centile for box and whiskers is 1.5× IQR (above and below), which gives the upper and lower adjacent values. PEEP, positive end expiratory pressure.

Data analysis

The analysis using Stata V.14 examined the mean and SD for PIP, PEEP, TV and mask leak. Analysis of variance for repeated measures was used to determine differences between device types, with estimated means reported with their SEs (table 1) with p values-adjusted F test using Box's conservative epsilon. p Values of <0.05 were considered significant. Pairwise comparisons between groups were assessed with Bonferroni correction. LMA insertion attempts were grouped as frequency counts of subjects (no failure, partial or complete failure), and participant impression of ease of insertion (dichotomised to easy or difficult) per device was examined by Fisher's exact test.

RESULTS

There were 12 415 inflations recorded and analysed. The i-gel LMA consistently and statistically outperformed all other devices tested (table 1). Box and whisker plots demonstrate the median and spread of values for primary outcomes (PEEP in figure 3 and leak in figure 4).

Figure 4 Box and whisker plot: mask leak. The median of 25–75th centile for box and whiskers is 1.5× IQR (above and below), which gives the upper and lower adjacent values.

Primary outcomes

The mask leak observed with i-gel LMA of 5.7% was dramatically lower than all other devices observed, including face mask (ranging from 37.5% to 65.4%, p<0.001) (table 1 and figure 4). The i-gel LMA provided the highest mean PIP of 28.9 cmH$_2$O compared with the other devices (ranging between 16.3 and 25.5 cmH$_2$O, p<0.001) (table 1). The i-gel delivered PEEP was closest to the set value of 5.0 cmH$_2$O, with others ranging from 0.6 cmH$_2$O to 3.6 cmH$_2$O (p<0.001) (table 1 and figure 3). Comparing the maximal average PIP for the four inflations targeting the highest peak inflation pressure possible, the i-gel LMA delivered a significantly higher mean value of 38.4 cmH$_2$O compared with other devices, ranging from 18.5 cmH$_2$O to 32 cmH$_2$O (p<0.001) (table 1).

Secondary outcomes

The i-gel LMA was regarded by all 40 (100%) subjects as easy to insert, which compared with the range of subjective ratings for the other LMA from 16 (40%) with AuroOnce LMA to 38 (95%) with the Supreme LMA. All subjects were able to insert the i-gel LMA the first time with minimal leak. All other LMAs had at least one first insertion failure and four LMA brands had two failures recorded for subjects ranging from 1 to 5 in total.

DISCUSSION

We believe this is the first comprehensive study in a dedicated LMA manikin model looking at the performance of several brands (n>3) of size 1 single-use LMA devices compared with face mask PPV. The laryngeal structure of the AirSim Baby manikin head is modelled on CT scans of anatomically normal infants aged 2 months (TruCorp, personal communication). Anatomically correct laryngeal structure is vital to train LMA use and allow assessment of LMA cushion leak. During the design phase, the RFM pneumotach positioned between the SIB and the LMA tube connector frequently indicated 100% leak with some brands of LMA when the test lung was clearly moving. Thus, a single point pneumotach measurement at the delivery device (SIB) may not account for differing leak characteristics during inspiratory and expiratory flow past the LMA seal. A second pneumotach sited at the test lung was required to assess system leak relative to TV (figure 2).

The superiority of the i-gel LMA to deliver ventilation compared with the other LMAs and face mask tested was unexpected given the group of experienced neonatal resuscitators. The ease of use of the i-gel LMA coupled with no insertion failures indicates the physical structure of the LMA does significantly influence performance at least in this manikin model. Many of the LMAs tested in our study were rated by the users as difficult to insert. This paralleled the higher rates of partial or complete failure to establish a useable laryngeal seal. The provision of an effective PEEP may well be desirable in preterm infants, with some degree of surfactant deficiency and abnormally low lung compliance. Intubation with endotracheal tubes is frequently associated with many adverse physiological changes including hypoxia, bradycardia and raised blood pressure (cerebral and arterial).[17] The use of LMAs to provide PPV may mitigate some of these unwanted and potentially harmful physiological effects associated with intubation.

Recent interest in delivering surfactant via LMA[18–20] has led to its experimental use below ILCOR recommended weight (>2000 g) or gestation (>34 weeks)[1] in multicentre RCT in the USA, with results awaited.[21] The LMA used in this multicentre RCT, the Unique LMA in our study, did not provide adequate

PEEP (2.3 cmH$_2$O) and had a lower mean and maximal achievable PIP of 20.7 and 25 cmH$_2$O compared with the i-gel LMA mean and maximal achievable PIP of 28.9 and 38.4 cmH$_2$O. Inability to achieve PIP higher than 24 cmH$_2$O with size 1 LMAs has been reported in other manikin studies.[22] This may be relevant as preterm infants who are surfactant-deficient may require rapid escalation of peak inflation pressures and adequate PEEP during resuscitation.[3]

Limitations are as with any manikin study in generalising device performance to human subjects. The AirSim Baby JR1001 (TruCorp) is designed for training airway management in infants 0–6 months of age. Our results may not be generalisable to other brands of LMA not tested in particular reusable devices or use with other brands of manikin. Visualisation of test lung movement by participants may not be the same as assessment of chest wall movement in a human or a full body newborn infant manikin.

Results in this dedicated manikin model concur with findings of recent two brand LMA prospective studies comparing i-gel conducted in infants and children undergoing elective surgery.[10–12 23] However one study by Drake-Brockman et al[24] found the PRO-Breathe LMA superior to i-gel LMA in a wide age range of children (0–16 years) with LMA size range of 1.5–3. A study by Lee et al[25] in a lower age range of infants (10 months–5 years) found the Classic LMA had similar leak pressure to the i-gel LMA but longer insertion time.

A further important aspect not explored in this study nor in other studies we are aware of is the use of a flow-dependent t-piece resuscitator with an LMA device. T-piece resuscitators, in particular the Neopuff (NPR), have become a very popular device to resuscitate preterm newborns and infants with the provision of PEEP.[1 26 27] Previous work has shown the NPR is slower to adjust up the PIP during resuscitation than with SIBs.[16 28 29] The ability to quickly adjust the inflation pressure to respond to inadequate clinical response most likely due to mask leak may be important with the use of LMA devices given the variance of ventilation performance seen in this study. The clear superiority of the i-gel LMA with dramatically less leak and stable PEEP levels, compared with the traditional face mask SIB in this study, was unexpected. In our view, this warrants human studies to confirm this finding. Our group is currently examining the comparative performance of SIB and NPR with LMA devices, and beginning human infant studies to compare leak with face mask and i-gel LMA in the resuscitation of moderately preterm infants.

The emergency use of LMA devices during resuscitation of newborn infants is a time-critical procedure. We did not examine the additional time required for manual cuff inflation. We speculate devices that do not require manual inflation (i-gel with a solid gel cushion and the auto inflate cuff with the air-Qsp LMA) may allow shorter time to PPV and be simpler to use overall. This may be important with less experienced LMA users in the clinical setting of delivery suite or home birth.

CONCLUSION

This study has shown dramatic performance differences in delivered ventilation, mask leak and ease of use among seven different brands of LMA tested. PPV with the i-gel LMA with a solid gel laryngeal cushion had superior performance characteristics than the triangular or round face mask. This coupled with no partial or complete insertion failures and ease of use suggests i-gel LMA may have an expanded role with newborn resuscitation as a primary resuscitation device.

Original article

REFERENCES

1 Perlman JM, Wyllie J, Kattwinkel J, *et al*. Part 7: Neonatal Resuscitation: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. *Circulation* 2015;132(Suppl 1):S204–S41.

2 Heard AM, Green RJ, Eakins P. The formulation and introduction of a 'can't intubate, can't ventilate' algorithm into clinical practice. *Anaesthesia* 2009;64:601–8.

3 Weiner GM, ed. *Textbook of neonatal resuscitation (NRP)*. 7th edn. Illinois, USA: American Academy of Pediatrics, 2016:p. 326.

4 Zanardo V, Simbi A, Micaglio M, *et al*. Laryngeal mask Airway for neonatal resuscitation in a developing country: evaluation of an educational intervention. neonatal LMA: an educational intervention in DRC. *BMC Health Serv Res* 2010;10:254.

5 Iain P, Neil J. Is an i-gel ® supraglottic airway useful for airway rescue in the community? *Br J Midwifery* 2014;22:338–42.

6 Gandini D, Brimacombe J. Manikin training for neonatal resuscitation with the laryngeal mask airway. *Paediatr Anaesth* 2004;14:493–4.

7 Grein AJ, Weiner GM. Laryngeal mask airway versus bag-mask ventilation or endotracheal intubation for neonatal resuscitation. *Cochrane Database Syst Rev* 2005;2:CD003314.

8 Trevisanuto D, Cavallin F, Nguyen LN, *et al*. Supreme Laryngeal Mask Airway versus Face Mask during Neonatal Resuscitation: a Randomized Controlled Trial. *J Pediatr* 2015;167:286–91.

9 Zhu XY, Lin BC, Zhang QS, *et al*. A prospective evaluation of the efficacy of the laryngeal mask airway during neonatal resuscitation. *Resuscitation* 2011;82:1405–9.

10 Sanket B, Ramavakoda CY, Nishtala MR, *et al*. Comparison of Second-Generation Supraglottic Airway Devices (i-gel versus LMA ProSeal) During elective surgery in Children. *Aana J* 2015;83:275–80.

11 Kayhan GE, Begec Z, Sanli M, *et al*. Performance of size 1 I-gel compared with size 1 ProSeal laryngeal mask in anesthetized infants and neonates. *ScientificWorldJournal* 2015;2015:1–6.

12 Kim MS, Lee JH, Han SW, *et al*. A randomized comparison of the i-gel with the self-pressurized air-Q intubating laryngeal airway in children. *Paediatr Anaesth* 2015;25:405–12.

13 Micaglio M, Trevisanuto D, Doglioni N, *et al*. The size 1 LMA-ProSeal: comparison with the LMA-Classic during pressure controlled ventilation in a neonatal intubation manikin. *Resuscitation* 2007;72:124–7.

14 Komasawa N, Ueki R, Yamamoto N, *et al*. Comparison of air-Q(®) and Soft Seal(®) laryngeal mask for airway management by novice doctors during infant chest compression: a manikin study. *Resuscitation* 2012;83:365–8.

15 Trevisanuto D, Parotto M, Doglioni N, *et al*. The Supreme Laryngeal Mask Airway™ (LMA): a new neonatal supraglottic device: comparison with classic and ProSeal LMA in a manikin. *Resuscitation* 2012;83:97–100.

16 Bennett S, Finer NN, Rich W, *et al*. A comparison of three neonatal resuscitation devices. *Resuscitation* 2005;67:113–8.

17 Maheshwari R, Tracy M, Badawi N, *et al*. Neonatal endotracheal intubation: how to make it more baby friendly. *J Paediatr Child Health* 2016;52:480–6.

18 Ali E, Abdel Wahed M, Alsalami Z, *et al*. New modalities to deliver surfactant in premature infants: a systematic review and meta-analysis. *J Matern Fetal Neonatal Med* 2016;29:3519–24.

19 Sadeghnia A, Tanhaei M, Mohammadizadeh M, *et al*. A comparison of surfactant administration through i-gel and ET-tube in the treatment of respiratory distress syndrome in newborns weighing more than 2000 grams. *Adv Biomed Res* 2014;3:160.

20 Trevisanuto D, Marchetto L. Minimally invasive approaches for surfactant administration. *Acta Biomed* 2013;84(Suppl 1):28–31.

21 Wanous AA, Wey A, Rudser KD, *et al*. Feasibility of laryngeal mask Airway Device Placement in Neonates. *Neonatology* 2017;111:222–7.

22 Micaglio M, Doglioni N, Parotto M, *et al*. Training for neonatal resuscitation with the laryngeal mask airway: a comparison of the LMA-ProSeal and the LMA-Classic in an airway management manikin. *Paediatr Anaesth* 2006;16:1028–31.

23 Maitra S, Baidya DK, Bhattacharjee S, *et al*. Evaluation of i-gel(™) airway in children: a meta-analysis. *Paediatr Anaesth* 2014;24:1072–9.

24 Drake-Brockman TF, Ledowski T, Hegarty M, *et al*. A comparison of the i-gel(™) and the PRO-Breathe(®) laryngeal mask during pressure support ventilation in children. *Anaesthesia* 2015;70:1412–7.

25 Lee JR, Kim MS, Kim JT, *et al*. A randomised trial comparing the i-gel (TM) with the LMA Classic (TM) in children. *Anaesthesia* 2012;67:606–11.

26 Hinder M, Jani P, Priyadarshi A, *et al*. Neopuff T-piece resuscitator: does device design affect delivered ventilation? *Arch Dis Child Fetal Neonatal Ed* 2017;102:F220–F224.

27 Tracy M, Maheshwari R, Shah D, *et al*. Can Ambu self-inflating bag and Neopuff infant resuscitator provide adequate and safe manual inflations for infants up to 10 kg weight? *Arch Dis Child Fetal Neonatal Ed* 2017;102:F333–F338.

28 Hartung JC, Dold SK, Thio M, *et al*. Time to adjust to changes in ventilation settings varies significantly between different T-piece resuscitators, self-inflating bags, and manometer equipped self-inflating bags. *Am J Perinatol* 2014;31:505–12.

29 Kattwinkel J, Stewart C, Walsh B, *et al*. Responding to compliance changes in a lung model during manual ventilation: perhaps volume, rather than pressure, should be displayed. *Pediatrics* 2009;123:e465–e470.

FN

How do different brands of size 1 laryngeal mask airway compare with face mask ventilation in a dedicated laryngeal mask airway teaching manikin?

Mark Brian Tracy, Archana Priyadarshi, Dimple Goel, Krista Lowe, Jacqueline Huvanandana and Murray Hinder

Arch Dis Child Fetal Neonatal Ed published online August 11, 2017

Updated information and services can be found at:
http://fn.bmj.com/content/early/2017/08/11/archdischild-2017-312766

These include:

References   This article cites 28 articles, 4 of which you can access for free at:
**http://fn.bmj.com/content/early/2017/08/11/archdischild-2017-312766
#BIBL**

Email alerting service   Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

Appendix V

Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features

I contributed to the signal processing of and feature extraction of the normalised frequency features.

ORIGINAL ARTICLE

CrossMark

Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features

Nima Befrui[1] · Jens Elsner[2] · Achim Flesser[3] · Jacqueline Huvanandana[2] · Oussama Jarrousse[1,2] · Tuan Nam Le[1] · Marcus Müller[2] · Walther H. W. Schulze[1,2,4] ⓘ · Stefan Taing[2] · Simon Weidert[1]

Abstract

Vibroarthrography is a radiation-free and inexpensive method of assessing the condition of knee cartilage damage during extension-flexion movements. Acoustic sensors were placed on the patella and medial tibial plateau (two accelerometers) as well as on the lateral tibial plateau (a piezoelectric disk) to measure the structure-borne noise in 59 asymptomatic knees and 40 knees with osteoarthritis. After semi-automatic segmentation of the acoustic signals, frequency features were generated for the extension as well as the flexion phase. We propose simple and robust features based on relative high-frequency components. The normalized nature of these frequency features makes them insusceptible to influences on the signal gain, such as attenuation by fat tissue and variance in acoustic coupling. We analyzed their ability to serve as classification features for detection of knee osteoarthritis, including the effect of normalization and the effect of combining frequency features of all three sensors. The features permitted a distinction between asymptomatic and non-healthy knees. Using machine learning with a linear support vector machine, a classification specificity of approximately 0.8 at a sensitivity of 0.75 could be achieved. This classification performance is comparable to existing diagnostic tests and hence qualifies vibroarthrography as an additional diagnostic tool.

✉ Jens Elsner
  je@munich-innovation.com

  Walther H. W. Schulze
  ws@evolunis.com

  Simon Weidert
  simon.weidert@med.uni-muenchen.de

[1] Trauma Surgery Department, University Hospital of Munich, Munich, Germany

[2] Munich Innovation Labs, Grünwald, Germany

[3] CPE GmbH, Willich, Germany

[4] Evolunis UG (haftungsbeschränkt), Knesebeck, Germany

🙋 Springer

1 Introduction

Osteoarthritis (OA) is a clinical syndrome characterized by the progressive degeneration of articular cartilage and potentially involving the entire joint, including the synovium and underlying bone. It is among the most common causes of pain and disability in the middle-aged and elderly demographic and the leading cause of impairments in activities of daily life [8, 21]. Although joint biology and the pathogenetic mechanisms that lead to OA are not sufficiently understood, evidence suggests that age-related structural alteration in articular cartilage, as well as abnormal joint anatomy and pathological reaction patterns of chondrocytes to injury, contributes to the degeneration of articular cartilage and subsequent joint reaction [20, 48].

