The Impact of Apnoea on Physiology in Preterm Infants



Yiru Chen

Department of Paediatrics

University of Oxford

Linacre College

yiru.chen@paediatrics.ox.ac.uk

October 22, 2026

Abstract

Preterm infants are born with immature respiratory control and commonly experience apnoea, which can cause clinically important changes in heart rate, oxygen saturation, and cerebral oxygenation. These changes inform bedside intervention in neonatal units, where timely responses may help prevent life-threatening events, and more personalised monitoring could reduce alarm fatigue and clinical burden. However, the physiological consequences of apnoea vary substantially between infants, suggesting that cardiorespiratory instability is not determined by apnoea duration alone. This thesis investigates how apnoea-related instability is modulated by demographic, clinical, and dynamic factors, and whether routinely recorded respiratory signals can characterise respiratory maturity and functional respiratory state in preterm infants.

The first study was a systematic review and meta-analysis of 42 papers, including 1483 neonates and 2399 study sessions. Longer apnoeas were significantly associated with greater reductions in heart rate, oxygen saturation, and cerebral oxygenation, but responses varied widely. Postmenstrual age significantly modulated the relationship between apnoea duration and oxygen desaturation, with younger infants more likely to desaturate after shorter apnoeas.

The second study analysed continuous physiological recordings from 146 preterm infants, identifying 181,511 apnoeas longer than 5 seconds. Longer apnoeas were associated with greater risk of cardiorespiratory instability, but variability remained substantial. Mixed-effects modelling showed that demographic, clinical, and dynamic factors, including age, sex, weight,

ventilation mode, baseline heart rate and oxygen saturation, and recent apnoea clustering, influenced instability risk. XGBoost models predicted apnoea-related instability with a balanced test accuracy of 75.8% using all features, compared with 66.0% using only clinical and demographic features.

The third study evaluated whether respiratory features derived from impedance pneumography and inter-breath interval signals could estimate respiratory maturity. Regression models predicted postmenstrual age only modestly, but classification models distinguished broad maturity groups, suggesting that respiratory dynamics encode maturational information. Morphine administration was also associated with significant respiratory feature changes and detectable shifts in functional respiratory age, particularly in infants with severe clinical reactions.

Together, these studies show that respiratory instability and maturity in preterm infants are shaped by multiple interacting factors. Overall, this thesis provides a foundation for personalised approaches to neonatal respiratory monitoring.

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Caroline Hartley, Mauricio Villarroel, and Luc Berthouze, for their continuous guidance, encouragement, and support throughout my DPhil. Your expertise, patience, and generosity have shaped both my research and my development as a researcher. Without your mentorship, I would not have been able to complete this DPhil.

I am also deeply grateful to all my colleagues in the Paediatric Neuroimaging Group. Thank you for your collaboration and kindness, both professionally and personally. I have truly enjoyed working with all of you, and the time we have shared will remain one of the most treasured memories of my life.

My heartfelt thanks go to all the neonates and their families in the neonatal units who participated in the research projects that contributed to the datasets used in this thesis. Without your participation, advances in neonatal research would not be possible.

To my husband, thank you for always being by my side. Thank you for taking such good care of me and for supporting me through every challenge along this journey. Your love, patience, and unwavering belief in me gave me the strength to overcome difficult moments, celebrate every achievement, and keep moving forward.

This DPhil has been one of the most important journeys of my life. I joined the Department of Paediatrics in 2023 because I believed that babies represent the hope of the future. Helping to

protect a baby from illness, pain, or death is not only about saving a single life, but also about giving hope and happiness to an entire family. Although my own contribution is only a small part of a much larger effort, I believe that, through the collective dedication of researchers around the world, neonatal care and clinical research will continue to advance and improve.

Wherever my future career may lead, I sincerely hope that paediatric research will continue to flourish across the world, and that more families in neonatal units will be able to return home with hope and joy.

Contents

1 Introduction **1**

1.1 Thesis motivation 1

1.2 Thesis structure 3

1.3 Background 5

1.3.1 Neonatal respiratory system 5

1.3.2 Apnoea of prematurity as a marker of respiratory immaturity 9

1.3.3 Physiological variability and individualised assessment 11

1.3.4 Cerebrovascular regulation and vulnerability in preterm infants 13

1.3.5 Clinical monitoring in the Neonatal Unit 15

1.3.6 Age and functional maturity 16

1.3.7 Challenges and considerations in neonatal data analysis 19

1.3.8 Machine learning models in clinical studies 21

1.4 Summary 23

1.5 Aims and hypotheses 24

2 General methods **26**

2.1 Data Collection 26

2.2 Datasets 27

2.2.1 Dataset 1 - Apnoea data 27

2.2.2 Dataset 2 - NeoVital data 31

2.2.3 Dataset 3 - POPPI data 32

2.3 Data pre-processing 34

2.4 Operational definition of bradycardia in preterm infants 35

3 Apnoea duration and changes in cardiorespiratory and cerebrovascular responses 37

3.1 Introduction 37

3.2 Methods 39

3.2.1 Eligibility criteria 39

3.2.2 Information sources and search strategy 40

3.2.3 Selection process 40

3.2.4 Data collection process and data items 42

3.2.5 Study risk of bias assessment (quality assessment) 43

3.2.6 Data synthesis 44

3.3 Results 46

3.3.1 Relationship between apnoea duration and change in heart rate 52

3.3.2 Relationship between apnoea duration and change in oxygen saturation 57

3.3.3 Relationship between apnoea duration and change in cerebral oxygenation 60

3.3.4 Relationship between apnoea duration and change in cerebral blood volume 61

3.3.5 Factors which modulate the physiological response to apnoea 61

3.3.6 Quality assessment 68

3.4 Discussion 70

4 Risk of apnoea-related cardiorespiratory instability 108

4.1 Introduction 108

4.2 Methods 110

4.2.1 Study design and data collection 110

4.2.2 Data analysis 112

4.3 Results 121

4.3.1 Do the results of the systematic review hold in a contemporary data set? 121

4.3.2 Factors modulating the apnoea-associated cardiorespiratory instability . 123

4.3.3 Cardiorespiratory instability from apnoea can be predicted using machine learning 128

4.4 Discussion 135

5 Assessing preterm infants' functional respiratory age 146

5.1 Introduction 146

5.2 Methods 149

5.2.1 Study design and modelling framework 149

5.2.2 Data processing 150

5.2.3 Correlation analysis 151

5.2.4 Cross-validation strategy 152

5.2.5 Prediction of PMA 153

5.2.6 Feature selection 158

5.2.7 Assessing the effect of morphine on respiratory function 159

5.3 Results 162

5.3.1 Relationship between respiratory features and PMA 162

5.3.2 Can we predict PMA from respiratory features? 166

5.3.3 Morphine affects respiratory function in preterm infants 176

5.4 Discussion 181

6 Conclusion and future work 189

6.1 Summary of the main results 189

6.2 Limitations and future directions 192

6.3 Conclusions 197

A Searching strategy for the systematic review 199

B Respiratory features extracted from IP and IBI signals **206**

Bibliography **222**

List of figures

1.1 Schematic representation of respiratory control mechanisms and their developmental differences in infancy. 7

2.1 Data collection flow chart. 29

2.2 Relationship between PMA and heart rate characteristics during and after the start of the apnoea. 36

3.1 Prisma flow chart showing the study selection process 47

3.2 Histograms of the number of neonates and recordings in the included studies. . 48

3.3 Geographic regions and study designs of the included studies. 49

3.4 The percentage change in heart rate during apnoea compared with apnoea duration. 54

3.5 The percentage change in heart rate during apnoea compared with apnoea duration. 54

3.6 The percentage of apnoea events accompanied by bradycardia against apnoea duration. 55

3.7 The percentage of apnoea events accompanied by bradycardia against apnoea duration. 56

3.8 The percentage change in oxygen saturation during apnoea compared with apnoea duration. 58

3.9 The percentage change in oxygen saturation during apnoea compared with apnoea duration. 59

3.10 The change in cerebral haemoglobin difference (cHbD) during apnoea against apnoea duration. 62

4.1 Flowchart of the modelling procedure. 113

4.2 Physiological responses compared with apnoea duration. 122

4.3 Influence of demographic and clinical factors on apnoea-related bradycardia and/or desaturations. 124

4.4 Influence of dynamic variables on apnoea-related bradycardia and/or desaturations. . . . 126

4.5 Confusion matrices of the XGBoost model predicting apnoea-related bradycardia events. . . . 129

4.6 Confusion matrices of the XGBoost model predicting apnoea-related desaturation events. . . . 130

4.7 Confusion matrices of the XGBoost model predicting apnoea-related bradycardia and/or desaturation events. . . . 131

4.8 Pairwise heatmaps of the predicted risk of bradycardia and/or desaturation related to apnoeas between 5 and 10 seconds 133

4.9 Pairwise heatmaps of the predicted risk of bradycardia and/or desaturation related to apnoeas between 10 and 20 seconds 134

4.10 Pairwise heatmaps of the predicted risk of bradycardia and/or desaturation related to apnoeas longer than 20 seconds) 135

4.11 Example usage of the personalised apnoea alarm tool. . . . 136

5.1 Scatter plots showing the relationship between PMA and respiratory features that were significantly associated with PMA in the mixed-effects models after FDR correction. . . . 164

5.2 Mean respiratory feature values plotted against PMA after grouping recordings into 2-week PMA intervals for features that were significantly associated with PMA in the mixed-effects models after FDR correction. . . . 165

5.3 Predicted versus actual PMA obtained using six machine learning regression models. . . . 167

5.4 Predicted versus actual PMA obtained using six machine learning regression models with calibration. . . . 169

5.5 Predicted versus actual PMA obtained using six machine learning regression models with select k-best features selection and calibration. . . . 171

5.6 Predicted versus actual PMA obtained using six machine learning regression models with PCA features selection and calibration. . . . 172

5.7 Predicted versus actual PMA obtained using six machine learning regression models for one-hour epoched data, with calibration. . . . 174

5.8 Predicted versus actual PMA obtained using six machine learning regression models for one-hour epoched data, with calibration. . . . 175

5.9 Confusion matrices for the six classification models predicting maturational category from respiratory features. 177

5.10 Selected respiratory features plotted against time relative to drug administration. 180

5.11 Functional respiratory age gap against time relative to drug administration . . . 181

List of tables

2.1 Demographic and clinical variables of all infants and recording sessions with eligible apnoea. 30

2.2 Basic demographic information for all infants and recording sessions included in Dataset 2. 32

2.3 Demographic information for all infants included in Dataset 3. 33

3.1 Demographic details for the studies associated with each outcome measures. . . 50

3.2 Definition of apnoea, desaturation and bradycardia used across the studies. . . . 52

3.3 Results of the statistical analysis investigating if apnoea type modulates the relationship between bradycardia incidence and apnoea duration. 64

3.4 Results of the statistical analysis investigating if apnoea type modulates the relationship between oxygen saturation and apnoea duration. 65

3.5 Results of the statistical analysis investigating if PMA and apnoea type modulates the relationship between oxygen saturation and apnoea duration. 67

3.6 Results of the statistical analysis investigating if PMA modulates the relationship between heart rate drops and apnoea duration. 68

3.7 Overall risk of bias assessment. 70

3.8 Summary of included studies reporting the relationship between apnoea duration and change in heart rate. 77

3.9 Summary of included studies reporting the relationship between apnoea duration and change in oxygen saturation. 92

3.10 Summary of included studies reporting the relationship between apnoea duration and change in cerebral oxygenation. 101

3.11 Summary of included studies reporting the relationship between apnoea duration and change in cerebral blood volume. 102

3.12 JBI Critical Appraisal Checklist for the included cross-sectional studies (Part A). 103

3.13 JBI Critical Appraisal Checklist for the included cross-sectional studies (Part B). 104

3.14 JBI Critical Appraisal Checklist for the included cohort studies. 105

3.15 JBI Critical Appraisal Checklist for the included case control studies. 107

3.16 JBI Critical Appraisal Checklist for the included case report study. 107

4.1 Summary of the three binary classification models. 117

4.2 The parameters and instances explored during the hyperparameter optimisation. 117

4.3 Details of all included apnoea episodes. 121

4.4 Results from univariate mixed-effects models examining the effects of individual factors and their interactions with apnoea duration on the likelihood that an apnoea is associated with bradycardia/desaturation. 127

4.5 Performance of the optimised XGBoost models. 128

4.6 Comparison of model features between apnoeas with the highest and lowest 10% predicted risk of bradycardia and/or desaturation at duration between 10 and 20 seconds.. 132

4.7 Comparison of model features between apnoeas with the highest and lowest 10% predicted risk of bradycardia and/or desaturation at duration between 10 and 20 seconds. 133

4.8 Comparison of model features between apnoeas with the highest and lowest 10% predicted risk of bradycardia and/or desaturation at duration longer than 20 seconds. 134

5.1 Correlation between respiratory features and PMA for the 15 features that were found to be statistically significant. 163

5.2 Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, without calibration. 166

5.3 Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, with calibration. 168

5.4 Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, with Select K-Best feature selection and calibration. 170

5.5 Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, with PCA feature selection and calibration. 170

5.6 Result metrics for the six regressors used to predict PMA from respiratory features for recordings with one-hour epochs, without calibration. 173

5.7 Result metrics for the six regressors used to predict PMA from respiratory features for recordings with one-hour epochs, with calibration. 174

5.8 Performance metrics of the three-class classification models for predicting maturational category from respiratory features. 177

5.9 Results of paired t-tests comparing respiratory feature values before and after morphine administration. 179

Glossary

Acronyms/Abbreviations

AOP apnoea of prematurity

Apnoea a pause in breathing

BPD bronchopulmonary dysplasia

Bradycardia an abnormally slow heart rate

CPAP continuous positive airway pressure

ECG electrocardiogram

FiO_2 fraction of inspired oxygen

FRC functional residual capacity

GA gestational age

HR heart rate

Hypoxaemia an abnormally low level of oxygen in the blood

IBI inter-breath-interval

IP impedance pneumogram

NICU neonatal intensive care unit

NIRS near-infrared spectroscopy

PMA post-menstrual age

PNA post-natal age

RDS respiratory distress syndrome

SpO_2 peripheral oxygen saturation

SVIA self-ventilating in air

List of publications

The work presented in this thesis has resulted in the following publications.

Chapter 3 is primarily based on a manuscript published in *Pediatric Research*, for which I was the first author [1].

Chen, Y., Zandvoort, C. S., Baxter, L., *et al.* (2025). *Apnoea duration and changes in cardiorespiratory and cerebrovascular responses in preterm neonates: A systematic review and meta-analysis*. *Pediatric Research*. https://doi.org/10.1038/s41390-025-04496-x

Chapter 4 is primarily based on a second manuscript for which I was the first author [2]. This manuscript is currently under review at *Pediatric Research*, and a minor revision decision has been received.