1.1 Early detection of OA

The detection of cartilage defects at an early stage remains an area of interest: X-ray imaging of the knee joint only provides an approximation of the articular cartilage as soft

tissue is not depicted, and visible joint narrowing and osteophytes typically manifest in the later stages of OA. Although knee arthroscopy is the current gold standard method for detection of degenerative changes in articular cartilage, its use for OA diagnosis has been abandoned due to its invasive nature and inefficacy as a therapeutic tool [7, 22]. Recent studies have shown that magnetic resonance imaging (MRI), although a commonly used diagnostic tool for identification of cartilage damage, provides moderate sensitivity with high specificity [32]. Moreover, these methods are unable to characterize the functionality of articular cartilage, which could be detected during dynamic knee movements, where roughness, softening, degeneration, or the state of lubrication can be assessed [17, 27, 38]. Table 1 provides an overview of the advantages and disadvantages of different diagnostic methods used for articular cartilage degeneration detection [18, 36, 42, 57]. A non-invasive routine diagnostic method for early detection of cartilage damage is crucially important for early optimizing treatment, clinical management, and delaying the progression of joint degeneration.

1.2 Vibroarthrography

Vibroarthrography (VAG) is emerging as a candidate for non-invasive, cost-efficient, and dynamic detection of knee joint disorders.

It is expected that pathological conditions in the knee joint, such as the degeneration of articular cartilage, will correspond to variations in VAG signals, these being the vibrations or acoustic signals emitted from the knee joint during active movement [24, 26, 31, 44]. Current research has demonstrated the potential of VAG as a method for distinguishing between normal and pathological knees with progressive degeneration of articular cartilage. Vibrations generated in the joint during active movement of the knee can offer insight into the articular cartilage and its state of

roughness, degeneration, or lubrication [52]. An elaborate and comprehensive introduction to the technology can be found in [62].

The idea of diagnosing joint pathologies through the evaluation of acoustic joint emissions dates back to 1885 when Heuter described the localisation of loose bodies in the knee joint by use of a stethoscope [9]. Similar stethoscope evaluations of the joint pathologies followed (Blodgett in 1902 [5]; Bircher in 1913 [4]; Walters in 1929 [58]). In the following years, doctors transitioned from evaluating pure sound intensity to analyzing sound frequency, wavelength, and quality by use of microphones, thereby eliminating subjective interpretation (Erb in 1933 [15]; Fischer and Johnson in 1961 [16]).

During the 1980s to 1990s, the use of vibration sensors became prevalent and sophisticated signal analysis was now possible. During this time, several acoustic phenomena linked to physiological biomechanical motion as well as pathological joint degeneration were discovered. Moussavi et al. found that signals obtained from symptomatic knees exhibit an overall higher energy [33]. Observation of the frequency domain also revealed distinctions in the activity range, where normal signals generally expressed characteristics between 0–100 Hz. This range was considerably lower than that of pathological signals which extended beyond 200 Hz [33]. This was also supported by the findings by McCoy et al., who found signals associated with degeneration of articular cartilage to be "bizarre" and "irregular," predominantly in the range of 300–600 Hz [31].

Active knee extension and flexion has been shown to produce certain acoustic characteristics, including the patellar click [31]. This is an isolated, transient signal found to be maximal directly over the patella. McCoy et al. observed that the incidence of this click was related to the cycle speed of the knee. Faster cycles were shown to produce an increased amplitude and occurrence of the patellar clicks. The angle at which these clicks occurred ranged from 19

Table 1 Qualitative comparison of different clinical diagnosis methods

	CT	MRI	Ultrasound	Arthroscopy	VAG
Operating costs	Low	Very high	Very low	Medium	Very low
Applicability	Limited	Limited	Broad	Medium	Broad
Investment costs	Medium	High	Low	Low	Very low
Inter-observer variability	Low	Medium	Very high	High	Low
Duration of examination	2 min	30 min	10 min	60 min	5–10 min
Invasivity	Ionizing radiation	None	None	Invasive, operative risk	None
Documentation capability	Good	Good	Very poor	Good	Good
Output data	Image	Image	Image	Image	Sound
Sensitivity to artifacts	High, i.a. metal	High, i.a. metal	Very high, manifold	Almost no artifacts	High
Measurement condition	Static	Static	Static	Static	Dynamic

to 55º, approximately corresponding to the position at which the iliotibial tract crosses the lateral femoral condyle. Speculation remains as to the origin of the acoustic signal, although the most probable cause is the changing patellar orientation and its area of contact with adjacent facets [31]. These clicks manifested as sharp bursts in the time domain as well as wide-band activity in the frequency domain [33]. Their use as a distinguishing feature in knee injury diagnoses is also uncertain due to the prevalence of these features in normal joints [64]. Another acoustic characteristic is that of physiological patellofemoral crepitus (PPC), a manifestation of patellar slip-stick friction produced when the extension-flexion cycle speed of the knee joint is less than 5º per second [3, 23]. PPC and its quality, amplitude, and distribution in the movement cycle can be useful indicators for cartilage integrity, where increased cartilage deterioration is associated with higher amplitudes and greater incidence of the crepitus signal [23, 31]. Schindler et al. also observed a signal associated with the synovial plicae (see Table 2), a fold in the synovial lining of the knee joint capsule which is often found in the knee joint but is mostly asymptomatic [49]. It may produce acoustic characteristics during impingement on the femoral condyles throughout the extension-flexion cycles [31]. In patients with meniscal injuries, McCoy et al. also identified the meniscal signals [31], which were characterized by a large displacement, being largest on the affected side and appearing at approximately the same angle in each extension-flexion cycle. Table 2 shows a summary of acoustic characteristics of the knee joint mentioned above in the literature review.

In the 2000s, several advanced methods of signal analysis were proposed with encouraging results: Krishnan et al. used adaptive time-frequency analysis to screen for knee cartilage pathologies [26] and Umapathy et al. described the use of local discriminant bases to classify their test subjects [55]. Further signal analysis methods include probability density functions using Parzen windows [46], radial basis functions [44] and power spectral analysis (Rangayyan et al., 2013 [43]), and high frequency acoustic emissions (Prior, Mascari [41]), as well as bivariate feature distribution estimation, maximal posterior probability decision (Wu et al., 2013 [61]), and the use of entropy or envelope amplitude measures [63]. Various techniques have been developed for processing features extracted from VAG signals in order to characterize normal and abnormal knees. These techniques include linear prediction [33, 54], adaptive muscle contraction interference cancelation [65], adaptive time-frequency analysis [26], wavelet decomposition [55], and statistical parameters [44, 45]. Using these signal analysis techniques, the authors achieved impressive results in the distinction of normal and abnormal joints with classification accuracies ranging from 77.5% [26] and 78.7% [55] to 91.4% [24], and areas under the curve (AUCs) of receiver operating characteristics in the ranges of 0.82 [44], 0.91 [61], and 0.92 [43, 45] to 0.96 and even 1.0 [43].

Building on the promising results reported in the prior art, the purpose of this work was to develop a measurement system and procedure that are both highly sensitive and suitable to the clinical environment. The intention was to verify and improve upon the classification accuracy of features obtained from healthy knee joints and joints affected by chondromalacia and osteoarthrosis by analyzing multiple sensor signals that were gained from a larger subject test group. We further chose to focus on earlier stages of cartilage degeneration by only including patients presenting MRI scans of their knee joints, whose beginning pathological joint lesions may not have been visible in X-ray, as to test the efficiency of VAG as a screening tool for those patients who may still benefit from treatment to slow down the cartilage degeneration.

This work presents a classification approach that uses linear support vector machines together with simple yet normalized and therefore amplitude-robust features that are based on relative high-frequency components.

2 Methods

For detection of osteoarthritis, we developed a setup of acoustic sensors (see Section 2.1). After semi-automatic segmentation, we calculated simple and robust feature vectors (see Section 2.2) which we assessed in their ability to distinguish between healthy and non-healthy knees. With the feature vectors, we performed a classification study (see Section 2.3) using support vector machines.

Table 2 Acoustic characteristics of the knee joint

Name	Source
Patellofemoral crepitus	Patella
Patellar click	Patella, medial, and lateral femoral condyle (maximum over the patella)
Lateral band signal	Lateral femoral condyle
Meniscal signal	Patella, medial, and lateral femoral condyle (maximum on the affected side)
Synovial plica signal	Medial and lateral femoral condyle

2.1 Materials

To obtain a clinical database of vibroarthrography signals, the measurement setup was installed to study the recruited cohort of patients and healthy volunteers.
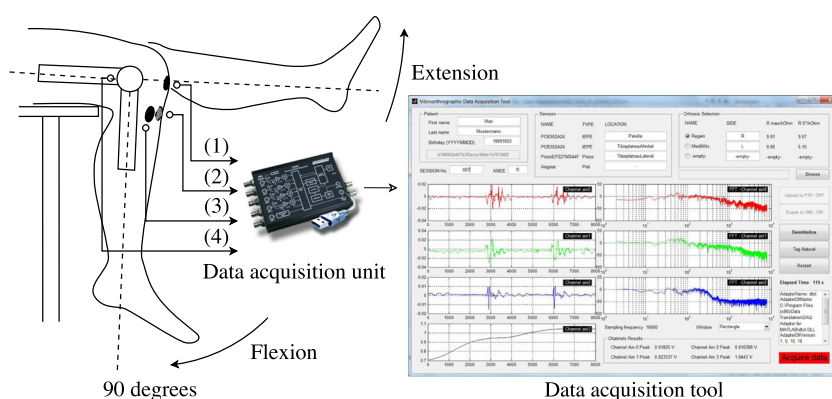
2.1.1 Measurement system

The measurement system was developed to acquire mechanical vibrations emitted from the human knee joint, as well as to track the relative joint angle during extension and flexion movement. Two miniature accelerometers and one piezoelectric disk were attached to each knee joint by using two mounting techniques: (1) direct adhesive mounting where sensors were directly fixed to the surface of the skin using medical glue as in Figs. 1b and (2) mounting via electrocardiography (ECG) pads where sensors were mounted to the top side of a custom-built adapter, whereas the bottom side of the adapter contains a snap button that

connects to common ECG pads as in Fig. 1c. This mounting technique proved to be more convenient in clinical application, providing easier and faster sensor mounting, as well as higher sensor stability during the measurement.

The technical details of the two sensor types are given in Table 3. The sensors were attached to the skin over the patella, the medial tibial plateau, and the lateral tibial plateau (see Fig. 1b, c). These specific anatomical locations offer an optimal contact area, which is the closest to the bone surface, minimizing the influence of skin and subchondral soft tissue on the propagating acoustic signal. In order to stabilize the range of motion of the knee joint during measurements, various sizes of knee orthoses were developed to account for inter-individual variations in leg size. Each of these orthosis prototypes incorporates an angle potentiometer that is aligned with the axis of rotation of the orthosis and located on the lateral side of the orthosis. The VAG signals and the relative knee angular signal are simultaneously acquired at 16 kHz sampling frequency and

Fig. 1 **a** Schematic diagram of the VAG measurement system. Channels (1) to (3) are used to record the VAG knee signals, while channel (4) records the corresponding angular signal. **b**, **c** Vibration sensors are attached to the skin of a volunteer wearing the developed knee orthosis: two accelerometers (channels (1) and (2)) are placed over the knee patella and over the medial tibial condyle, a piezoelectric disk (channel (3)) over the lateral tibial condyle. A potentiometer is located at the center of the orthosis rotation to track the angle of the active movement (channel (4))
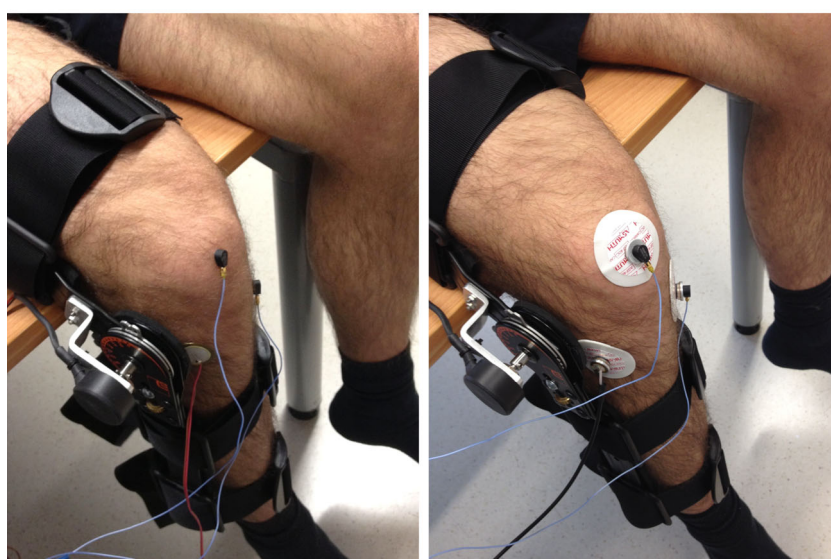
Table 3 Technical details of hardware components used for the measurement system

Type	Main specification
Accelerometer[1]	Frequency range 1–8000 Hz ($\pm$ 5%), sensitivity 10.09 mV/(m/s^2), measurement range 50 m/s^2, resonant frequency 38,300 Hz, constant current excitation 2–20 mA
Piezoelectric disk[2]	Resonant frequency 4400 Hz, impedance 300Ω.
Potentiometer[3]	Resistance 10kΩ, linearity $\pm$ 5%, tolerance 3%, mechanical angle 320 $\pm$ 5°.
Data acquisition[4]	Four single-ended simultaneous channels with 24-bit resolution, support for IEPE inputs with current source 4 mA, high-pass filter 0.5 Hz

[1] Model 352A24, PCB Piezotronics, Inc., USA

[2] Model EPZ-27MS44F, Elektrotechnik Karl-Heinz Mauz GmbH, Germany

[3] Model SW22E-10K, ETI Systems, Inc., USA

[4] Model DT9837, Data Translation, Inc., USA

24 bits/sample via a four-channel data acquisition unit. The high sampling frequency was chosen in line with the frequency range of the accelerometer, as the upper limit of the actual VAG signals was not known at the beginning of the study. The recordings of the four channels were processed using a MATLAB-based signal acquisition tool.

2.1.2 Measurement protocol

The measurement protocol comprises a calibration phase and an active movement phase. During the calibration phase, the leg of the volunteer rests at an angle of 90° in the knee joint for 3–5 s, then the leg is extended towards the 0° knee angle and held at the position of the greatest/maximal extension the patient is able to reach for 3–5 s. During the subsequent active movement phase, the volunteer performs 15 repeated leg extension-flexion cycles with pauses of approximately 3–5 s between cycles. One cycle consists of extending the leg from the 90° knee angle to fully extended position which is then followed by flexion back to the neutral position of 90° joint angle (see Fig. 1a). Volunteers were asked to complete each extension-flexion cycle in 3–5 s. For all volunteers, two sets of measurements for each knee were recorded in a measurement session.

2.1.3 Clinical database

The clinical database used in the present work consists of data from 30 healthy subjects and 39 patients with articular cartilage damage at different stages of the disease; see Table 4.

All patients selected for the study presented MRI scans no more than 12 months old. The MRI findings, which

Table 4 Clinical database

	Healthy	Patients
Female	14	24
Male	16	15
Age	26.9 $\pm$ 4.18	55.2 $\pm$ 13.5
BMI[1]	22.6 $\pm$ 2.80	27.4 $\pm$ 6.34
OKS[2]	47.7 $\pm$ 1.14	27.5 $\pm$ 11.32

[1] Average body mass index, where available

[2] Average Oxford Knee Score, where available

were selected as the reference diagnostic for categorizing the knee conditions, were subsequently assessed by a radiologist, based on the Outerbridge classification system [34, 35] and additional parameters such as location and size of the findings as well as other pathological findings within the joint were determined. Table 5 summarizes the standardized patients' diagnoses and Fig. 2 reveals how scores from the patient questionnaire correlate with the diagnoses. Additional metadata including their age, gender, and body mass index (BMI) was also obtained. Furthermore, the influence of other demographic and clinical factors, such as occupation, sport activities, and medical history associated with pathological conditions, was investigated. The Oxford Knee Score (OKS), a standardized joint-specific patient questionnaire consisting of 12 questions covering function and pain associated with the knee, was collected for all volunteers [13, 14]. The healthy volunteers were characterized by being below the age of 36 and having no knee-related symptoms such as pain or instability (OKS > 45), no history of major knee joint trauma or pathologies such as Osgood-Schlatter disease, and no previous invasive clinical treatment of the knee joint. In doing so, the prevalence of a healthy joint condition could be assumed and unnecessary imaging could be avoided.