Chen, Y., Ketheeswaranathan, V., Fordington, S., Baxter, L., Stevens, F., Zandvoort, C. S., Gawthorpe, R., Villarroel, M., Berthouze, L., Hartley, C. *Risk of apnoea-related cardiorespiratory instability in preterm infants is modulated by clinical, demographic and dynamic indicators*. Under review at *Pediatric Research* (minor revision).

A preprint of this manuscript is available on medRxiv (2026): https://doi.org/10.64898/2026.05.13.26353101

Chapter 1

Introduction

1.1 Thesis motivation

Globally, approximately 10% of neonates are born preterm (before 37 weeks of gestation) [3]. Preterm birth is prevalent and associated with increased morbidity and mortality in the neonatal period [4]. Among many physiological systems affected by preterm birth, the respiratory system is particularly vulnerable due to the immaturity of neural control and lungs and chest walls. Therefore, respiratory instability is very common in preterm infants. Clinical monitoring and management of neonatal respiratory function are a key focus in neonatal units.

Apnoea, also known as breathing cessation, is a common respiratory complication affecting over 50% of preterm infants, especially those who are born extremely preterm (before 28 weeks of gestation) [5]. Apnoea can occur multiple times per day and can lead to significant physiological changes, including alterations in heart rate, blood oxygen saturation, cerebral oxygenation, and cerebral blood volume [6–9]. While most episodes resolve spontaneously or with minimal intervention, recurrent or prolonged episodes may have clinical significance and require appropriate monitoring and management. Additionally, apnoea has been linked to long-term complications, such as an increased incidence of retinopathy of prematurity (ROP) [10]

and cognitive deficits later in life [11]. These may be related to the extent of apnoea-induced hypoxia or brain activity changes [12].

In current clinical practice, respiratory function and physiological instability are usually assessed according to population-based definitions and fixed thresholds. For example, apnoea is usually defined as the cessation of breathing for more than 20 seconds, or more than 10 seconds if accompanied by bradycardia (an abnormally slow heart rate) or low saturation [13]. The definition is not adjusted according to the infant's age, weight, sex, ventilation mode, or the medications taken. However, it has been shown that short pauses in breathing can sometimes lead to large changes in physiology, and other times much longer pauses in breathing not leading to significant instability [7]. This suggests that the current clinical practice may not reflect the significant variability observed within and between individual infants. Infants with different demographic status may exhibit different respiratory patterns and trajectories of respiratory maturation.

Factors such as the age of the neonate may alter physiological responses to an apnoea, with older babies able to tolerate longer periods of apnoea without changes in physiology. Other co-morbidities and medication can also have an impact. For example, caffeine reduces the incidence of apnoea and intermittent hypoxaemia (an abnormally low blood oxygen levels) [14–16]. Hypoxaemia is associated with impact on brain maturity and longer term neurodevelopmental outcome. A better understanding of the factors which modulate the impact of apnoea on physiology is crucial to identify neonates at the highest risk of adverse physiological changes. Ultimately, this could lead to the development of predictive models that can guide treatment or facilitate closer monitoring for high-risk neonates. In addition to apnoea, broader respiratory

function can vary considerably between infants, and within the same infant as they grow and develop.

In parallel, respiratory maturation in preterm infants is often inferred indirectly through gestational age (GA), postmenstrual age (PMA), or broad clinical milestones, such as requirement for respiratory support and medication. However, many clinical milestones are driven by local guidelines and treatment pathways and do not always necessarily reflect the condition of the baby themselves. Therefore, such measures provide limited insight into the functional state of the respiratory system and may fail to capture individual developmental trajectories. Advances in signal processing and data-driven modelling offer an opportunity to quantify the respiratory function and maturity through utilisation of physiological parameters which are already routinely recorded for most infants, without causing any burden to them and the clinical team.

Overall, the motivation of this thesis is to improve the understanding and assessment of respiratory maturation and physiological instability in preterm infants through the analysis of cardiorespiratory signals. By focusing on individual variability rather than population averages, this work aims to support the development of more personalised approaches to neonatal respiratory monitoring, potentially enabling more timely and individualised care.

1.2 Thesis structure

This thesis is organised into six chapters, each addressing a specific aspect of neonatal respiratory monitoring and analysis, with a particular focus on apnoea and respiratory function in preterm infants.

Chapter 1 introduces the clinical and physiological background underpinning this research. It summarises the current understanding of the neonatal respiratory system, with emphasis on respiratory control and lung development in preterm infants. The chapter also reviews the pathophysiology and clinical significance of apnoea of prematurity, alongside its associated cardiorespiratory consequences. In addition, existing methods for monitoring neonatal respiratory function in the neonatal intensive care unit are discussed, highlighting their limitations and motivating the need for improved analytical approaches.

Chapter 2 describes the datasets utilised throughout this thesis and outlines the general methodological framework adopted. This chapter details the characteristics of the study populations, data acquisition procedures, and signal processing pipelines. It also introduces common data preprocessing steps and analytical methods that are shared across subsequent chapters, providing a methodological foundation for the experimental work presented later in the thesis.

Chapter 3 presents a systematic review and meta-analysis examining the relationship between apnoea duration and associated changes in cardiorespiratory and cerebrovascular responses in neonates. This chapter synthesises existing evidence from the literature, quantifies reported associations, and identifies key gaps and inconsistencies in current knowledge.

Chapter 4 investigates factors influencing cardiorespiratory instability in preterm infants, with a particular focus on predicting apnoea-related bradycardia and oxygen desaturation. Using physiological signals and demographic features, this chapter explores predictive modelling approaches and evaluates their ability to identify clinically significant apnoea events.

Chapter 5 focuses on the estimation of infants' functional respiratory age using respiratory

signals. This chapter examines whether respiratory dynamics alone can provide meaningful insight into respiratory function, and assesses the potential of data-driven models to capture developmental trajectories in preterm infants.

Finally, Chapter 6 provides an overall synthesis of the findings presented in this thesis. It discusses the potential clinical and methodological implications of the results, outlines key limitations of the work, and offers recommendations for future research directions in neonatal respiratory monitoring and analysis.

1.3 Background

1.3.1 Neonatal respiratory system

A primary function of the human lung is gas exchange, where oxygen diffuses from air into the pulmonary capillary blood and carbon dioxide diffuses from the blood into the alveoli for elimination [17]. Effective gas exchange depends on adequate breathing (ventilation), a spontaneous and automatic motor behaviour generated within the brainstem. Neural networks located in the medulla oblongata, including the respiratory rhythm-generating centres, produce and maintain the basic respiratory rhythm. Ventilation is continuously adjusted through an integrated control system to maintain arterial oxygen (PaO_2), carbon dioxide ($PaCO_2$), and pH within narrow physiological ranges [18].

Central chemoreceptors, located mainly in the ventral side of the medulla oblongata, regulate breathing by sensing changes in the pH of cerebrospinal fluid (CSF), which reflects arterial

carbon dioxide (CO_2) levels [18, 19]. Carbon dioxide can easily diffuse through the blood-brain barrier, forming carbonic acid and increasing hydrogen ion (H^+) concentration in CSF. This rise in H^+ stimulates central chemoreceptors, increasing respiratory rate and tidal volume to enhance CO_2 elimination and restore acid–base balance [19–21].

Peripheral chemoreceptors, located in the carotid and aortic bodies, detect changes in PaO_2, $PaCO_2$, and arterial pH. These receptors transmit afferent signals via the glossopharyngeal and vagus nerves to the medullary respiratory centres. In response to hypoxaemia, peripheral chemoreceptors rapidly increase ventilatory drive, while also contributing to responses to hypercapnia (elevated arterial carbon dioxide levels) and metabolic acidosis (a decrease in blood pH due to the accumulation of metabolic acids or loss of bicarbonate), thereby supporting cardiopulmonary homeostasis [17, 20, 22].

Ventilation is achieved through a coordinated contraction of the diaphragm and other respiratory muscles, generating pressure gradients that drive airflow [17, 23]. The lungs and chest wall together form the mechanical pump of respiration. Mechanical properties include characteristics such as lung compliance, airway resistance, chest wall compliance, and functional residual capacity (FRC), the volume of air remaining in the lungs at the end of passive expiration. Adequate FRC is essential for maintaining stable gas exchange by helping to avoid atelectasis (collapse of alveoli) between breaths [23, 24]. Figure 1.1 shows the respiratory control mechanisms in infants.

Lung development progresses through distinct stages during gestation. Alveoli, which are very small air sacs in the lungs, begin formation during the late canalicular stage (16–25 weeks'

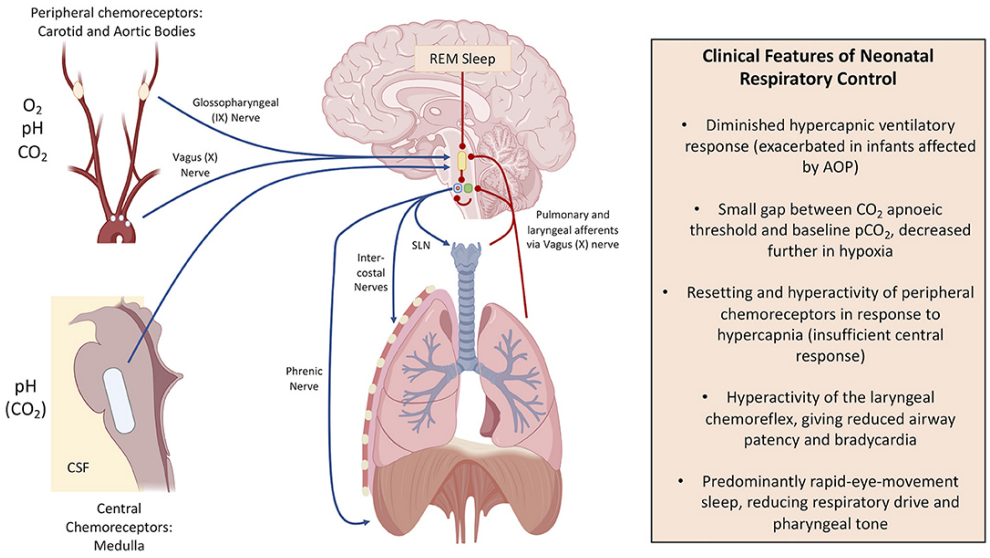

Figure 1.1: Schematic representation of respiratory control mechanisms and their developmental differences in infancy. Reproduced from "Apnoea of prematurity and neurodevelopmental outcomes: Current understanding and future prospects for research" by M Williamson, R Poorun, and C Hartley, 2021, Frontiers in Pediatrics, 9:755677, (https://doi.org/10.3389/fped.2021.755677). Copyright 2021 by the Authors. Licensed under CC BY 4.0 [25].

gestation) and saccular stage (24–36 weeks' gestation), and accelerates during the alveolar stage, which extends from approximately 36 weeks' gestation into early childhood [26, 27]. It has been found that lung volume increases by four times between 29 weeks of gestational age and full term age, and further doubles within four months after birth [28]. Surfactant is a mix of mainly phospholipids produced by Type 2 pneumocytes from around 24 weeks' gestation, peaking around 32 weeks, and reduces the risk of atelectasis by decreasing the surface tension of alveoli. Surfactant deficiency is common in preterm infants, and contributes to the work of breathing in preterm infants, as seen in respiratory distress syndrome (RDS) [29]. Therefore, respiratory system in preterm infants remain structurally and functionally immature at birth for potentially a considerable amount of time, depending on the degree of prematurity [30].

In addition, the neonatal chest wall is highly compliant due to incomplete ossification of the

ribs and reduced muscle strength. In preterm infants, this compliance is even greater. A highly compliant chest wall combined with relatively stiff immature lungs reduces FRC and increases the work of breathing [31–33]. Neonatal ribs are more elevated and horizontal relative to the adult ribcage, reducing their effect on intra-thoracic volume. Also, the neonatal diaphragm is relatively flat and has a smaller attachment zone compared to the adult, further reducing their ability to increase intra-thoracic volume. These mechanical constraints increase susceptibility to atelectasis and contribute to breath-to-breath variability in tidal volume and ventilation efficiency.

In preterm infants, both neural control and mechanical properties are immature. Central and peripheral chemoreceptive pathways exhibit reduced or unstable sensitivity to hypoxia and hypercapnia, and integration of sensory input within the brainstem is not fully developed [34, 35]. The respiratory system is less stable, meaning that small perturbations in blood gases or sensory feedback may result in exaggerated variability or transient cessation of breathing [36].

Because neural regulation, chemosensitivity, lung structure, and respiratory muscle strength continue to mature after birth, respiratory function in preterm infants is highly dynamic. Changes in respiratory timing, variability, and stability may occur over days to weeks as these systems develop [34, 37]. Therefore, within the same infant, their respiratory patterns may look very different at different time points. In addition, different infants of the same GA or PMA may demonstrate markedly different respiratory trajectories, reflecting heterogeneous maturation of neural control, lung structure, and the potential impacts of different clinical exposures [7, 38].

1.3.2 Apnoea of prematurity as a marker of respiratory immaturity

Apnoea of prematurity (AOP) is one of the most frequently diagnosed conditions in neonatal units and reflects respiratory immaturity in preterm infants [13]. It is commonly classified into three types based on the presence or absence of upper airway obstruction [39, 40]. Central apnoea is defined as a cessation of breathing effort. Obstructive apnoea occurs when airflow is blocked despite continued respiratory effort, typically due to upper airway obstruction or collapse, particularly at the pharyngeal level. Mixed apnoea comprises features of both central and obstructive types, typically beginning as central apnoea and subsequently developing an obstructive component, although the reverse sequence may also occur.

The incidence and severity of AOP increase with decreasing gestational age, reflecting the developmental vulnerability of respiratory structure, function and regulatory processes in early life [41]. The key stages of lung maturation, including alveolar formation, vascular development and the production of surfactants, continue even after the preterm birth [26, 42]. Therefore, apnoea events can be regarded as the result of the interaction between ongoing lung development, limited physiological reserves and immature neurological control.

A defining feature of apnoea of prematurity is its marked heterogeneity. Infants of similar GA or PMA may exhibit substantial differences in apnoea frequency, duration, clustering behaviour, and associated physiological responses [7, 43]. Some infants tolerate prolonged apnoeic pauses with minimal cardiorespiratory disturbance, whereas others develop pronounced bradycardia and/or oxygen desaturation following relatively brief events. This variability likely reflects differences in maturational trajectories, baseline respiratory stability, and the influence

of antenatal and postnatal factors such as infection and other comorbidities, respiratory support, and pharmacological treatment [13, 44–46].

In addition to its direct impact, AOP is also associated with potential long-term consequences. Recurrent episodes of hypoxaemia and bradycardia may expose the developing brain to intermittent reductions in oxygen delivery, which may lead to impaired neurodevelopmental outcomes [11]. Several studies have reported associations between frequent intermittent hypoxaemia in preterm infants and adverse cognitive, behavioural, and visual outcomes later in childhood [38, 47, 48]. Although causality remains difficult to establish due to confounding by overall prematurity and complications of prematurity, these findings suggest that the physiological burden of apnoea may extend beyond the neonatal period. Consequently, understanding not only the frequency and duration of apnoea events but also their physiological impact may be important for evaluating longer-term vulnerability.