An evaluation of the MRI-based Outerbridge classification against the outcome of the patient questionnaire in Fig. 3 clearly revealed that patients were not always aware of their chondromalacia. In our study, 36% of the OA stage 4 or manifest osteoarthrosis cases actually classified themselves as having an OKS of $\geq$ 30 out of 48. According to

Table 5 Diagnoses of knee diseases

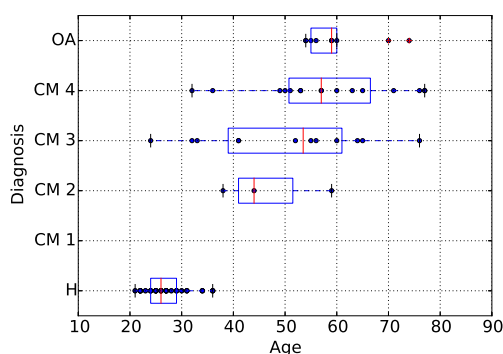| Diagnosis	Number of cases (knees)
Chondromalacia grade II	3
Chondromalacia grade III	12
Chondromalacia grade IV	16
Manifest OA	9

Fig. 2 Evaluation of the MRI-based Outerbridge classification against age. Each sample represents one knee

[60, Chapt. 25], such scores may indicate "mild to moderate arthritis." Twelve percent even had an OKS of 40 and above, which may indicate "satisfactory joint function." In some patients, the chondromalacia was not associated with a positive OKS result despite having clinical symptoms that led to the initial orthopedic evaluation and imaging in the first place. The mismatches showcase that many OA patients experience no clear subjective symptoms or impairments, revealing the need for patient screening with easy-to-use and inexpensive technology.

2.2 Signal processing

From the VAG signals, we extracted extension and flexion cycles using semi-automatic segmentation, which were used to calculate knee-specific feature vectors for classification.

2.2.1 Segmentation

As described in the previous sections, each VAG signal consists of a calibration phase followed by 15 leg extension-flexion cycles. The segmentation of the extension-flexion cycles was conducted using the relative angular signal
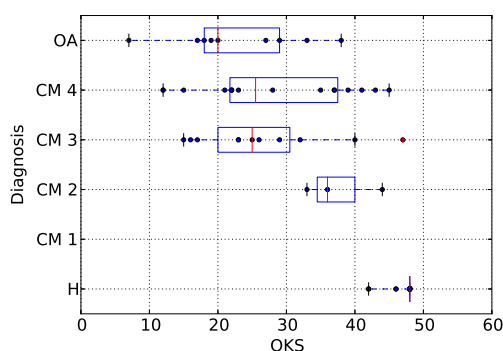

Fig. 3 Evaluation of the MRI-based Outerbridge classification against the outcome of the patient questionnaire (Oxford Knee Score). Each sample represents one knee

acquired from the embedded angle potentiometer. This relative angular signal was negatively proportional to the knee angle. Therefore, peaks in the signal correspond to full knee extension. The analog signal was low-pass filtered (RC filter, cut-off frequency 4.5 Hz) to suppress signals from small-scale muscular movement or tremor and to ensure that only large-scale extension and flexion movements are evaluated.

A local maxima search was executed to locate all possible maxima $t_{max,i}$. Then, a local minima search directly to the left of a certain maximum $t_{max,i}$ defined the start $t_{s,i}$, and another local minima search directly to its right defined the end $t_{e,i}$ of an extension-flexion cycle, where i is the cycle index. These points determined three kinds of segments of each cycle: the extension segments $[t_{s,i}, \ldots, t_{exte,i}]$, the flexion segments $[t_{fles,i}, \ldots, t_{e,i}]$, and the cycle segments $[t_{s,i}, \ldots, t_{e,i}]$, which are illustrated in Fig. 4. $t_{exte,i}$ and $t_{fles,i}$ were initially set to $t_{max,i}$. After this basic annotation, the extension and flexion segments were then shortened to exclude the resting phase around $t_{max,i}$ with angular velocity below an experimentally trained threshold, resulting in $t_{exte,i} \leq t_{max,i} \leq t_{fles,i}$. After the automatic segmentation steps, unsuccessful segmentations were corrected manually in a graphical user interface and the VAG knee signals acquired from the three sensors were segmented according to the indices $t_{s,i}, t_{exte,i}$ for extension and $t_{fles,i}, t_{e,i}$ for flexion phases as shown in Fig. 4.

2.2.2 Censoring

Artifacts originating from incidents such as mechanical pull on the sensor cable during measurement were identified using two main criteria. The cycles were discarded if they contained values at the saturation limit of the A/D converter, while those of artifact-free VAG signals were observed to be substantially below this limit. The second criterion entailed a comparison of the variances; a cycle was excluded if its variance exceed twice that of the overall segment. Cycles shorter than a second were also excluded as these did not adhere to the specified measurement protocol.

2.2.3 Spectrum normalization and feature extraction

In this work, we studied either exclusively the extension cycles or the flexion cycles of the VAG measurements. To represent the characteristics of the measurements, the power spectra of all extension or flexion cycles in a measurement were calculated, normalized, and then averaged. To facilitate averaging, power spectra of 3.9 Hz/bin resolution were estimated using Welch's method [59] using a Hann window with a segment length of 4096 samples and 50 percent overlap. The estimated power spectra of each cycle $\hat{S}_{xx}[f]_i$ were first normalized so that their power
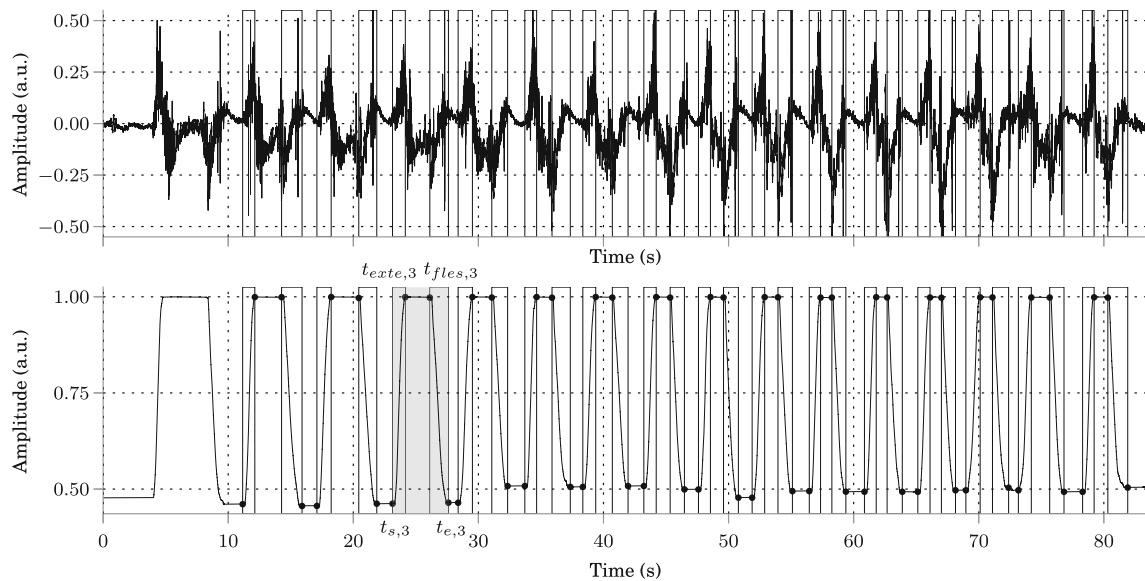
Fig. 4 VAG raw signal of the patella sensor (top; normalized) and angular raw signal (bottom; normalized, negatively proportional to the knee angle) of a measurement, which includes the calibration phase followed by an active movement phase. The vertical lines represent extension and flexion cycles determined by local maxima and minima of the angular signal. The segmentation procedure for the ith cycle with $i = 3$ (gray shadow) is here highlighted

between 10 and 8000 Hz was one and then averaged. This normalization was found to be essential, pushing our results from an area under the ROC curve (AUC) of 0.66 for training and testing on the full population (full pop.) and an AUC of 0.57 for fivefold cross-validation (fivefold CV) to AUCs of 0.89 and 0.85, respectively (for the latter results, see Fig. 11).

Similar to [28], the partial sum of the power spectrum components between a lower frequency limit l and an upper limit u was then defined as

$$R_{l,u} = \sum_{f=l}^{u} \hat{S}_{xx,M}[f], \tag{1}$$

where f runs over the bins contributing to the power contained in the signal between l and u.

On top of providing an amplitude-independent feature, normalization on cycle level ensures that no cycle can contribute disproportionally to the spectrum of the measurement, which confines the impact of strong signal artifacts.

The VAG signals are caused by the movement of the leg, and the velocity of that movement may correlate strongly with the pathological condition of the knee. This led to the suspicion that VAG frequency characteristics were also influenced by the rotational speed of the leg rather than the pathological condition alone. We therefore explored the possibility of classification based on the movement duration, i.e., the duration of the leg movement between flexion and extension or extension and flexion. We considered the movement duration to be a robust substitute

for computing the average absolute angular velocity, which is susceptible to noise as a temporal derivative signal. The movement durations of the extension or flexion cycles in a measurement were averaged to form the mean movement duration feature D_c.

2.2.4 Knee-specific feature vectors for classification

The combination of features at the position of a sensor was structured as a feature vector and stored for each measurement. Three sensors were present on the knees during the measurements; see Fig. 1a. When information from multiple sensors was taken into consideration for classification, their feature vectors were concatenated to form a vector of higher dimension.

Then, feature vectors from multiple measurements of the same knee were averaged. In line with the protocol, two measurements were recorded for each knee—except for five cases where only one measurement was acquired and three cases where three measurements were recorded. The resulting feature vector that was used for classification purposes was hence knee-specific; i.e., healthy volunteers and patients were represented by up to two cases of diagnosed knees each in the classification study. Case data was collected for both knees of the healthy volunteers except for one (30 healthy volunteers, 59 knees in the classification study). Only for one patient, the pathological condition of both knees was known (39 patients, 40 knees in the study).

2.3 Classification

The knee-specific feature vectors were used for classification experiments with a linear support vector machine (SVM) [12, 56], a machine learning algorithm.

2.3.1 Support vector machine

The aim of the SVM approach, which is a supervised learning technique based on the theory of structural risk minimization, is to discriminate d-dimensional data $x_k \in \mathbb{R}^d$ into two classes $\{-1, 1\}$ using a hyperplane decision surface $w^T x + b = 0$. In the linearly separable case, the hyperplane is optimized such that the margin $\frac{2}{\|w\|}$ between the data points $w^T x_+ + b = 1$ and $w^T x_- + b = -1$ (support vectors) and the hyperplane is maximized, which is achieved by maximizing the following cost function $\forall k$:

$$\max_{w,b} \frac{2}{\|w\|} \text{ s.t. } \begin{cases} w^T x_k + b \geq 1 \text{ if } y_k = 1 \\ w^T x_k + b \leq -1 \text{ if } y_k = -1 \end{cases}, \quad (2)$$

where $y_k \in \{-1, 1\}$ assigns classes to the samples k in the training data x_k.

The advantages of SVMs are their high performances in practical applications, effectiveness in high dimensional data, robustness and flexibility in dealing with noisy features [6, 19, 50]. For linear SVMs, in case the data is not linearly separable, Eq. 2 has no solution. In this work, we therefore utilized the standard linear SVM classifier (C-support vector classification) as implemented in the *scikit-learn* modules for Python [6, 10, 12, 37], which solves a soft-margin variant of the classification problem that allows for mis-classifications and margin violations to happen. To achieve this, slack variables ζ_k and a regularization parameter C are introduced. In a dual problem of Eq. 2,

$$\min_{w,b} \|w\| \text{ s.t. } y_k \left(w^T x_k + b \right) \geq 1 \forall k, \quad (3)$$

ζ_k is introduced to allow for mis-classifications and margin violations to happen, and an additional cost term is introduced to penalize these, which is weighted with C:

$$\min_{w,b,\zeta_k} \|w\| + C \sum_k \zeta_k \text{ s.t. } y_k \left(w^T x_k + b \right) \geq 1 - \zeta_k \forall k. \quad (4)$$

As result of a training, the SVM produces a hyperplane decision surface $w^T x + b = 0$, which is then used to compute scores for test data samples that predict class membership. With a classifier that provides such class membership scores, it is possible to produce very sensitive or very specific classifications, depending on the threshold that decides on the class membership. One can naturally produce ideal true positive rates (TPRs) of 1 (ideal sensitivity) when this threshold is set to the lowest score in the dataset, and all test samples are classified positive. A false positive rate (FPR) of 0 (ideal specificity) is achieved

for a threshold at the maximum score in the dataset when all of the samples are classified negative. To produce curves of the receiver operating characteristics (ROC) such as Figs. 11 or 10, the discrimination threshold was varied across all possible thresholds, illustrating the performance of the classifier and the bandwidth of application scenarios. In this process, single samples may cause a step in the curve in either TPR or FPR.

Before processing the datasets, the knee-specific feature vectors were scaled to prevent features of great range to dominate those with smaller range in the SVM optimization problem and to avoid related numerical difficulties [19, Sect. 2.2]. In particular, the components in the feature vectors were centered to zero mean across the dataset, and each component was then standardized to have unit variance across the samples in the dataset.

An exhaustive grid-search procedure was performed in order to identify the optimal regularization parameter $C \in \{10^{-3}, 10^{-2}, \cdots, 10^3\}$ [19, Sect. 3.2], as no consensus exists on an analytical choice [11, Sect. 2]. Parameter C controls the trade-off between smoothness of the decision boundary and classification error on the training set, i.e., between false positives and false negatives. According to [11, 51], C can, e.g., be chosen in the range of the SVM output values, which are $\in \{0, 1\}$ in the present training and test dataset, justifying our initial search in $\{10^{-3}, 10^{-2.5}, \cdots, 10^3\}$. Optimization of C, however, led to ROCs with almost exactly the same areas under the curve as we obtained with $C = 1$. For this reason, C was set to 1 for the results we present in this work.

We weighted the regularization parameter to balance the classes in the training of the SVM as our dataset was biased towards healthy knees (59, against 40 patient knees). A heuristic balancing approach from the *scikit-learn* module was used for this purpose that weights C differently for each class c with a weight of $w_c = \frac{\text{average class size}}{\text{class size}}$, as proposed in [25]. Weights were computed from the class members as presented with each training set in the cross-validation.

By using the kernel trick, the concept of SVMs can be extended to perform non-linear classifications. In case the data x_k is not linearly separable in the input space, non-linear mapping functions $\varphi(x_k)$ can be used to transform it into a higher dimensional space where such separation is possible. Kernels are the inner product of the underlying transformation $k(x_k, x_l) = \langle \varphi(x_k), \varphi(x_l) \rangle$, which represents the data in the dual quadratic programming problem. However, SVMs with non-linear transformations are prone to over-fitting [2, 53]. To illustrate this behavior on the present data (see Fig. 12), we used the *scikit-learn* implementations of radial basis function kernels ($k(x_k, x_l) = \exp(-\gamma \|x_k - x_l\|^2)$) and polynomial kernels ($k(x_k, x_l) = (\gamma \langle \varphi(x_k), \varphi(x_l) \rangle + r)^d$). Parameters $\{\gamma \in \{10^{-5}, 10^{-4}, \cdots, 10^1\}\}$ and $\{d \in \{2, 3, 4\}, r \in \{10^{-2}, 10^{-1},$

10^0}} were optimized in a grid-search procedure. The regularization parameter C was set to 1.

The two non-linear SVMs were indeed found to facilitate an over-fitting on the present data (see Fig. 12). As they had only marginally better performance than linear SVMs in full population training, the focus of this study was put on the use of linear SVMs. In other words, the main underlying pathology-related phenomenon was assumed to be sufficiently linearly separable in the features space.