As respiratory control and pulmonary function mature, the overall burden of apnoea typically declines [41]. However, this maturation does not follow a uniform or linear course. Transitional periods of instability may occur, resulting in fluctuating apnoea characteristics over time [49]. These dynamic patterns make the static definition of the severity of apnoea based on population-level thresholds insufficient. A more comprehensive definition that is tailored to an individual baby at a given point in their development would be beneficial.

Linking apnoea events to potential maturity rather than treating them as isolated events is of great significance for risk stratification, monitoring strategies, and interpretation of respiratory signals in preterm infants. This approach may facilitate more timely assessment and support

individualised care. It also supports the motivation of the subsequent chapters of this thesis, which study the physiological consequences of apnoea and explore data-driven methods for personalised respiratory assessment.

1.3.3 Physiological variability and individualised assessment

Physiological variability is a fundamental characteristic of biological systems and reflects the dynamic interaction between regulatory mechanisms, environmental influences, and developmental processes [50]. In preterm infants, such variability is often pronounced due to immaturity of control systems and rapid postnatal adaptation [51]. Rather than representing noise alone, variability can provide important insight into underlying system function and maturation.

Cardiorespiratory signals in preterm infants exhibit substantial inter- and intra-individual variability across multiple timescales [49, 52, 53]. This heterogeneity reflects differences in neural maturation, lung mechanics, and clinical course. However, several commonly used diagnostic definitions and treatment thresholds remain based on population-derived criteria. For example, apnoea of prematurity is typically defined using fixed duration thresholds (e.g. cessation of breathing for more than 20 seconds, or more than 10 seconds if accompanied by bradycardia or desaturation), caffeine therapy is often initiated according to standardised clinical definitions (e.g often related to GA and PMA), and chronic lung disease is defined by oxygen requirement at a specified PMA. While such criteria are practical and widely adopted, they assume comparable physiological baselines among infants of similar GA, PMA, and/or birth weight.

Increasing evidence suggests that identical cardiorespiratory events (e.g. apnoea, hypoxaemia and bradycardia) may have markedly different physiological consequences across infants [7, 38]. Therefore, threshold-based criteria may fail to identify some clinically significant events from mild variations in some infants.

Individualised assessment approaches seek to account for this variability by interpreting physiological measurements relative to an infant's own baseline and developmental trajectory. In other areas of neonatal and critical care, personalised monitoring strategies, such as heart rate variability–based surveillance for early detection of sepsis and predictive monitoring of cardiorespiratory deterioration, have demonstrated improved sensitivity for identifying adverse physiological events compared with fixed population-based thresholds [54]. The KP sepsis calculator provides clinicians with an individualised interpretation of physiological features based on an infant's pre-existing demographic and antenatal risk factors to improve identification of early onset sepsis [55]. These approaches rely on patient-specific signal characteristics rather than static cut-offs, and align with the broader shift towards precision medicine, which utilises individual-level data to refine risk stratification and clinical decision-making [56].

Quantitative analysis of physiological variability provides a framework for such individualised assessment. Features derived from continuous signals, including breathing timing and variability indices, may offer insight into functional maturity that cannot be captured by static age-based indicators alone [50]. Within the context of this thesis, physiological variability is therefore treated as an informative signal rather than a confounder. By examining individual respiratory dynamics in relation to developmental maturity, this work aims to move beyond one-size-fits-all definitions towards more personalised, physiology-driven assessment frame-

works for preterm infants.

1.3.4 Cerebrovascular regulation and vulnerability in preterm infants

Cerebrovascular regulation plays a critical role in maintaining adequate cerebral oxygen delivery [57]. In neonates, cerebral blood flow is governed by an integrated control system involving systemic arterial blood pressure, arterial oxygen and carbon dioxide tensions, metabolic demand, and neurovascular coupling [58, 59]. In mature individuals, one of the principal protective mechanisms within this system is cerebral autoregulation, which maintains relatively stable cerebral blood flow across a range of systemic blood pressures by adjusting cerebrovascular resistance [60].

In preterm infants, several components of this regulatory system are immature. Cerebral autoregulatory capacity is often impaired or inconsistent, and cerebrovascular responses to changes in carbon dioxide and oxygen tension may be exaggerated or unstable [58, 61]. Therefore, the cerebral circulation of many preterm infants shows pressure passivity, which means that fluctuations in surface blood pressure are more directly transmitted to cerebral blood flow. In addition, variations in oxygenation and carbon dioxide concentration can lead to disproportionate changes in cerebrovascular tone. Consequently, cardiorespiratory instability may increase the risk of cerebral hypoperfusion or hyperperfusion, thereby contributing to the vulnerability of the preterm brain [62].

Respiratory pauses can lead to hypoxaemia, hypercapnia, and bradycardia, which may lead to an alteration of cerebral blood flow and oxygen delivery. Experimental and clinical studies

have shown that the changes in arterial oxygen saturation and carbon dioxide levels in preterm infants are closely related to the cerebral haemodynamic response [63, 64]. The response of cerebral blood vessels to these disturbances is highly variable, reflecting differences in maturity status, baseline brain autoregulation, and other clinical conditions.

The vulnerability of the preterm brain to fluctuations in cerebral oxygenation has significant clinical significance. Periods of impaired cerebral perfusion or oxygen delivery are associated with an increased risk of brain injury, including intraventricular haemorrhage and white matter damage [62, 65]. Even though many apnoeic events are transient and self-resolving, they may still have significant impacts on cerebrovascular function, which may be underestimated when their assessment is purely focused on cardiorespiratory parameters [66, 67].

Near-infrared spectroscopy (NIRS) can perform non-invasive and continuous cerebral oxygenation monitoring at the bedside, providing valuable insights into the cerebrovascular responses of preterm infants [68]. Studies using NIRS have demonstrated significant inter-individual variability in cerebral oxygenation responses to hypoxaemia, blood pressure changes, and respiratory instability [67, 69].

These findings suggest that cerebrovascular responses provide complementary and clinically relevant information beyond routine cardiorespiratory measurements. Apnoea-related fluctuations in cerebral blood volume and oxygenation may offer additional insight into physiological vulnerability in preterm infants, particularly given the recognised susceptibility of the developing brain to haemodynamic instability. For this reason, cerebrovascular outcomes were included in the systematic review conducted in this thesis, in order to provide a broader under-

standing of the physiological consequences of apnoea.

1.3.5 Clinical monitoring in the Neonatal Unit

Continuous respiratory monitoring is a central component of care for preterm infants admitted to the NICU. Due to immaturity of respiratory control and lung function, preterm infants are at increased risk of cardiorespiratory instability, including apnoea, bradycardia, and oxygen desaturation [7], all of which can potentially lead to acute and chronic adverse consequences. Early detection of such events is therefore critical to clinical decision-making, escalation of care, and the prevention of adverse outcomes.

In current clinical practice, respiratory monitoring in the NICU typically relies on a combination of impedance pneumography, electrocardiography, pulse oximetry, and bedside observation [13, 38]. Impedance-based methods estimate respiratory activity from changes in thoracic electrical impedance [70], while pulse oximetry provides indirect information on ventilation through peripheral oxygen saturation trends [45]. These methods are widely available and non-invasive, making them suitable for continuous monitoring in vulnerable populations.

However, conventional monitoring methods have important limitations. Impedance pneumography reflects changes in thoracic volume rather than direct airflow or tidal volume and may not reliably distinguish between apnoea and shallow breathing [71]. Signal quality may be affected by motion artefact, electrode displacement, poor skin contact, or interference during nursing care. Pulse oximetry detects downstream changes in oxygen saturation and may lag behind alterations in ventilation or respiratory effort. Moreover, oxygen saturation is influenced by

factors such as perfusion, haemoglobin concentration, and supplemental oxygen administration, potentially obscuring the primary respiratory disturbance.

In addition to technical constraints, alarm systems are commonly configured using fixed thresholds, such as predefined apnoea durations or oxygen desaturation limits applied uniformly across patients [38]. While such threshold-based approaches are simple to implement, they do not account for the substantial variability in respiratory patterns and physiological responses among preterm infants. Consequently, clinically significant events may be missed in some infants, while others may experience frequent alarms that do not require intervention, thereby causing alarm fatigue among clinical staff [72].

Collectively, these limitations suggest that conventional respiratory monitoring provides an incomplete representation of underlying physiological processes. Advances in digital signal acquisition now allow access to high-resolution physiological waveforms in neonatal units. Emerging research indicates that more sophisticated analyses of respiratory and cardiorespiratory signals may improve detection of clinically relevant events, better characterise respiratory instability, and support more individualised assessment of respiratory development, which are tailored to individual infants' demographic, clinical and physiological background [54, 73]. This evolving perspective underpins the data-driven approach adopted in this thesis.

1.3.6 Age and functional maturity

Chronological age is commonly used in neonatal medicine as a proxy for developmental maturity. However, increasing evidence across biomedical research suggests that chronological age

does not always accurately reflect the functional state of dynamic biological systems. Therefore, the concept of biological age emerged as a framework for estimating maturity based on structural, physiological or molecular characteristics rather than time alone [74–76]. Biological age models aim to quantify how developed or aged a system is relative to expected age-related trajectories.

Across medicine, biological age approaches have been developed for multiple organ systems. Examples include epigenetic clocks derived from DNA methylation patterns [77]; heart age or cardiovascular risk age models based on vascular and haemodynamic parameters [78, 79]; and signal-derived physiological age metrics based on heart rate variability, electrocardiographic features, or other functional measures [80–82]. These models show that dynamic biological signals contain information that cannot be predicted by chronological age alone.

The broader implication of biological age models is that ageing or maturity should be treated as a continuous and dynamic process, which is impacted by numerous contributing variables, rather than a fixed attribute determined solely by chronological age. This perspective is widely accepted in ageing models for adults, but is also relevant in preterm infants, where developmental trajectories may diverge substantially from population averages.

In early life, neurodevelopment involves rapid and highly dynamic changes in brain structure and function [83, 84]. In preterm infants, this developmental process unfolds partially outside the intrauterine environment, exposing the immature brain to altered sensory input, physiological stress, and clinical interventions. As a result, two infants of the same PMA may differ substantially in neurodevelopmental maturity. This variability has motivated the development

of quantitative biomarkers capable of estimating functional brain maturity more precisely than chronological age alone.

The concept of brain age represents a neonatal application of the biological age framework. Brain age is typically derived by modelling age-related patterns in structural or functional neuroimaging data and estimating the age that best corresponds to an individual's brain characteristics [85]. Deviations between estimated brain age and chronological age have been associated with altered developmental trajectories and adverse neurodevelopmental outcomes, particularly in preterm populations [86, 87]. In neonatal research, brain age estimation has most commonly been pursued using magnetic resonance imaging (including functional MRI) metrics such as cortical folding, white matter microstructure, and network connectivity [88]. More recently, electroencephalography (EEG)-based approaches have demonstrated promise for assessing functional brain maturity at the bedside, offering high temporal resolution and feasibility for longitudinal monitoring [89, 90].

It is reasonable to extend this framework beyond the brain to other physiological systems, including the respiratory system, which may also exhibit measurable functional maturity that can be quantified independently of PMA. In this thesis, the primary approach to estimating functional respiratory age is based on the assumption that, in clinically stable preterm infants, PMA is associated with a recognisable trajectory of respiratory maturation. Under this assumption, respiratory signal characteristics are used to learn the characteristic respiratory patterns associated with different PMAs, enabling an age-equivalent estimate of functional respiratory maturity to be derived from an infant's respiratory signals. By comparing respiratory-derived maturity estimates with PMA, this work seeks to evaluate whether alternative biological age

markers provide additional information about respiratory maturation beyond that expected from PMA alone.

1.3.7 Challenges and considerations in neonatal data analysis

The analysis of neonatal physiological data presents a number of methodological challenges that distinguish it from adult and paediatric studies. Preterm infants represent a highly heterogeneous population, characterised by rapid developmental change, physiological immaturity, and substantial inter-individual variability. They also often have a high level of medical intervention and often have a number of clinical interventions occurring at any one time, which can impact on their physiology. These factors complicate both the interpretation of physiological signals and the development of robust analytical models.

A key challenge arises from the dynamic nature of neonatal physiological signals. As infants mature, respond to clinical interventions, or experience brief instabilities, respiratory, cardiac, and oxygen saturation signals evolve within a short period of time. As a result, assumptions of stationarity or consistent baseline behaviour are often violated, limiting the applicability of conventional analytical approaches [50]. Signal characteristics may change not only between infants but also within the same infant over time. Therefore, it is necessary to develop methods that are sensitive to developmental trajectories.

Data quality is a further major consideration. Physiological signals recorded in the neonatal intensive care unit are frequently affected by artefacts due to motion, electrode displacement, nursing care, and equipment limitations [70, 91, 92]. Missing data and variable recording du-

rations are common, particularly in long-term monitoring studies. In preterm infants, low signal amplitudes and shallow breathing can further reduce the reliability of standard respiratory measurements, increasing the risk of misclassification or biased feature extraction [71]. The accuracy of clinical and demographic data is often depending on the quality and consistency of documentation, which can differ between clinicians and is subject to human error.

Cohort size and data imbalance also pose significant challenges in neonatal research [93]. Studies involving preterm infants are often limited by small sample sizes, while within-infant data may be unevenly distributed across developmental stages or clinical conditions. This imbalance can lead to overfitting and overly optimistic model performance if not appropriately addressed. Careful selection of validation strategies is therefore essential, particularly when multiple recordings are available from the same infant. Ensuring independence between training and testing data is critical when developing predictive models from neonatal physiological signals. Failure to account for repeated measurements from the same individual can result in information leakage and inflated performance estimates. Nested cross-validation and patient-level validation strategies are recommended to provide more realistic assessments of model generalisability [94].

Interpretability and clinical relevance are additional considerations when applying advanced analytical techniques, including machine learning. While complex models may achieve high predictive performance, their utility in neonatal care depends on transparency, robustness, and alignment with physiological understanding [95]. Models that produce clinically untrustworthy or unexplainable results may hinder rather than support clinical decisions. For practical implementation, results should be readily understandable to clinicians and easily integrated

into routine clinical workflows.

These challenges underscore the need for careful methodological design when analysing neonatal physiological data. Within this thesis, particular attention is paid to data quality, validation strategy, and physiological interpretability. This thesis aims to develop robust and clinically meaningful analytical approaches for studying respiratory function and maturation in preterm infants.

1.3.8 Machine learning models in clinical studies

The increasing availability of large clinical and physiological data has led to growing interest in the application of machine learning models within clinical research. Traditional statistical methods usually rely on predefined assumptions about data distribution and relationships, while machine learning methods aim to directly identify complex nonlinear patterns from data. This capability makes them particularly well suited to clinical settings, where physiological systems are highly dynamic, heterogeneous, and influenced by multiple interacting factors [96, 97].