2.3.2 Features evaluation with SVMs

To assess the ability of different features to discriminate between patient knees and healthy knees, we plotted the histograms for the respective groups. To understand how well they would perform in a threshold-based classification scenario, we conducted experiments by fitting SVM decision hyperplanes on our training dataset, which were initially both trained and tested on the full population (full pop.); see Figs. 5 and 8.

Full training and testing has two advantages: it is an exhaustive experiment that does not rely on the choice of a random cross-validation subset, and it further produces scores that are comparable across the tested samples, which is important when making ROC curves. However, it is not suitable to estimate how well the machine learning would perform on new data; i.e., it is not suitable to estimate how well the performance figures can be generalized. To access this capability for the proposed SVM, cross-validation was used. A natural choice would be leave-one-out cross-validation. However, if only one sample is excluded from each training, the models that are trained to make predictions are very similar, and the outcome of the

cross-validation experiment is then not reliable (if multiple such experiments were to be conducted, their outcomes would have great variance).

Therefore, we decided to have less overlap between the training datasets and split the dataset into only five random test subsets of approximately the same size (fivefold cross-validation), and we averaged the ROC curves of these five splits; see Figs. 10 and 11. Finally, to find the hyperplane that is best based on our knowledge to date, we then trained the SVM on the full population.

3 Results

The characteristics of the frequency feature of Eq. 1 were assessed for a lower limit of 25 Hz and an upper limit of 8000 Hz in Fig. 6, which shows the distribution of the summed and averaged relative high-frequency components of all three sensors per subject, while Fig. 5 shows the individual components for all sensor signals. The distribution clearly shows a difference in distribution of healthy subjects and patients during flexion, while the difference in distribution is less pronounced during extension. Though the exact cause remains speculative,
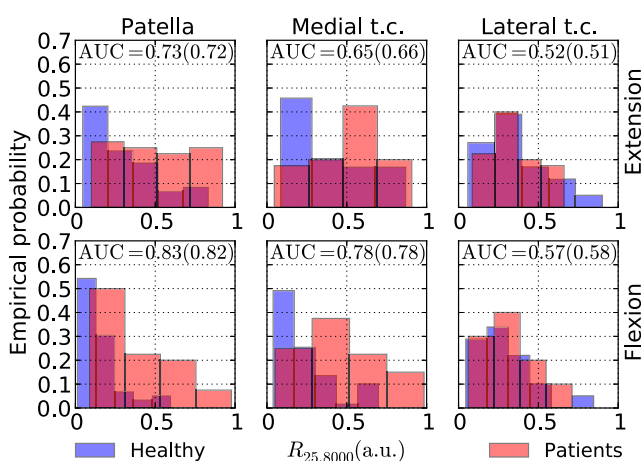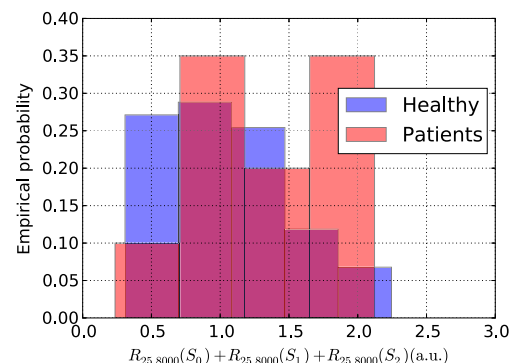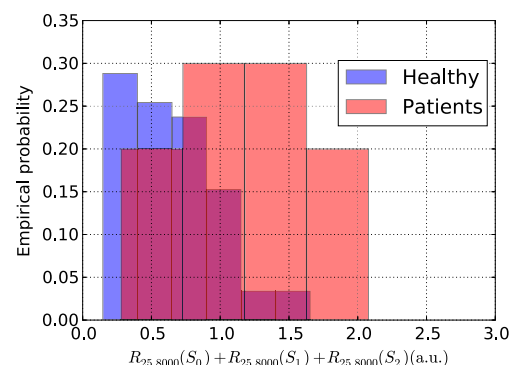

(a) Extension

(b) Flexion

Fig. 6 Histograms of the sum of the frequency features $R_{25,8000}$ (patella) $+$ $R_{25,8000}$(medial tibial condyle) $+$ $R_{25,8000}$(lateral tibial condyle). Each sample represents one knee. **a** Extension. **b** Flexion

Fig. 5 Histograms of frequency feature $R_{25,8000}$ for sensors from the patella, medial tibial condyle (t.c.), and lateral tibial condyle (t.c.). The SVM is trained and tested on the full population, and the AUC is calculated (in brackets: AUC of averaged ROC curve of a fivefold cross-validation). Each sample represents one knee

we assume that this is due to the greater muscle tension necessary to raise the leg to full extension compared to slowly dropping it to the neutral position, which likely produces strong mechanical excitations of the knee joint during the extension phase, rubbing cartilage against tibia and femur.

Figure 5 provides an overview of the $R_{25,8000}$ frequency feature for each individual sensor. Again, a difference in distribution between healthy volunteers and patients can be seen during flexion cycles for the patella and medial tibial plateau sensors, with patients tending to present higher relative frequency components. The lateral tibial plateau sensor shows a greater overlap of healthy subjects and patients, as do all three sensors during the extension cycle.

Figure 7 shows the distribution of the relative high frequency components for all subjects per sensor during flexion. As can be seen for the patella and medial tibial plateau sensor, patients' measurements are concentrated close to 0 for the $R_{75,8000}$ components and also have less contribution to the $R_{25,75}$ component. This indicates good linear separability. For the lateral tibial plateau, there is no such clear distinction.
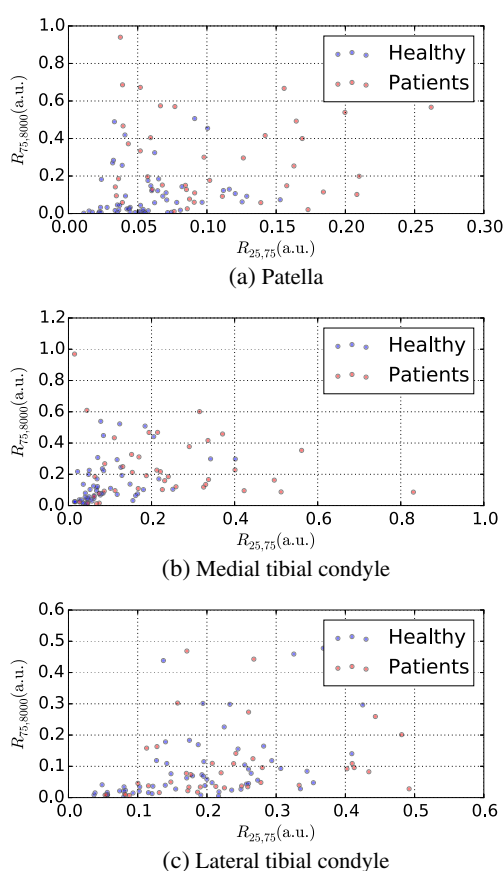

(a) Patella

(b) Medial tibial condyle

(c) Lateral tibial condyle

Fig. 7 Scatter plot of the $R_{75,8000}$ feature against the $R_{25,75}$ feature for flexion. Each plot represents a sensor; each sample represents one knee. **a** Patella. **b** Medial tibial condyle. **c** Lateral tibial condyle

Optimal frequency boundaries were derived from a comprehensive parameter study. In order to quantify the ability of the frequency features to classify into patient knees and non-patients knees, the SVM model was trained and tested on the full population. The optimal frequency boundary of $c = 25$ Hz for $R_{c,8000}$ was obtained from training a three-dimensional linear SVM on the $R_{c,8000}$ features of the three sensors. Figure 8b reveals that for features derived from the flexion phase, a lower limit of 25 Hz results in the best area under the ROC curve (AUC) of 0.85 for $R_{c,8000}$ when $c \in \{25, 50, \ldots, 100, 250, \ldots, 1000\}$. Again, a comparison with Fig. 8a reveals a superior performance of features from the flexion phase. The frequency boundaries for combinations of two features were obtained from training a six-dimensional linear SVM, where two features of three sensors were combined to identify the best performing parameter combination, in this case with an AUC of 0.89 for $R_{a,b}$ and $R_{b,8000}$ with $a = 25$ and $b = 75$, where optimization was performed on $a, b \in \{25, 50, \ldots, 100, 250, \ldots, 1000\}$. Figure 9 shows the resulting estimated probabilites for each knee when the SVM model is trained and tested on the full population with optimal parameters.

Based on the features shown in Fig. 7, a six-dimensional linear SVM was then trained and cross-validated using fivefold cross-validation; see Fig. 10 where an average AUC of 0.85 was obtained for the random split shown. In Fig. 11, ROC curves are further shown for the optimal features of the three-dimensional case and the six-dimensional case, first when trained on the full population and second when trained using fivefold cross-validation. It can be seen that the SVM that operates on $R_{25,75}$ and $R_{75,8000}$ performs slightly better than the SVM that makes use of $R_{25,8000}$. The similar performance in the fivefold cross-validation and in the training on the full population suggests that results may be generalized.

Based on the knowledge to date, i.e., based on the two trainings on the full population shown in Fig. 11, the parameters in Table 6 were found for the optimal hyperplane decision surface

$$w^T(x - m^T)\text{diag}(s)^{-1} + I = 0, \qquad (5)$$

where m^T and s are parameters that represent the scaling of the data before the SVM classification. Values are provided for both the three-dimensional case ($R_{25,8000}$) and the six-dimensional case ($R_{25,75}$, $R_{75,8000}$).

Given the high sampling frequency used in this work (VAG signals are usually recorded in the range of below 1000 Hz [1]), it may be assumed that a significant fraction of the acquired high-frequency components are dominated by noise, i.e., signals that do not originate from the physiological or pathophysiological physics of the knee. To assess this assumption, the study of Fig. 8 was repeated

Fig. 8 Extension and flexion: area under the ROC curve for different frequency boundaries a and b in linear SVM setups with features $R_{a,b}$ and $R_{b,8000}$—or with feature $R_{c,8000}$ alone. In the SVM training, the regularization parameter was set to $C = 1$. The SVM was trained and tested on the full population. **a** Extension. **b** Flexion

(a) Extension
(b) Flexion
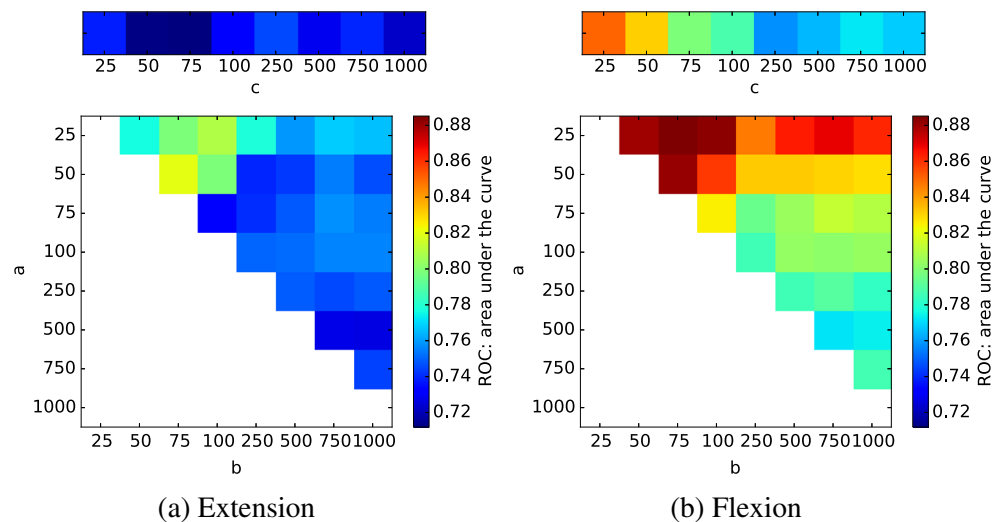
with the upper limit of the frequency ranges set to 1250 Hz instead of 8000 Hz and with the frequency features normalized with respect to the 10–1250 Hz power spectrum instead of 10–8000 Hz. It was revealed that this changed the characteristics of Fig. 8a, b only marginally (deviations of the areas under the ROC curves were on average 0.01 and up to 0.05, and the characteristic patterns were maintained—especially the situation and magnitude of the maxima between 25 and 10 Hz), revealing (a) that high-frequency components beyond 1250 Hz were indeed dominated by noise and (b) that the proposed methods are not susceptible to such noise. For clinical applications, a lower sampling frequency would further reduce the price of a measurement device.

Classification with the two non-linear SVMs was found to lead to an over-fitting on the present data: in the results of Fig. 12, similar or slightly improved areas under the curve with respect to Fig. 11 were achieved for training on the full population (full pop.) (0.88 and 0.94 over 0.89),

but worse results for fivefold cross-validation (fivefold CV) (0.83 and 0.83 compared to 0.85) revealed an over-fitting on the training population compared to the linear SVM.

When challenging the validity of the results in this work, one may naturally suspect the average speed of knee movement to be the root cause of differences in the frequency-based features we present. Lower speeds may mean lower frequencies, and patients may have moved their knees faster or slower than healthy volunteers. However, with the histograms of Fig. 13, it is apparent that the average duration of extending or flexing the leg D_c, i.e., the reciprocal of the average speed of movement, is distributed very similarly for both groups, with the exception of one outlier. This could further be underlined in a classification based on feature D_c alone: classification with linear SVMs
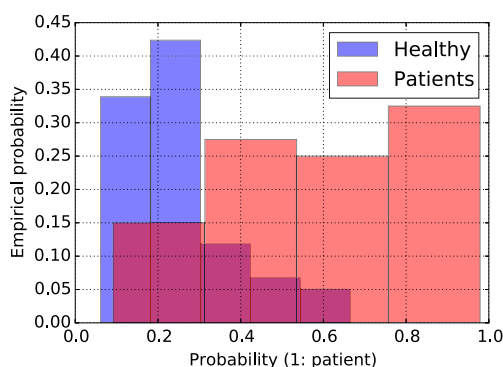

Fig. 9 Histogram of the estimated probabilities for the testing data of the SVM with features $R_{25,75}$ and $R_{75,8000}$ flexion. The SVM is trained and tested on the full population and probabilities are assigned to the samples using Platt scaling as implemented in LIBSVM [10, 29, 40]. Each sample represents one knee
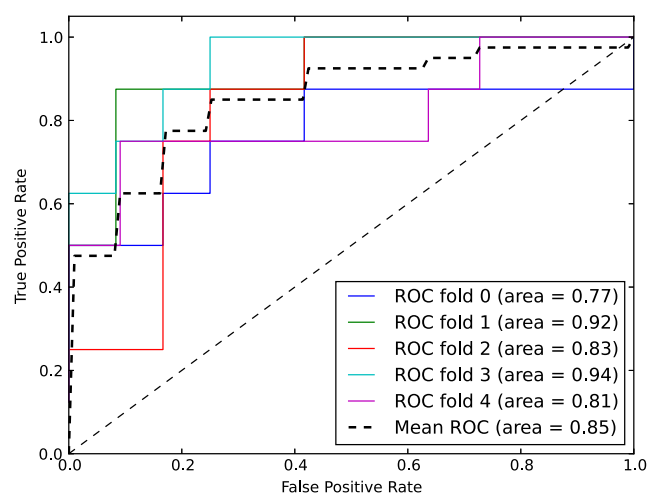

Fig. 10 Receiver operating characteristics (ROC) curves of the SVM with features $R_{25,75}$ and $R_{75,8000}$ for fivefold cross-validation (flexion). The dashed curve represents the interpolated average of the five ROC curves obtained in the cross-validation
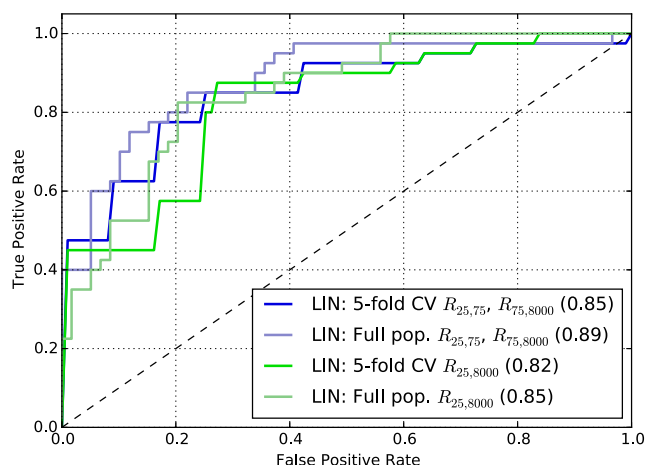
177

Fig. 11 Receiver operating characteristics (ROC) curves: SVM with linear (LIN) kernels with features $R_{25,75}$ and $R_{75,8000}$, as well as $R_{25,8000}$ for flexion. ROC curves are shown for training and testing on the full population (full pop.), as well as for fivefold cross-validation (5-fold CV), where the curves represent the average of the five ROC curves obtained

Fig. 12 SVM with non-linear kernels (RBF: radial basis function, POLY: polynomial); results for features $R_{25,75}$ and $R_{75,8000}$, flexion

was found to be only slightly better than guessing (AUCs for D_c alone were 0.60 for extension, 0.58 for flexion), revealing that such bias does not exist or influence the results in our study significantly.

4 Discussion

Figure 2 shows the age of the subject versus the Outerbridge classification. As chondromalacia typically affects middle-age to elderly patients, the distribution of patients is skewed towards older age. Although the participants of the study were selected with great care and emphasis on achieving highly comparable groups of healthy volunteers and patients in order to minimize the likelihood of confounders, the groups present noticeable differences in variables such as their average age and BMIs. Age differences between both groups however are hard to avoid, since the disease we focused on in this study, the degeneration of articular cartilage, is linked to higher age. Since chondromalacia and osteoarthritis mainly affect middle-aged and elderly patients, volunteers for the healthy group were excluded if they were above the age of 35 as to minimize the
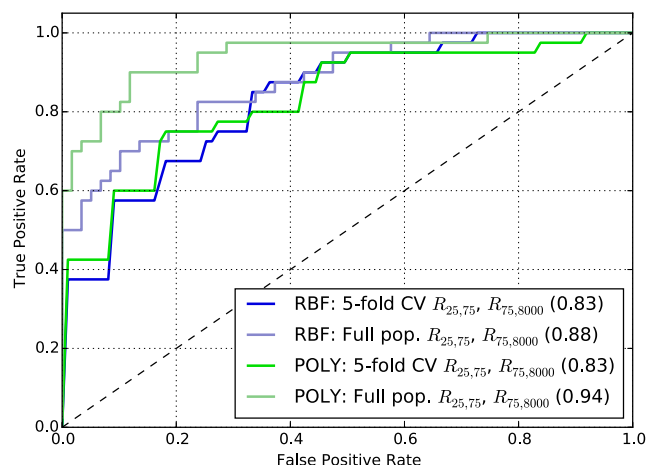
likelihood of them presenting cartilage damage which they were unaware of and which would falsify the measurements of the control group. Although the presence of cartilage damage could not be ruled out in the presumably healthy subjects by MRI due to logistical and financial reasons, as well as issues of the compliance of study participants, exclusion criteria for participants (see Section 2.1.3) were designed to only include those for whom optimal joint health could be assumed. Further, it could be shown that there is no linear correlation between the age and the sum of the features we used in the classification.

It may further be hypothesized that the amount of soft tissue under the sensors, especially in patients with higher BMIs, had an influence on the measurements' frequency characteristics. However, an assessment of the relation between the BMI and the sum of the frequency features $R_{25,8000}$(patella) $+ R_{25,8000}$(medial tibial condyle) $+ R_{25,8000}$(lateral tibial condyle) revealed a correlation of only 23%. This suggests that no linear dependence exists and underlines that the choice of relative features and/or the placement of sensors on areas with good bone contact were successful measures for minimizing measurement bias from BMI-related knee characteristics.

A weak point of this study is the prevalence of various additional knee disorders in the patient group: Many patients presented ligament injuries or meniscal damage and

Table 6 Parameters of the SVM hyperplane decision surfaces (Eq. 5) for training on the full population. Order of features in the vector dimensions is patella, medial t.c., and lateral t.c. for $R_{25,8000}$ (3-D) or for $R_{25,75}$ and then $R_{75,8000}$ (6-D)

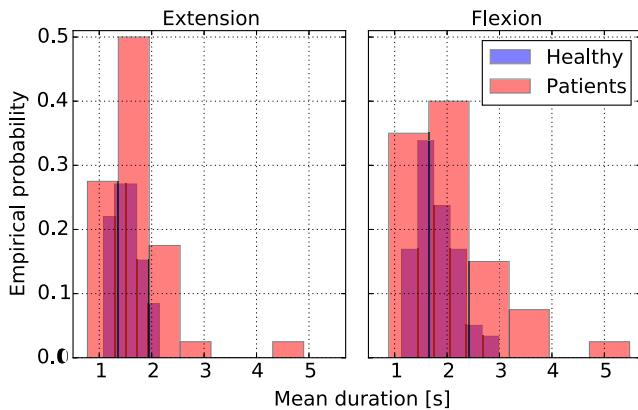| SVM	w^T	I	m^T	s
3-D	$w^T = (1.10\ 0.93 - 0.20)^T$	$I = 0.10$	$m^T = (0.25\ 0.32\ 0.30)^T$	$s = (0.21\ 0.24\ 0.18)$
6-D	$w^T = (0.83\ 0.67\ 0.63\ 0.09 - 0.12 - 0.51)^T$	$I = -0.04$	$m^T = (0.08\ 0.17\ 0.15\ 0.18\ 0.21\ 0.10)^T$	$s = (0.05\ 0.20\ 0.14\ 0.17\ 0.11\ 0.11)$

Fig. 13 Study of the mean duration of extending or flexing the leg as a potential bias: histograms of feature D_c. Each sample represents one knee

joint pathologies which heavily correlate with chondromalacia and/or osteoarthrosis. Whether the differences in the observed signals indeed stem from these problems or degeneration of cartilage remains an open question. However, the assumption that an increase in bone-to-bone contact causes an increase in relative high-frequency components is justified due to the underlying mechanics. Furthermore, the assessment of MRI scans by only one radiologist to determine the ground truth for the signal classification may be seen as problematic, especially since recent studies suggest moderate sensitivity of the MRI in detection of knee joint cartilage degeneration, as well as a moderate interrater reliability in the Outerbridge classification [47]. Further studies involving patients undergoing knee arthroscopy may improve results by providing diagnoses and CM/OA grades that are not restricted by the technical limitations of MRI and less prone to subjective interpretation.

Nonetheless, the discrepancy between MRI and/or clinical diagnoses in our patients group and their OKS suggests that many patients may not be fully aware of the

developing cartilage degeneration, while others experience great discomfort from comparably minor cartilage lesions. This proves the need for a cost- and time-efficient objective diagnostic tool to screen for cartilage degeneration, so early treatment can be initiated for those patients who require it.

Figure 14 shows the sum of all frequency features versus the clinical diagnosis, directly underlining the usefulness of the proposed features. By using features based on relative high-frequency components that are little sensitive to variances in signal gain, the SVM achieved a good linear separation and classification of healthy subjects and patients with a specificity of, e.g., 0.8 and a sensitivity of 0.75 when choosing an appropriate point on the ROC curve. Other medical diagnostic tools and tests such as MRI perform similarly well, with comparable values, cf., e.g., [30, 39]. Because of its performance, non-invasiveness and ease-of-use, we view vibroarthrography as a valid low-cost method for the detection of cartilage degeneration. A potential clinical application of the technology would be the quick objective assessment of knee joint status and screening for cartilage degeneration while in the examination room in order to assist the doctor in his decision, whether further diagnostic measures (i.e., MRI) or treatment are necessary. VAG may also enable consistent monitoring of the knee condition which can play a crucial role combined with conservative treatment and lifestyle adjustments, eventually leading to better early management of OA.

5 Conclusions

As was shown, simple features based on relative high-frequency components allow to distinguish between healthy and non-healthy subjects. Due to the proposed normalization, classification improved by an increase in the AUC of more than 0.20. We think this is rooted in that normalized features are robust against all signal acquisition or attenuation steps that influence the signal gain, such as fat tissue and variance in acoustic coupling due to sensor placement.

Machine learning techniques are able to linearly separate and classify cartilage defects. The sum of relative high frequency components can directly serve as an indication to the physician to justify further screening of the knee with more elaborate and expensive methods.

The presented method achieves a similar automated classification performance (0.80 sensitivity/0.75 specificity) as human interpretation of MRI images: McCauley et al. [30] report 0.86 sensitivity/0.74 specificity in detection of chondromalacia patellae using MRI, respectively. Pihlajamaki et al. [39] report 0.83 sensitivity/0.84 specificity for MRI for stage III chondromalacia. In the authors' opinion, this qualifies vibroarthrography using the presented robust features
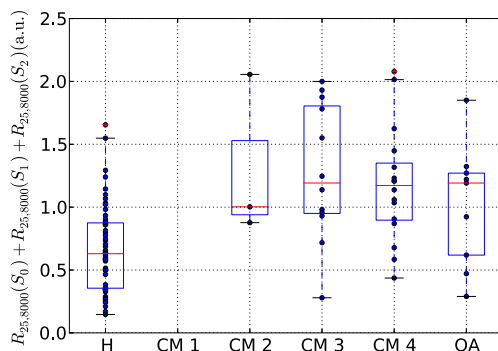

Fig. 14 Evaluation of the MRI-based Outerbridge classification against the sum of relative high-frequency components for the flexion phase. Each sample represents one knee

as a valid method for the detection of cartilage degeneration in the knee joint.

The features used in this paper are heuristical. Future research should focus on modeling the knee joint mechanics to derive acoustic features from a realistic model. This will result in theoretically founded features that incorporate both temporal and frequency aspects and could improve the accuracy of vibroarthrography further.

Compliance with Ethical Standards

References

1. Andersen RE, Arendt-Nielsen L, Madeleine P (2016) A review of engineering aspects of vibroarthography of the knee joint. Critical Reviews in Physical and Rehabilitation Medicine 28(1–2):13–32

2. Ben-Dor A, Bruhn L, Friedman N, Nachman I, Schummer M, Yakhini Z (2000) Tissue classification with gene expression profiles. J Comput Biol 7(3-4):559–583

3. Beverland D, Kernohan G, McCoy G, Mollan R (1985) What is physiological patellofemoral crepitus? Med Biol Eng Comput 23(2):1249–1250

4. Bircher E (1913) Zur diagnose der meniscusluxation und des meniscusabrisses. Zentralbl f Chir 40:1852–1857

5. Blodgett WE (1902) Auscultation of the knee joint. The Boston Medical and Surgical Journal 146(3):63–66

6. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory COLT 92 6(8):144–152

7. Brooks S, Morgan M (2002) Accuracy of clinical diagnosis in knee arthroscopy. Ann R Coll Surg Engl 84(4):265–8

8. Buckwalter JA, Mankin HJ (1998) Articular cartilage: degeneration and osteoarthritis, repair, regeneration, and transplantation. Instr Course Lect 47:487–504

9. Carl H (1885) Grundriss der chirurgie, 3rd edn. FCW Vogel, Leipzig

10. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2:27:1–27:27. software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

11. Cherkassky V, Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw 17(1):113–126

12. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

13. Dawson J, Fitzpatrick R, Murray D, Carr A (1998) Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg (Br) 80(1):63–9

14. Dunbar M, Robertsson O, Ryd L, Lidgren L (2001) Appropriate questionnaires for knee arthroplasty. Bone & Joint Journal 83(3):339–344x

15. Erb KH (1933) ÜBer die möglichkeit der registrierung von gelenkgeräuschen. Deutsche Zeitschrift fü,r Chirurgie 241(11):237–245

16. Fischer H, Johnson E (1961) Analysis of sounds from normal and pathologic knee joints. Arch Phys Med Rehabil 42:233

17. Frank CB, Rangayyan RM, Bell GD (1990) Analysis of knee joint sound signals for non-invasive diagnosis of cartilage pathology. IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society 9(1):65–8

18. Guermazi A, Roemer FW, Hayashi D (2011) Imaging of osteoarthritis: update from a radiological perspective. Curr Opin Rheumatol 23(5):484–91

19. Hsu CW, Chang CC, Lin CJ (2010) A practical guide to support vector classification. Bioinformatics 1(1):1–16

20. Hudelmaier M, Glaser C, Hohe J, Englmeier KH, Reiser M, Putz R, Eckstein F (2001) Age-related changes in the morphology and deformational behavior of knee joint cartilage. Arthritis Rheum 44(11):2556–61

21. Jackson DW, Simon TM, Aberman HM (2001) Symptomatic articular cartilage degeneration: the impact in the new millennium. Clin Orthop Relat Res 391:S14–S25

22. Jackson RW, Abe I (1972) The role of arthroscopy in the management of disorders of the knee. An analysis of 200 consecutive examinations. The. J Bone Joint Surg (Br) 54(2):310–22

23. Jiang CC, Liu YJ, Yip KM, Wu E (1993) Physiological patellofemoral crepitus in knee joint disorders. Bull Hosp Jt Dis (New York, N.Y.) 53(4):22–6

24. Kim KS, Seo JH, Kang JU, Song CG (2009) An enhanced algorithm for knee joint sound classification using feature extraction based on time-frequency analysis. Comput Methods Prog Biomed 94(2):198–206

25. King G, Zeng L (2001) Logistic regression in rare events data. Polit Anal 9(2):137–163

26. Krishnan S, Rangayyan RM, Bell GD, Frank CB (2000) Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology. IEEE Trans Biomed Eng 47(6):773–83

27. Krishnan S, Rangayyan RM, Bell GD, Frank CB (2001) Auditory display of knee-joint vibration signals. J Acoust Soc Am 110(6):3292–304

28. Lee TF, Lin WC, Wu LF, Wang HY (2012) Analysis of vibroarthrographic signals for knee osteoarthritis diagnosis. In: Proceedings - 2012 6th international conference on genetic and evolutionary computing, ICGEC 2012, pp 223–228

29. Lin HT, Lin CJ, Weng RC (2007) A note on Platt's probabilistic outputs for support vector machines. Mach Learn 68(3):267–276

30. McCauley TR, Kier R, Lynch KJ, Jokl P (1992) Chondromalacia patellae: diagnosis with MR imaging. Am J Roentgenol 158(1):101–105

31. McCoy GF, McCrea JD, Beverland DE, Kernohan WG, Mollan RA (1987) Vibration arthrography as a diagnostic aid in diseases of the knee. A preliminary report. J Bone Joint Surg (Br) 69(2):288–93

32. Menashe L, Hirko K, Losina E, Kloppenburg M, Zhang W, Li L, Hunter DJ (2012) The diagnostic performance of MRI in osteoarthritis: a systematic review and meta-analysis. Osteoarthritis and cartilage / OARS. Osteoarthritis Research Society 20(1):13–21

33. Moussavi ZM, Rangayyan RM, Bell GD, Frank CB, Ladly KO, Zhang YT (1996) Screening of vibroarthrographic signals via adaptive segmentation and linear prediation modeling. IEEE Trans Biomed Eng 43(1):15–23

34. Outerbridge RE (1961) The etiology of chondromalacia patellae. J Bone Joint Surg (Br) 43-B:752–7

35. Outerbridge R (1964) Further studies on the etiology of chondromalacia patellae. J Bone Joint Surg (Br) 46(2):179–190

36. Palmer AJR, Brown CP, McNally EG, Price AJ, Tracey I, Jezzard P, Carr AJ, Glyn-Jones S (2013) Non-invasive imaging of cartilage in early osteoarthritis. The bone & joint journal 95-B(6):738–46

37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

38. Peylan A (1953) Direct auscultation of the joints; preliminary clinical observations. Rheumatism 9(4):77–81

39. Pihlajamäki HK, Kuikka PI, Leppänen VV, Kiuru MJ, Mattila VM (2010) Reliability of clinical findings and magnetic resonance imaging for the diagnosis of chondromalacia patellae. JBJS 92(4):927–934

40. Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. MIT Press, Cambridge, pp 61–74

41. Prior J, Mascaro B, Shark L, Stockdale J, Selfe J, Bury R, Cole P, Goodacre J (2010) Analysis of high frequency acoustic emission signals as a new approach for assessing knee osteoarthritis. Ann Rheum Dis 69(5):929–930

42. Quatman CE, Hettrich CM, Schmitt LC, Spindler KP (2011) The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: a systematic review. Am J Sports Med 39(7):1557–68