In clinical studies, machine learning models have been applied across a wide range of tasks, including disease diagnosis, risk stratification, outcome prediction, and phenotyping [98–100]. Supervised learning methods, such as regression and classification models, are commonly used to predict clinical outcomes based on physiological signals, electronic health records, or imaging data [101]. Unsupervised and semi-supervised methods are also employed to reveal the underlying structures in complex datasets [102]. This can enable the identification of new patient subgroups or development trajectories that may not be obvious in traditional analyses

[100, 103].

Physiological time-series data present both opportunities and challenges for machine learning. Continuous signals such as respiratory signals, heart rate, and oxygen saturation contain rich temporal information but are often noisy, non-stationary, and subject to missing data. Machine learning models, especially ensemble methods and tree-based algorithms, have demonstrated robustness against these challenges and are increasingly being applied in clinical physiology research [104, 105].

In neonatal and critical care settings, machine learning has shown promise for early detection of clinical deterioration, prediction of adverse events, and personalised risk assessment [54, 106, 107]. These applications are especially relevant for preterm infants, whose physiological signals exhibit large variability and rapid developmental change. By leveraging individual-level data, machine learning models offer the potential to move beyond population-based thresholds and towards more personalised interpretations of physiological measurements.

Neonatal units also represent favourable environments for the development of machine learning models, as large volumes of routinely collected physiological data are available. In addition, compared with most other intensive care settings, neonatal units include not only infants with complex medical conditions but also a substantial population of clinically stable infants who are establishing feeding and growth. This relatively 'healthy' subgroup provides an important reference population for model training and comparative analysis.

Within this thesis, machine learning models are employed as analytical tools to characterise respiratory-related physiology and maturation in preterm infants. Emphasis is placed on phys-

iologically informed modelling and rigorous evaluation, with the aim of using machine learning to enhance understanding of neonatal respiratory development while supporting clinically meaningful insight.

1.4 Summary

When an infant is born preterm, immaturity of respiratory control and lung mechanics leads to highly dynamic respiratory behaviour in preterm infants [30], while the immaturity of cerebrovascular regulation makes the developing brain sensitive to cardiorespiratory instability [63]. Therefore, apnoea in preterm infants may have a significant impact on their cardiorespiratory and cerebrovascular functions as well as long-term neurodevelopment [7, 44, 62].

However, preterm infants of similar GA or PMA may exhibit significantly different breathing patterns, apnoea frequencies, and physiological responses [49, 52, 53]. These observations pose challenges to the sufficiency of population-based definitions and fixed clinical thresholds. In particular, the cardiorespiratory and cerebrovascular consequences related to apnoea events may not be based solely on the duration of apnoea, and may vary significantly among infants and within the same infant. Clinical respiratory monitoring in neonatal units mainly relies on threshold-based alarms integrated with impedance pulmonary or pulse oximeters [13, 38]. Although these methods can provide convenient and continuous monitoring, they offer limited physiological specificity and are vulnerable to artifacts and false alarms [71].

Physiological variation provides an opportunity to go beyond static definitions and move towards personalised assessment. Continuous respiratory signals contain rich temporal infor-

mation, reflecting the maturity and stability of the functional system. Advancements in neuroimaging and neurophysiological analysis have led to the development of the brain's age framework [85], reinforcing the concept that maturity is a continuous and individualised process. Extending this view to the respiratory system can serve as a marker of physiological development to estimate functional respiratory age [83, 87].

Machine learning and data-driven analytical methods are powerful tools for addressing the complexity of neonatal physiological data [96, 97]. By adapting to nonlinearity, heterogeneity and high-dimensional temporal structures, these methods have shown promise in predicting adverse events, characterising physiological instability and supporting personalised risk assessment in neonatal settings. With rigorous validation, machine learning methods have the potential to enhance scientific understanding and optimise clinical care.

The work proposed in this thesis aims to enhance the understanding of physiological responses related to apnoea, identify the factors modulating cardiorespiratory and cerebrovascular instability, and develop data-driven methods to assess the respiratory function maturity of preterm infants. By focusing on individual differences rather than the average population level, this thesis aims to provide a more personalised and physiologically informed approach for neonatal respiratory monitoring and assessment.

1.5 Aims and hypotheses

The overarching aim of this thesis is to quantitatively characterise respiratory instability and maturation in preterm infants through the analysis of cardiorespiratory and cerebrovascular

signals. In particular, this work seeks to define the relationship between apnoea duration and physiological consequences, identify factors that modulate these responses, and develop data-driven models to capture individual variability in respiratory function.

The specific objectives and associated hypotheses of this thesis are as follows:

- To systematically evaluate the relationship between apnoea duration and associated changes in cardiorespiratory and cerebrovascular physiology in neonates. Based on existing literature, I hypothesise that longer apnoea duration is generally associated with greater physiological disturbance, but the strength and consistency of this relationship vary substantially across studies and populations.

- To identify factors contributing to variability in physiological responses to apnoea events in preterm infants. I hypothesise that demographic and dynamic factors, such as PMA, GA, weight z-score, ventilation mode, medication, baseline heart rate and oxygen saturation, and past apnoea history, modulate the physiological response of apnoea.

- To develop and evaluate predictive models for apnoea-related bradycardia and oxygen desaturation using demographic and dynamic features. I hypothesise that outcomes of apnoea can be predicted using demographic and clinical factors. The model would provide a better insight than the apnoea alarms based on static thresholds.

- To investigate whether respiratory signals alone can be used to estimate functional respiratory age in preterm infants. In clinically stable preterm infants, I hypothesise that features of the respiratory signal can be used to predict PMA.

Chapter 2

General methods

2.1 Data Collection

This thesis involves 3 datasets that will be detailed in Section 2.2. All the datasets were secondary datasets, where participants were selected from a database of data previously recorded by our research group between 2017 and 2025 at the John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. Ethical approval was obtained from the National Research Ethics Service (references: 12/SC/0447, 23/NW/0155, and 19/LO/1085), and informed written parental consent was obtained prior to each study. All studies were in accordance with the standards of the Declaration of Helsinki and adhered to Good Clinical Practice guidelines.

In this database, each infant had continuous vital signs monitoring as part of standard clinical care. Monitoring was conducted using Philips IntelliVue MX800 or MX750 monitors, and data were continuously downloaded using an electronic data capture system (iXtrend, ixitos, Germany). Physiological data including heart rate, oxygen saturation, and respiratory rate (as calculated by the monitor) were recorded at a sampling rate of 0.97 Hz. Additionally, high-resolution physiological signals were captured: the ECG, used to determine heart rate, was

sampled at 250 Hz and recorded via three chest electrodes; the impedance pneumography (IP), used to assess respiratory effort, was sampled at 62.5 Hz via the same chest electrodes; and the photoplethysmography (PPG), used to monitor oxygen saturation and pulse, was sampled at 125 Hz from a probe placed on the infant's hand or foot.

Of note, some infants receiving high-flow oxygen received this using automated oxygen control (delivered via the Vapotherm oxygen assist module (OAM)), which automatically titrates the FiO_2 according to the infant's oxygen saturations to achieve a preset target saturation. This was initiated in our unit from 2021 onwards, and is recommended in all preterm infants $<$32 weeks (until 36 weeks PMA) and infants with PPHN (persistent pulmonary hypertension of the newborn), under clinician discretion. For the rest of the infants receiving high-flow oxygen, FiO_2 is manually controlled by clinical staff.

2.2 Datasets

2.2.1 Dataset 1 - Apnoea data

Dataset 1 was used in Chapter 4, which builds predictive models of apnoea-related bradycardia and desaturation events. I included data collected from previous research studies in which continuous vital-sign recordings were obtained from preterm infants. The eligibility criteria for recording sessions and apnoea episodes were:

- Age: Postmenstrual age (PMA) $<$37 weeks at the time of recording (study focus: preterm infants).