43. Rangayyan RM, Oloumi F, Wu Y, Cai S (2013) Fractal analysis of knee-joint vibroarthrographic signals via power spectral analysis. Biomed Signal Process Control 8(1):23–29

44. Rangayyan RM, Wu YF (2008) Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions. Med Biol Eng Comput 46(3):223–32

45. Rangayyan RM, Wu Y (2009) Analysis of vibroarthrographic signals with features related to signal variability and radial-basis functions. Ann Biomed Eng 37(1):156–63

46. Rangayyan RM, Wu Y (2010) Screening of knee-joint vibroarthrographic signals using probability density functions estimated with Parzen windows. Biomed Signal Process Control 5(1):53–58

47. Reed ME, Villacis DC, Hatch GFR, Burke WS, Colletti PM, Narvy SJ, Mirzayan R, Vangsness CT (2013) 3.0-tesla MRI and arthroscopy for assessment of knee articular cartilage lesions, vol 36

48. Sandell LJ, Aigner T (2001) Articular cartilage and changes in arthritis. An introduction: cell biology of osteoarthritis. Arthritis Res 3(2):107–13

49. Schindler OS (2004) Synovial plicae of the knee. Curr Orthop 18(3):210–219

50. Scholkopf B, Smola A, Williamson R, Bartlett P (2000) New support vector algorithms. Neural Comput 12(5):1207–45

51. Schölkopf B., Burges CJ (1999) Advances in kernel methods: support vector learning. MIT Press, Cambridge

52. Shen Y, Rangayyan RM, Bell GD, Frank CB, Zhang YT, Ladly KO (1995) Localization of knee joint cartilage pathology by multichannel vibroarthrography. Med Eng Phys 17(8):583–594

53. Slonim DK (2002) From patterns to pathways: gene expression data analysis comes of age. Nat Genet 32:502–508

54. Tavathia S, Rangayyan RM, Frank CB, Bell GD, Ladly KO, Zhang YT (1992) Analysis of knee vibration signals using linear prediction. IEEE Trans Biomed Eng 39(9):959–70

55. Umapathy K, Krishnan S (2006) Modified local discriminant bases algorithm and its application in analysis of human knee joint vibration signals. IEEE Trans Biomed Eng 53(3):517–23

56. Vapnik V (1999) An overview of statistical learning theory. IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council 10(5):988–99

57. Wakefield RJ, Kong KO, Conaghan PG, Brown AK, O'Connor PJ, Emery P (2003) The role of ultrasonography and magnetic resonance imaging in early rheumatoid arthritis. Clin Exp Rheumatol 21(5 Suppl 31):S42–9

58. Walters C (1929) The value of joint auscultation. The Lancet 213(5514):920–921

59. Welch PD (1967) The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 15(2):70–73

60. Wise CH (2015) Orthopaedic manual physical therapy from art to evidence. In: Wise CH (ed). F.A. Davis Company, Philadelphia

61. Wu Y, Cai S, Yang S, Zheng F, Xiang N (2013) Classification of knee joint vibration signals using bivariate feature distribution estimation and maximal posterior probability decision criterion. Entropy 15(4):1375–1387

62. Wu Y (2015) Knee joint vibroarthrographic signal processing and analysis. Springer, Berlin

63. Wu Y, Chen P, Luo X, Huang H, Liao L, Yao Y, Wu M, Rangayyan RM (2016) Quantification of knee vibroarthrographic signal irregularity associated with patellofemoral joint cartilage pathology based on entropy and envelope amplitude measures. Comput Methods Prog Biomed 130:1–12

64. Zhang YT, Frank CB, Rangayyan RM, Bell GD (1992) Mathematical modeling and spectrum analysis of the physiological patello-femoral pulse train produced by slow knee movement. IEEE transactions on biomedical engineeringmedical engineering 39(9):971–9

65. Zhang YT, Rangayyan RM, Frank CB, Bell GD (1994) Adaptive cancellation of muscle contraction interference in vibroarthrographic signals. IEEE Trans Biomed Eng 41(2):181–91
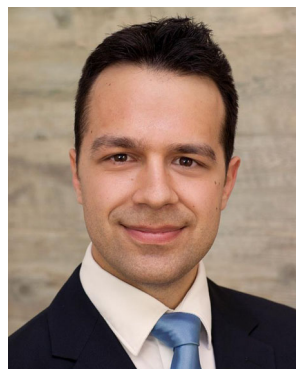
Nima Befrui is an orthopaedic trauma surgeon at the Ludwig-Maximilians-University Hospital of Munich, Germany.

Dr Jens Elsner received M.Sc. and Ph.D. degrees in Electrical Engineering from Karlsruhe Institute of Technology in 2007 and 2012, respectively. He is partner and Chief Engineer at Munich Innovation Labs.

Achim Flesser MBA + Master in Mechanical Engineering. Since 07/ 2004 Head of R&D, CEO of SME cpengineering GmbH, innovation of products, feasibility studies, design, construction and simulation, prototypes and serial production.

Marcus Müller Former Innovation Research Expert at Munich Innovation Labs, currently active as Scientific and Education Assistant at the Communications Engineering Lab, Karlsruhe Institute of Technology.

Dr Jacqueline Huvanandana received a B. Eng. degree in Biomedical and Electrical Engineering from the University of Sydney. Former analyst at Munich Innovation Labs, current PhD candidate at Woolcock Institute of Medical Research.

Dr Walther HW Schulze received a Ph.D. in Biomedical Engineering from the Karlsruhe Institute of Technology. He is the founder of Evolunis, and his research interests include applications of illposed inverse problems to biomedical research.

Dr Oussama Jarrousse received a Ph.D. in Biomedical Engineering from the Karlsruhe Institute of Technology. He works as Senior Data Scientist at Munich Innovation Labs.

Dr Stefan Taing has studied Computer Sciences and Business Engineering in Vienna, Birmingham and Taipei and holds a Ph.D. from the LMU Munich. He serves as Partner at Munich Innovation Labs.

Tuan Nam Le worked as a Research Associate at the Trauma Surgery Department, Klinikum Innenstadt of LMU Munich, where he focused on biomedical signal processing at the Navigated Augmented Reality Visualization System (NARVIS) Lab.

Dr Simon Weidert Orthopeadic trauma specialist and research group leader for computer aided surgery at the university hospital of LMU Munich. His research focuses on the intersection of surgery and computer sciences.

Appendix VI

How do different LMA brands compare to face-mask ventilation in a LMA teaching manikin?

I contributed to the signal processing of the ventilator data.

among babies of Indigenous mothers and teenagers in Australia. Confirmation with a longer time series is needed.

Reference: [1]AIHW 2016. Monitoring the health impacts of mandatory folic acid and iodine fortification. Cat. no. PHE 208. Canberra: AIHW.

THE EFFECTIVENESS DURING MULTITASKING AND WITH BACKGROUND NOISE OF NOVEL PULSE OXIMETER SONIFICATIONS DESIGNED FOR MONITORING PREMATURE NEONATES

Kelly Hinckfuss[1]*, Penelope Sanderson[1], Robert Loeb[2], Birgit Brecknell[1], David Liu[1], Helen Liley[3]
[1]University of Queensland, the [2]University of Arizona, the [3]Mater Mothers Hospital

Background: Previous research indicates that the pulse oximeter auditory display could be improved by adding features that indicate the general clinical range of neonatal SpO_2. We investigated the impact of background noise and a demanding arithmetic task on participants' monitoring accuracy with different pulse oximetry auditory displays.

Method: We classified SpO_2 into five ranges suggested for premature neonates on oxygen support: Very Low (80–83); Low (84–89); Target (90–95); High (96–98), Very High (99–100). Three sonifications were tested: 1) the Beacon84 sonification played a beacon (intermittent reference tone) before every 8th or 4th pulse tone, depending on severity of target range deviation; 2) the Tremolo37 sonification added a vibrating quality to the pulse tones (3 or 7 cycles, depending on severity); 3) Control + Alarms used conventional variable-pitch pulse tones with alarm limits of 96% and 88%. Eight non-clinicians completed all sonification conditions in counterbalanced order. Background hospital sound (50 to 65 dBA) played continuously. After training, participants completed thirty 30-second trials while performing arithmetic tasks, and indicated (a) when SpO_2 transitioned into or out of target range, and (b) the SpO_2 range in which each trial finished.

Results: Accuracies at identifying transitions and SpO_2 range were Beacon84 (76%, 70%), Tremolo37 (83%, 67%), Control + Alarms (49%, 45%). T-tests revealed significant performance differences with each novel sonification versus the control sonification (p <.01 for all comparisons, Bonferroni corrections applied). There were no differences in arithmetic task accuracy.

Conclusions: Novel pulse oximeter sonifications improve monitoring during multitasking compared with conventional variable-pitch pulse tones plus alarms.

CAN AMBU SELF-INFLATING BAG AND NEOPUFF INFANT RESUSCITATOR PROVIDE ADEQUATE AND SAFE MANUAL INFLATIONS FOR INFANTS UP TO 10 KGS WEIGHT?

Mark Tracy[1]*, Rajesh Maheshwari[1], Dharmesh Shah[1], Murray Hinder[1]
[1]Westmead Hospital Neonatal ICU

Background: Manual resuscitation devices for infants and newborns must be able to provide adequate ventilation in a safe and consistent manner across a wide range of patient size (0.5 to 10kgs) and differing clinical states. There is little comparative data assessing performance of common infant manual resuscitation devices such as Neopuff t-piece resuscitator (NTPR) and Ambu self-inflating bag (SIB) across the manufacturers recommended operating weight range.

Method: Five experienced clinicians delivered targeted ventilation to three lung models differing in compliance, delivery pressures and inflation rates; Pre-term (0.5 mL/cmH$_2$O, 25/5cmH$_2$O, 60 per minute), Term (3 mL/cmH$_2$O, 30/5 cmH$_2$O, 40 per minute) and Infant (9 mL/cmH$_2$O, 35/5 cmH$_2$O, 30 per minute). The NTPR was examined with three gas inflow rates (5, 10 and 15LPM).

Results: 3309 inflations were collected and analysed. The NTPR was unable to reach set peak inflation pressures and exhibited seriously elevated PEEP with all inflow gas rates (p < 0.001) in the infant model. The Ambu SIB accurately delivered targeted pressures in all three models.

Conclusions: The Ambu SIB was able to accurately deliver targeted pressures across all three models from preterm to infant. The NTPR infant resuscitator was unable to deliver the targeted pressures in the infant model developing clinically significant levels of inadvertent PEEP which may pose risk during infant resuscitation.

HOW DO DIFFERENT LMA BRANDS COMPARE TO FACEMASK VENTILATION IN A LMA TEACHING MANIKIN?

Mark Tracy[1]*, Archana Priyadarshi[1], Krista Lowe[1], Dimple Goel[1], Jacqueline Huvanandana[2], Murray Hinder[1]
[1]Westmead Hospital Neonatal ICU, the [2]Dept Engineering and IT, BMET Institute, Sydney University

Introduction: ILCOR and AAP recommend the use of Laryngeal Mask Airway (LMA) with newborn infants (?34wk gestation or >2Kg weight) requiring resuscitation when face mask ventilation or tracheal intubation is unsuccessful. No data exists to allow broad LMA device comparison. We compared 7 brands of size1 LMA devices with two brands of facemask using Ambu self-inflating bag (SIB).

Methods: 40 experienced neonatal staff trained in resuscitation provided 2 min PPV cycles using Ambu SIB with manometer and PEEP valve (set to 5cmH2O) for each LMA, and facemask randomly sequenced. All subjects received prior training in LMA insertion. A specialised LMA newborn training manikin head was attached to a leak free test lung with compliance of 2 ml/cmH2O. Subjects were required to provide sufficient pressure to observe test lung inflation.

Results: 12,415 recorded inflations for 7 brands of LMA and 2 face masks were analysed using ANOVA for repeated measures. Leak detected was lowest with I-Gel brand; mean 5.7% compared to facemask (triangular 42.7% round 35.7%) and other LMAs (means ranging from 45.5% to 65.4%) p < 0.0001. PIP was higher with I-Gel; mean 29cmH20 compared to facemask (triangular 22.8, round 25.9) and other LMAs (means ranging from 14.4 to 22.0) p < 0.0001. PEEP was higher with I-Gel; mean

5.0cmH20 compared to facemask (triangular 3.0, round 3.6) and other LMAs (means ranging from 0.6 to 2.6) p < 0.0001.

Conclusion: I-Gel LMA performed better than all other LMAs and facemasks/SIB evaluated in this study.

NEOPUFF T-PIECE RESUSCITATOR: DOES DEVICE DESIGN AFFECT DELIVERED VENTILATION?

Murray Hinder[1*], Pranav Jani[2], Archana Priyadarshi[1], Alistair McEwan[3], Mark Tracy[1]

[1]Westmead Hospital Neonatal ICU, the [2]Westmead Hospital, the [3]Dept Engineering and IT, BMET Institute, Sydney University

Background: The T-piece resuscitator (TPR) is in common use worldwide to deliver positive pressure ventilation during resuscitation of infants <10 kg. Ease of use, ability to provide positive end-expiratory pressure (PEEP), availability of devices inbuilt into resuscitaires and cheaper disposable options have increased its popularity as a first-line device for term infant resuscitation. Research into its ventilation performance is limited to preterm infant and animal studies. Efficacy of providing PEEP and the use of TPR during term infant resuscitation are not established.

Method: A single operator experienced in newborn resuscitation provided positive pressure ventilation in a randomised sequence to three different Crs models (0.5, 1 and 3 mL/cmH2O) at three different set PIP (20, 30 and 40 cmH2O). Set PEEP (5 cmH2O), gas flow rate and inflation rate were the same for each sequence.

Results: A total of 1087 inflations were analysed. The delivered mean PEEP was Crs dependent across set PIP range, rising from 4.9 to 8.2 cmH2O. At set PIP 40 cmH2O and Crs 3 mL/cmH2O, the delivered mean PIP was significantly lower at 35.3 cmH2O.

Conclusions: As Crs increases, the Neopuff TPR can produce clinically significant levels of auto-PEEP and thus may not be optimal for the resuscitation of term infants with healthy lungs.

PROGESTERONE THERAPY DURING PREGNANCY INCREASES MATERNAL BUT NOT FETAL CORTISOL OR NEUROSTEROID CONCENTRATIONS IN GUINEA PIGS

Julia Shaw[1], Hannah Palliser[1], Jonathan Hirst[1*]

[1]University of Newcastle

Background: Progesterone (PROG) therapy in pregnancy is used to reduce the risk of preterm labour however the effects on the steroid profile of the fetus remain unclear. We have previously found PROG therapy in male preterm guinea pig neonates leads to increased cortisol and reduced myelination. Hence, we aimed to determine the effects of PROG treatment on steroid profiles and neurodevelopmental markers in guinea pig fetuses.

Method: Pregnant guinea pig dams were administered PROG or vehicle orally (GA29-61) and saliva collected daily. Fetal brains and plasma were collected at GA62 (preterm) and GA69 (term). We measured salivary PROG and plasma cortisol by EIA and the neurosteroid, allopregnanolone (ALLO), by RIA. The myelination marker, myelin basic protein (MBP) and astrocyte marker, glial fibrillary acidic protein (GFAP), were quantified by immunohistochemistry.

Results: PROG treatment significantly increased both maternal salivary PROG and cortisol levels. Plasma ALLO levels were unchanged. Term female fetuses exhibited raised PROG levels whilst both cortisol and ALLO levels remained unaffected. Fetal MBP and GFAP expression were also largely unaffected with only preterm males displaying reduced GFAP expression in the hippocampal CA1.

Conclusions: The placental barrier adequately protected the fetus from increased maternal cortisol following maternal PROG administration, however the treatment was not sufficient to raise ALLO levels.

DOES THE HOSPITAL ENVIRONMENT INFLUENCE TEST ORDERING PATTERN AND FREQUENCY IN PREMATURE NEONATES?

Hodge TL[*1,2], Widjaja CE[1,3], Holberton J[3], Fan WQ[1,2]

[1]University of Melbourne, Melbourne, Australia
[2]The Northern Hospital (TNH), Melbourne, Australia
[3]The Mercy Hospital for Women (MHW), Melbourne, Australia
Email: tayhodge@gmail.com

Background: This study investigates differences in premature neonate test ordering patterns and frequencies between the special care nurseries of two hospitals. Institution specific factors may influence test ordering. Reducing unnecessary tests can impact patient burden and hospital expenditure.

Method: Data was retrospectively collected on neonates born over two-years from TNH (n = 267) and MHW (n = 391). Neonates were aged between 32–36 weeks gestation (GA), weighed >1500 g at birth (BW), were seizure free and required no respiratory support.

Results: TNH's mean GA (34.69 ± 0.99 weeks) and BW (2303.79 ± 442.64 g) were similar to MHW's (GA 34.67 ± 1.24 weeks and BW 2294.50 ± 461.32 g): p = 0.80 and p = 0.82 respectively. TNH's outcomes did not differ significantly from MHW's (mean ± standard error lengths of stay of 11.6 ± 0.41 vs 11.6 ± 0.40 respectively, p = 0.44). TNH's mean (± standard error) test-ordering rate of 1.11 ± 0.06 tests/patient/day was significantly lower than MHW's rate of 1.40 ± 0.07 tests/patient/day (p < 0.01). On average TNH ordered significantly fewer tests than MHW on days 1 and 2 of admission (2.54 ± 0.17 vs 3.62 ± 0.13, p < 0.01 and 1.63 ± 0.11 vs 2.17 ± 0.12, p < 0.01 respectively). Patterns of test types ordered and rates of tests requiring recollection differed between the hospitals.

Conclusions: Despite having demographically similar cohorts, significant differences in premature neonate test-ordering practices exist between TNH and MHW. Further investigation is warranted given potential patient and financial consequences.

GOLDEN HOUR MANAGEMENT PRACTICES FOR INFANTS <32 WEEKS GESTATIONAL AGE IN CANADA

Kate Hodgson[1*], Michael Dunn[2], Mary Seshia[3], Georg Schmoelzer[4], Vibhuti Shah[5]

[1]Royal Women's Hospital, the [2]Sunnybrook Health Sciences Centre, the [3]University of Manitoba, Winnipeg, the [4]Royal Alexandra Hospital, the [5]Mount Sinai Hospital, Toronto

Appendix VII

A comparison of low-cost body composition assessment using near-infrared interactance in infants aged 0-2 years

Model development for body composition assessment using near-infrared interactance in infants aged 0-2 years

J Huvanandana[1], P Jones[1], H Jeffery[1,2], A Carberry[1,2], S Norris[3], A McEwan[1]

[1] School of Electrical and Information Engineering, University of Sydney, NSW, Australia

[2] Sydney School of Public Health, University of Sydney, NSW, Australia

[3] MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa

Introduction

Inexpensive and accessible body composition measurement represents an important step in addressing global development and nutritional needs. Historically, measures of infant malnutrition and growth centred around weight percentiles for a given age, with the widespread use of reference charts from the World Health Organization (WHO). This lead to classifications of small, appropriate and large for gestational age infants, as demarcated by the 10th and 90th weight percentiles. Although weight-for-age offers an understanding of infant growth to a certain extent, it is not sensitive to stunted growth and thus does not capture infants who fail to reach linear growth potential. Another commonly used anthropometric measure is mid-upper arm circumference which often guides assessments of wasting. Though this measure is easily obtained, its reliability and the relationship to mortality in this population is not well-characterised.

A 2010 UNICEF/WHO report identified undernutrition as a contributing cause in over one third of child deaths [1]. The major proportion of these deaths occur within the first year of life, particularly in the neonatal period and in low-middle income settings [2]. Malnutrition has also shown impacts beyond infancy, with links to diabetes and obesity in later life [3]. Body fat also plays a neuroprotective role, with undernutrition in early life potentially affecting later neurodevelopmental outcomes [4, 5].

There are various approaches to body composition assessment, ranging from dual x-ray absorptiometry (DXA) to air-displacement plethysmography (ADP). When it comes to suitability

Of the non-cadaveric methods, the four-compartment model is considered the criterion method for body composition assessment. This multicompartment approach involves the independent assessment of various fat-free mass (FFM) components, allowing for a more robust albeit indirect estimation of body fat. The FFM is typically divided into water, protein and mineral mass components. The total body water component is assessed through deuterium dilution where saliva and urine samples collected prior to ingesting a dose of deuterium and following an equilibration period are analysed [6]. The bone mineral content is derived from dual x-ray absorptiometry (DXA). Total body protein mass can be measured using potassium counting. Some multicompartment approaches also require body volume which can be assessed through densitometry methods such as air-displacement plethysmography.

Anthropometric approaches may suitable for use in low and middle income settings, given their simplicity and scalability. Relevant metrics include birthweight, weight-for-length or body mass index (BMI: weight for length-squared) and their corresponding percentiles. Thought important measures, work by Carberry et al. [7] has demonstrated that measures of body fat percentage (BF%) may be more predictive in identifying a composite neonatal morbidity compared with birthweight percentiles. Mid-upper arm circumferences (MUAC) may also be used in screening

for acute malnutrition, though there is limited data on its reliability and predictive value for mortality in infants under 6 months.

Given that a large proportion of infant deaths under five occur in the neonatal period and the majority of these are in low and middle countries, there is a clear need for a simple, low-cost and portable method for rapid assessment of body composition. Near-infrared (NIR) interactance, has previously been explored and used as an input to body composition assessment, with several studies evaluating the agreement between commercially-available devices operating on these principles with other reference methods [8, 9] and the development of models surrounding body composition [10].

The near-infrared (NIR) region of the electromagnetic spectrum spans from approximately 600-1300 nm, where the absorption properties of tissue vary according to its constituents. These include water, fat and collagen. Early studies of NIR spectrometry by Conway et al. [11] identified visible peaks in the spectral profile measured from the triceps of adult subjects, corresponding to the pure fat absorption band at 930 nm and pure water absorption band at 970 nm [11].

Building on this work, the primary aim of this paper is to develop a model using features extracted from NIR interactance for the estimation of infant fat mass and BF%. The second aim is to compare the performance of this model against other accessible methods of body composition assessment for low and middle-income countries, using DXA as the reference method. The comparative measures for adiposity and/or nutritional status include mid-upper arm circumference, weight-for-length, weight percentiles and sum of skinfolds.

Methods

Study Population

This study was approved by the Ethics Committee at the University of Sydney (USyd) and the University of Witwatersrand (Wits) (USyd HREC number: Project No.: 2015/595; Wits HREC number: M150774). The study has also been registered on the Australia and New Zealand Clinical trials Registry (ANZCTR) number: ACTRN12615001318572. The infants and children enrolled were recruited from the Soweto, South Africa pregnancy and follow up clinics over a period of 9 months (April to December 2016). Inclusion criteria included well infants born at term (37-41 weeks) with an absence of major congenital anomalies, morbidity and chronic health problems from birth (within first 48 hours) up to 24 months of age. Other inclusion criteria included ≤ 24 months of age at measurement, mother's age > 18 yrs, singleton pregnancy, living within the study area as well as English literacy and comprehension. Infants that had undergone any DXA scans in the prior 12 months were also excluded from enrolment. Informed parental written consent was obtained and participation was voluntary.

Data Collection

The data in this study was collected as part of a population-based cross-sectional study of ages from birth (within 48 hours) to 24 months. Subject data included anthropometric measures such as weight, length, head and mid-upper arm circumferences. Weight was measured using electronic scales (Seca 376, Hamburg, Germany). Length was measured using an infantometer (Harpenden, Holtain Model 702) which has a fixed headboard and moveable footboard following a two-trained person technique. Head, abdomen, thigh and mid-upper arm circumferences were measured using a metal tape measure. Skinfold measurements were conducted using handheld callipers at four anatomical locations; triceps, subscapular, mid-thigh and flank. Other metadata including sex and ethnicity was also included.

For all participants, BF% was measured using dual x-ray energy absorptiometry (DXA) (Hologic DiscoveryA DXA S/N 86254 APEX software version 4.0.2, Hologic Inc., Waltham, MA, USA) with relevant paediatric software installed. The DXA measurement involved the infant lying flat on a scanner bed to measure bone mineral content (BMC) and total body fat %. Bone mineral content information was also collected to serve as inputs for the final 3-compartment model. Fat mass is determined from weight and BF%.

Prior to NIR measurements, reference scans against ambient light and a dark material were obtained. The spectrometer (QEPro, Ocean Optics) collected a range of wavelengths in the range of 350-1100 nm, with reflection profiles determined by customised software developed in-house using LabView, QEProInterface v3.1, adjusting the measured intensity for the dark and light calibration scans. The NIR measurements were conducted through the skin's surface at 4 the anatomical locations where skinfold thickness measurements were obtained; triceps, subscapular, mid-thigh and flank for approximately 30 seconds performed by a researcher (Appendix 1). Two sets of measurements were taken at each of the four sites.

All measurements were conducted in duplicate by two-trained researchers. Any pair of measurements falling outside the maximum allowed differences were repeated by both researchers. If this second pair of measurement values again exceeded the limits for the measurement, the researchers repeated the measurement for a third and final time.

Model Development

Pre-processing, model development and statistical analysis were completed using Python (Python Software Foundation, version 2.7 https://www.python.org/).

Feature selection and parameter fitting was undertaken on a randomised division (two thirds) of the dataset and subsequently evaluated on the remaining third. The division was stratified based on age, that is, similar distributions of ages across the training and testing subsets. Two population groups were defined for model development; 0-8 months and 9-25 months inclusive. The prime motivation for this age division was observed plateau or slight decline in BF% 6-9 months of age, as also observed by Butte et al. [12]. The model development process involved two stages; first, the feature selection and the second, model fitting on a subset (training and validation set) of the available data.

The wavelength range of interest spanned 850-1100 nm, with potential features being sampled from 10 nm intervals. Feature selection from this pool of 25 features for each measurement site (subscapular, flank, triceps and thigh) involved a backward elimination approach, recursive feature elimination (RFE) with cross-validation. A wrapper-based technique for feature selection, this RFE involved the continual removal of lowest-ranked features – the reflection % values at the respective wavelengths, from the model. The 2-fold cross-validated mean-squared error was evaluated at each point of elimination, with the optimal set of features defined by the minimised error. An example of this process if shown in Figure 1. Model fitting was subsequently undertaken on the selected features for male- and female- specific divisions of each age group.

Results

Of the 651 eligible term infants (37–42 weeks gestational age) aged from birth (48 hours) up to 24 months, 416 (63.9%) had available DXA measurements. Of these, 370 infants (88.9%) had complete NIR and anthropometric measurements and were included in the model development. The characteristics of the study population are summarised in Table 1.

Table 2 summarises the wavelengths selected from the recursive feature selection process. Linear coefficients are based on the regression model fitted to the training set. Equation (1) shows the

generalised example for the 0-8 month group, with n reflection-based features, weight for length (g/cm) and sex (male = 1, female = 0). Note that reflection values are expressed as the difference in reflection from that at wavelength x to that at 930 nm, normalised by the reflection at 970 nm (equation 2).

$$BF\% = \beta_1 R_1 + \beta_2 R_2 + \beta_3 R_3 \ldots \beta_n R_n + \beta \frac{W}{L} + \beta S + \varepsilon$$

Where ε is the intercept and R_x is derived using reflection (r) at wavelength x:

$$R_x = \frac{r_x - r_{970}}{r_{930}} (2)$$

Principal component analysis (PCA) of the various weight-length ratios (W/L, W/L^2 and W/L^3) was undertaken in each of the groups. A single component could account for 86.7% and 92.9% of the total variance in the 0-8 and 9-25 month age groups, respectively. This lower proportion could potentially be due to inaccuracies in length measurements in the younger group [13]. This motivated the inclusion of W/L in the 0-8 month group and W/L^2 in the 9-25 month group. In the former instance, W/L accounted for 55.8% of the variation in body fat, while W/L^2 explained 29.4% for the older group. The W/L and W/L^3 ratios explained 22.6% and 22.4% of the variance in percentage body fat for the older group.

Model evaluation and comparison with other anthropometric methods was undertaken on the remaining third (testing set) of the data. Correlation and Bland-Altman agreement for each of the models are summarised in Table 3 with Bland-Altman plots for all models across the trainings groups displayed in Figure 2.

The agreement of NIR model measurements with body fat % as measured by DXA was 0.32 (-9.39, 10.03) for the 0-8 month group and -0.63 (-9.58, 8.31) for the 9-25 month group. The transformed FM estimations from the product of BF% and weight (g) are also included, showing a greater R fit statistic. To better characterise the sensitivity of the model we also evaluated the variability of predicted % fat values between consecutive sets of NIR measurements. Mean (SD) difference in the BF% estimate was 1.84 (1.55) and 1.80 (1.58) for the 0-8 month and 9-25 month groups, respectively.

The precision of the developed models in estimating BF% in an independent sample was further evaluated using cross-validation [14]. The cross-validated estimation errors across each of the methods are shown in the boxplots for each population in Figure 3. The inclusion of NIR interactance features to the W/L and sex models for each group exhibited a lower distribution of error from cross validation. The next lowest errors were achieved by sum of skinfold thickness measurements.

Discussion

Estimation of body fat and comparison with other models

This study focused on the comparison of NIR interactance-based models for the estimation of body fat against other accessible methods for low-middle income settings and population (infants aged 0-2 years). These models were evaluated based on Bland-Altman agreement with DXA-derived BF% and FM, the variance explained by each of the models measured by R and the distribution of standard error across 100 repetitions of cross-validation.

Across both age groups (0-8 months and 9-25 months), NIR displayed narrower 95% limits of agreement from Bland-Altman analysis. There was also no important bias between any of the methods of comparison and DXA, with little evidence of curvature in the residuals or proportional

bias. This supports the inclusion of NIR interactance features in the estimation model for infant body fat, which may help to improve the robustness of basic anthropometric indices such as weight/length^2 which has been considered a potential proxy for adiposity.

The sum of skinfold thickness displayed the next highest agreement and correlation with DXA across both groups, and the highest (R = 0.835) for the prediction of FM in the 9-25 month group. However, it is also necessary to consider operator training requirements when comparing the methods. SFT measurements generally require a high degree of training and are conducted in duplicate with defined tolerance to ensure agreement [15]. There is also a tendency for skinfold thickness measurements to overestimate fat in lean infants and underestimate it in those with more fat [16].

Wavelength selection

The wavelength range for feature selection spanned 850-1100 nm. Recursive feature elimination identified wavelengths in the range of 900-1000 nm range as most relevant for the body fat estimation model. This range encompassed peaks of water (930 nm) and fat (970 nm) reported in literature [11].

Reproducibility

Additional analysis to characterise the reproducibility of the NIR models was undertaken using single sets of NIR measurements acquired by different observers. Mean (SD) absolute differences in the BF% estimates were 1.84 (1.55) and 1.80 (1.58) for the 0-8 month and 9-25 month groups, respectively.

Strengths and Limitations

One of the strengths of this study is the large population based sample across a wide range of age groups (0-2 years). We also used DXA, one of the gold standard methods for body composition, as the reference method for model development.

The model development for the estimation of body fat in infants aged 0-2 years was limited by the non-normal distributions of the infant data collected, as well as the relatively low representation of 'low fat' infants, with Z-score plots on WHO percentile charts indicating skewness towards the higher end across skinfolds.

The model developed used DXA as the criterion method, which is considered one of the gold standard techniques for measuring body composition. However one of the considerations and potential limitations is that in the paediatric population between gold standards such as DXA and ADP do not agree as well [17] as in other populations. This serves as motivation for further development and evaluation using a multi-compartment approach, which will be explored in future work using deuterium dilution as in input for the three-compartment (3C) model.

Conclusion

This study developed and evaluated models using near-infrared interactance features in comparison to other portable and accessible methods for infant body composition assessment in low-middle income settings. The two NIR models offered a more robust means of body fat percentage estimation in infants aged 0-25 months, showing a lower distribution of estimation error across multiple cross-validation repetitions (figure 3). The developed NIR models exhibited a mean difference with DXA BF% of 0.32 (-9.39, 10.03) in the 0-8 month group and -0.63 (-9.58, 8.31) in the 9-25 month group.
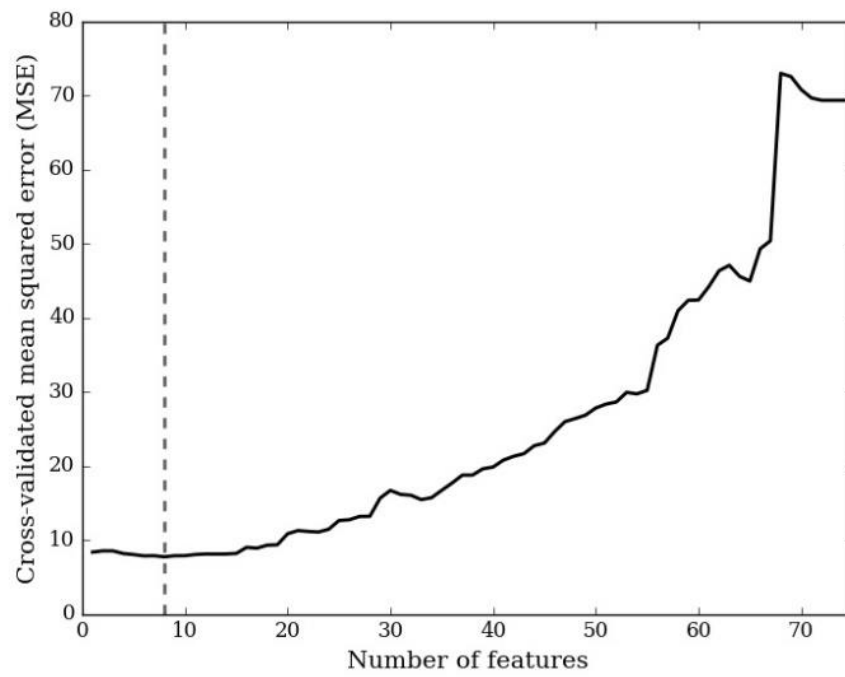
Figure 1 Recursive feature elimination for BF% estimation in 0-8 months. Figure shows the mean squared error minimisation across all 25 wavelengths at three locations; subscapular, triceps and thigh.

Table I Study population characteristics. Of the 651 infants aged 0-24 months enrolled, 416 (63.9%) had available DXA measurements. Of these, 370 (88.9%) had complete NIR and anthropometric measurements.

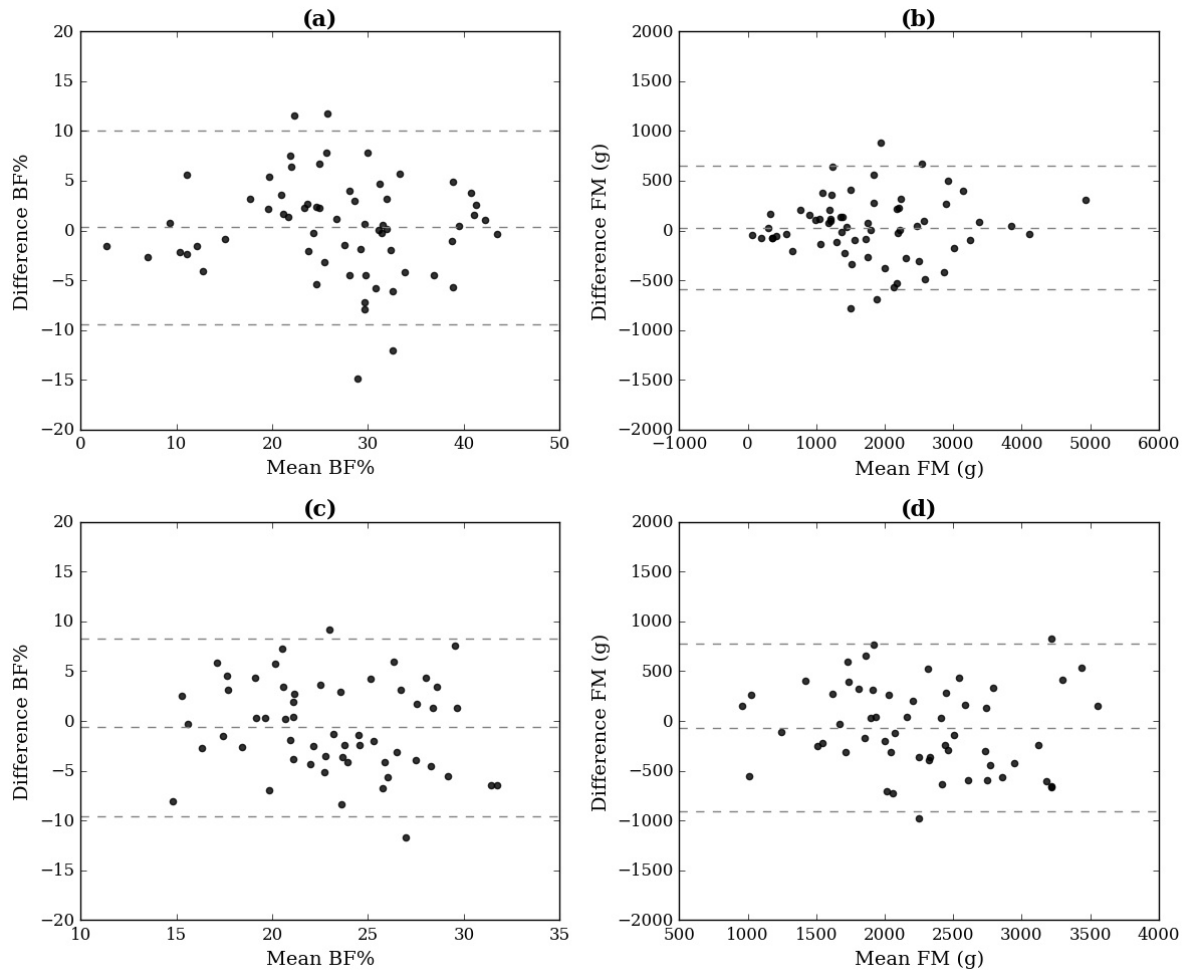| Variable	0-8 months	9-25 months
n	201	169
Age (months)	3.1 ± 2.1	15.4 ± 4.8
Male %	52.7	47.3
Weight (g)	6060 ± 1760	9790 ± 1580
Length (cm)	59.4 ± 6.9	75.8 ± 6.3
Fat mass (g) (DXA)	1724 ± 953	2365 ± 842
Body fat % (DXA)	26.4 ± 9.8	23.9 ± 6.6
Mid-upper arm circumference (cm)	13.1 ± 2.0	14.9 ± 1.8
Subscapular skinfold (cm)	9.0 ± 2.4	9.1 ± 4.7
Flank skinfold (cm)	10.6 ± 3.4	10.2 ± 3.0
Thigh skinfold (cm)	18.5 ± 5.1	18.5 ± 4.3
Triceps skinfold (cm)	9.0 ± 2.1	9.5 ± 2.0

Figure 2 Bland-Altman plots for body fat percentage and fat mass. Panels (a) and (b) correspond to the body fat percentage (BF%) agreement of NIR with DXA and fat mass (g), respectively for the 0-8 month group. Panels (c) and (d) are the corresponding plots for the 9-25 month group.

Table 2 Model wavelengths and coefficients for each age group. Scan locations are shown with wavelengths in nm are shown in parentheses. Reflection based features were generated by determining the difference with reflection at 970 nm and normalising to that at 930 nm (equation 2).

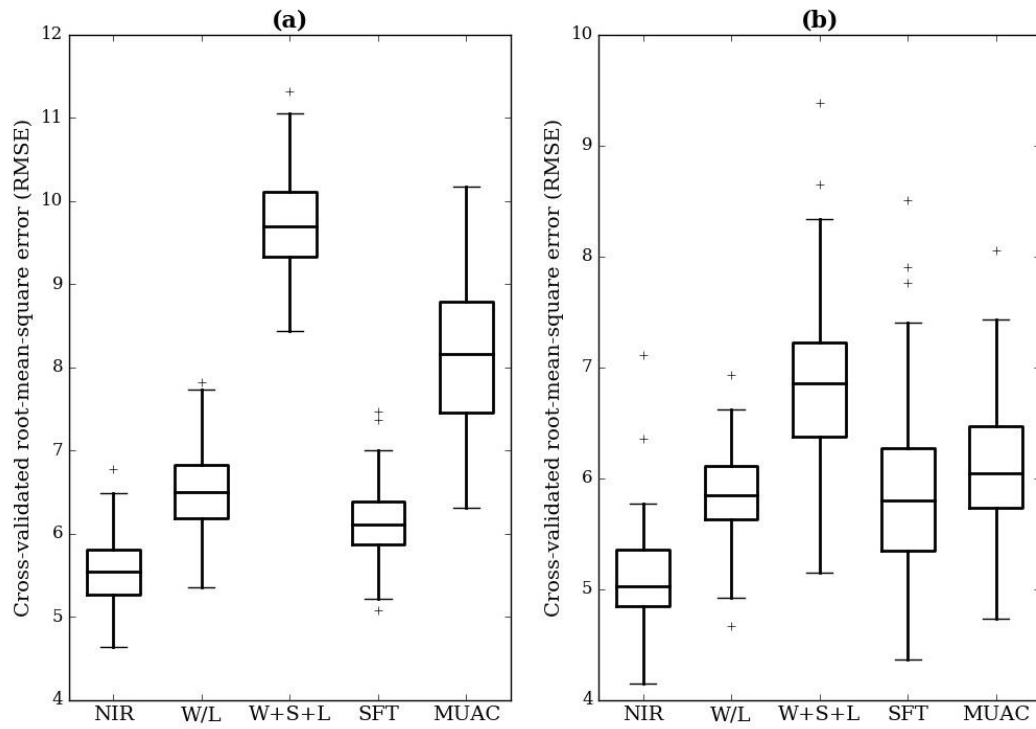|	0-8 months		9-25 months	
Feature 1	subscapular (980)	32.76	subscapular (900)	-6.48
Feature 2	subscapular (990)	-15.54	subscapular (910)	8.14
Feature 3	thigh (1000)	-12.75	subscapular (940)	-13.49
Feature 4	thigh (1010)	6.70	subscapular (950)	15.92
Feature 5	triceps (850)	-7.64	triceps (900)	17.81
Feature 6	triceps (860)	8.44	triceps (910)	-29.38
Feature 7	triceps (1000)	17.34	triceps (920)	10.95
Feature 8	triceps (1010)	-12.68	-	-
Feature 9	W/L	0.2482	W/L^2	13.915
Feature 10	Sex	-3.489	Sex	-1.910
Constant		10.91		11.18

Figure 3 Distribution of root mean squared error (RMSE) from 5-fold stratified cross-validation for (a) 0-8 months and (b) 9-25 months. The boxplots represent the median and interquartile range.

Table 3 Overview of correlation and Bland-Altman agreement. Methods for comparison include weight for length (W/L), a linear regression fit to weight, sex and length (W + S + L), sum of skinfold thicknesses (SFT) and mid-upper arm circumference (MUAC). The training set fit, including correlation (R) and Bland-Altman agreement are shown. Independent Testing set R and agreement are also included. Bland-Altman agreement is displayed as mean difference (95% confidence intervals). Model fit to body fat % was also transformed to fat mass by multiplying by weight (g).

	Target	Model	Train R	Train Bland-Altman	Test R	Test Bland-Altman
Body Fat %	0-8 months	NIR	0.83	0.02 (-10.83, 10.87)	0.863	0.32 (-9.39, 10.03)
		W/L	0.73	-0.00 (-13.29, 13.29)	0.787	0.48 (-11.14, 12.11)
		W+S+L	0.358	-0.00 (-18.16, 18.16)	-0.31	3.04 (-15.80, 21.89)
		SFT	0.773	-0.00 (-12.33, 12.33)	0.801	1.44 (-9.85, 12.73)
		MUAC	0.649	-0.00 (-14.79, 14.79)	0.582	1.38 (-16.22, 18.97)
	9-25 months	NIR	0.729	0.13 (-9.54, 9.80)	0.549	-0.63 (-9.58, 8.31)
		W/L	0.48	-0.00 (-12.38, 12.38)	0.475	-0.10 (-9.09, 8.88)
		W+S+L	0.218	-0.00 (-13.78, 13.78)	0.278	-0.60 (-10.57, 9.38)
		SFT	0.453	-0.00 (-12.59, 12.59)	0.648	0.16 (-7.83, 8.15)
		MUAC	0.435	-0.00 (-12.71, 12.71)	0.481	-0.22 (-8.99, 8.54)
Fat mass	0-8 months	NIR	0.924	-1.78 (-701.34, 697.78)	0.95	27.70 (-590.42, 645.82)
		W/L	0.885	-1.90 (-855.91, 852.10)	0.926	34.53 (-718.91, 787.98)
		W+S+L	0.834	-86.88 (-1156.33, 982.57)	0.924	54.14 (-990.89, 1099.17)
		SFT	0.9	-18.70 (-815.26, 777.85)	0.932	63.68 (-645.15, 772.51)
		MUAC	0.871	-32.11 (-932.79, 868.57)	0.861	70.33 (-977.71, 1118.37)
	9-25 months	NIR	0.849	15.38 (-927.15, 957.91)	0.79	-65.37 (-904.48, 773.74)
		W/L	0.726	17.56 (-1211.19, 1246.32)	0.77	11.78 (-894.34, 917.91)
		W+S+L	0.703	-40.36 (-1442.58, 1361.86)	0.733	-83.82 (-1028.63, 860.99)
		SFT	0.72	-11.13 (-1254.82, 1232.56)	0.835	1.69 (-754.65, 758.03)
		MUAC	0.698	-1.26 (-1277.15, 1274.63)	0.762	-20.70 (-887.45, 846.04)

Appendix I

Measurement conducted on the anterior thigh using the Ocean Optics QEPro NIR device

References

1.	UNICEF., *Facts for life*. 2010: UNICEF.
2.	Lawn, J.E., et al., *4 million neonatal deaths: when? Where? Why?* The lancet, 2005. **365**(9462): p. 891-900.
3.	Barker, D.J., et al., *Growth and living conditions in childhood and hypertension in adult life: a longitudinal study.* Journal of hypertension, 2002. **20**(10): p. 1951-1956.
4.	Levitsky, D.A. and B.J. Strupp, *Malnutrition and the brain: changing concepts, changing concerns.* The Journal of Nutrition, 1995. **125**(8): p. 2212S.
5.	Prado, E.L. and K.G. Dewey, *Nutrition and brain development in early life.* Nutrition Reviews, 2014. **72**(4): p. 267-284.
6.	IAEA, *Introduction to body composition assessment using the deuterium dilution technique with analysis of saliva samples by Fourier transform infrared spectrometry.* IAEA human health series, 2010(12).
7.	Carberry, A.E., et al., *Is body fat percentage a better measure of undernutrition in newborns than birth weight percentiles?* Pediatric Research, 2013. **74**(6): p. 730-736.
8.	Fukuda, D.H., et al., *Validity of near-infrared interactance (FUTREX 6100/XL) for estimating body fat percentage in elite rowers.* Clinical physiology and functional imaging, 2015.
9.	Cassady, S.L., et al., *Validity of near infrared body composition analysis in children and adolescents.* Medicine and science in sports and exercise, 1993. **25**(10): p. 1185-1191.
10.	Mustafa, F.H., et al., *Length-free near infrared measurement of newborn malnutrition.* Scientific reports, 2016. **6**.
11.	Conway, J.M., K.H. Norris, and C. Bodwell, *A new approach for the estimation of body composition: infrared interactance.* The American journal of clinical nutrition, 1984. **40**(6): p. 1123-1130.
12.	Butte, N.F., et al., *Body composition during the first 2 years of life: an updated reference.* Pediatric research, 2000. **47**(5): p. 578-585.
13.	Wood, A.J., et al., *Neonatal length inaccuracies in clinical practice and related percentile discrepancies detected by a simple length-board.* Journal of Paediatrics and Child Health, 2013. **49**(3): p. 199-203.
14.	O'Connor, D.P., et al., *The Bland-Altman method should not be used in regression cross-validation studies.* Research quarterly for exercise and sport, 2011. **82**(4): p. 610-616.
15.	West, J., et al., *Reliability of routine clinical measurements of neonatal circumferences and research measurements of neonatal skinfold thicknesses: findings from the Born in Bradford study.* Paediatric and perinatal epidemiology, 2011. **25**(2): p. 164-171.
16.	Olhager, E. and E. Forsum, *Assessment of total body fat using the skinfold technique in full-term and preterm infants.* Acta Paediatrica, 2006. **95**(1): p. 21-28.
17.	Wrottesley, S., et al., *A comparison of body composition estimates using dual-energy X-ray absorptiometry and air-displacement plethysmography in South African neonates.* European Journal of Clinical Nutrition, 2016. **70**(11): p. 1254-1258.