- Mode of ventilation: Infants were included if they were self-ventilating in air (SVIA) or receiving low-flow or high-flow oxygen therapy (the latter delivered via Vapotherm Precision Flow). Infants receiving other forms of respiratory support were excluded to ensure that extracted apnoea episodes and their physiological responses were not influenced by mechanical ventilation.
- Apnoea detection: A recording was included only if at least one apnoea (pause in breathing $>$5 seconds) was detected.
- Stable cardiorespiratory baseline: Apnoea episodes were excluded if, during the 90 seconds before apnoea onset, oxygen saturation fell below 85%, heart rate fell below 100 bpm, or heart rate exceeded 200 bpm. This requirement ensured that included apnoeas began from a reasonably stable cardiorespiratory baseline so subsequent changes could more plausibly be attributed to the apnoea itself.
- Medication: Infants receiving antibiotics, opioids, morphine or fentanyl at the time of the apnoea episode were excluded.

All IBIs lasting between 5 and 40 seconds that were identified by the machine learning algorithms (presented later in Section 2.3) as true apnoeas were included in this study as apnoea events. IBIs exceeding 60 seconds were manually checked before including as apnoea events, as clinical experience indicates these are more likely to be false positives than true apnoea episodes. IBIs between 40 and 60 seconds were manually reviewed if they were associated with less than 20% drop in either heart rate or oxygen saturation, as these were also more likely to be false positives. During the manual review, both the original and filtered IP signals were examined to determine whether the detected pauses in breathing were due to potential

connection issues causing abnormal signals, or genuine apnoeas characterised by a clear cessation of breathing. Whilst this leads to potential bias, with more false positives for shorter apnoeas, this approach enabled a balance between time (it was not feasible to review all apnoea episodes) and validating results against clinical experience (longer apnoeas likely to lead to instability).

Based on age, ventilation mode and medication, there were 277 recordings from 156 infants eligible. After including the apnoea criterion, the eligible sample size decreased to 257 recording sessions in 146 infants. 181,511 eligible apnoeas were detected from these included recordings (Figure 2.1).

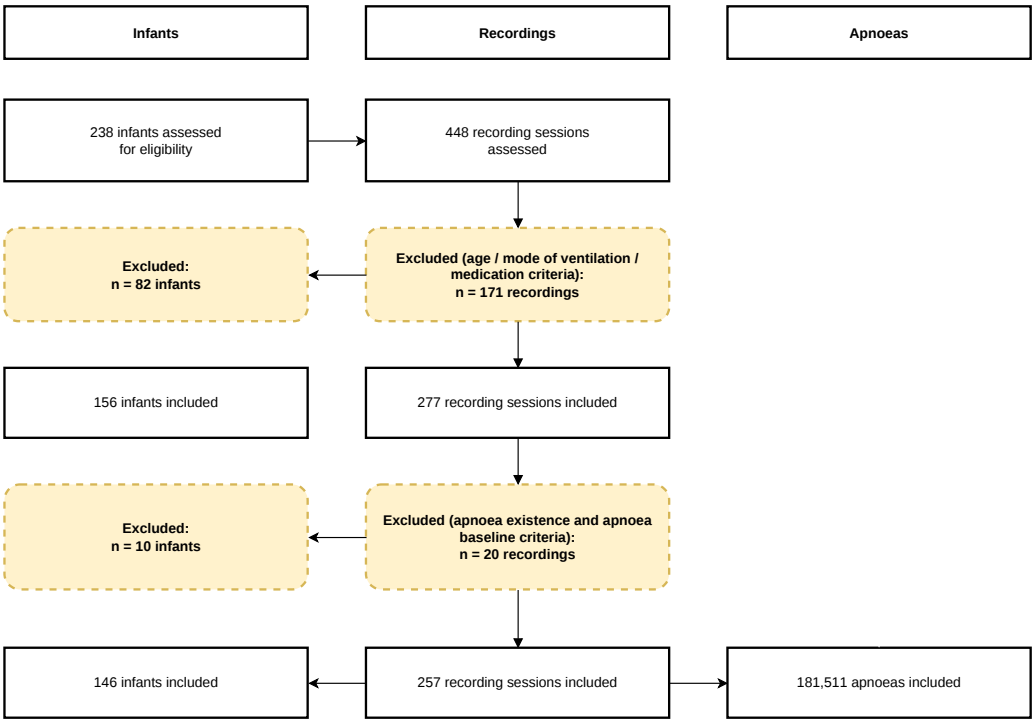

Figure 2.1: Data collection flow chart.

The 257 included recording sessions totalled 23,854 hours of vital signs recordings (median:

50.2, range: 0.7–634.3 hours per recording). Of the included 146 infants, 99 had one recording session only, 19 had two sessions, and 28 had three or more. The length of recording varied considerably due to differences in study design at the time of data collection. Some infants took part in a research study where multiple test occasions were conducted, with vital signs recorded continuously throughout the hospital stay. This subset included 39 infants, contributing 133 test occasions, with a median recording length of 155.9 hours per session (range: 0.9–634.3 hours). Other research studies recorded single epochs of vital signs, with a median duration of 19.6 hours (range: 0.7–242.0 hours). The demographic information of all included infants and recording sessions are presented in Table 2.1.

Variables	**Per infant (N = 146)**
Gestational age at birth (weeks)	28.9 (23.0 – 35.0)
Sex - Female	44.5%
Apgar score 5 min after birth	9 (3 – 10)

Variables	**Per recording session (N = 257)**
Recording length (hours)	50.2 (0.7 – 634.3)
Postmenstrual age at study (weeks)	33.2 (26.3 – 36.9)
Weight z-score at study	-0.8 (-3.7 – 2.3)
Number of apnoea per hour	6.8 (0.0 – 57.5)
Ventilation - SVIA	45.1%
Ventilation - Low Flow	14.4%
Ventilation - High Flow	40.5%

Table 2.1: Demographic and clinical variables of all infants and recording sessions with eligible apnoea. Value expressed as median (range) or percentage. Variables that do not change over time are reported per infant, whereas time-varying variables are reported per recording session.

2.2.2 Dataset 2 - NeoVital data

Dataset 2 was used in Chapter 5, which explores the functional respiratory age in preterm infants. To construct this dataset, available IP (impedance pneumogram) recordings were capped at 24 hours for each recording session. If the original recording was shorter than or equal to 24 hours, the entire recording was included. If the original recording exceeded 24 hours, an algorithm was developed to select an appropriate 24-hour epoch:

1. Set the starting time of the window as the beginning of the recording.
2. Select the 24-hour window from the starting time.
3. Calculate the signal coverage rate by dividing the number of available data points in the filtered IP signal by the expected number of data points (sampling frequency multiplied by the duration of 24 hours in seconds, which is 86,400 seconds).
4. If the selected window has a coverage rate greater than 80%, retain the window. If not, shift the starting time forward by 1 hour and repeat from Step 2.

Recording epochs were then selected according to the following inclusion criteria:

- Total recording duration $\geq$6 hours.
- Infants were $<$37 weeks' PMA at the time of the study.
- Infants were either SVIA or receiving low-flow or high-flow oxygen therapy at the time of the study.
- Infants were not receiving antibiotics (benzylpenicillin, amoxicillin, metronidazole, co-amoxiclav, chloramphenicol, gentamicin, cefotaxime, flucloxacillin, amikacin), corticos-

teroids (dexamethasone, budesonide, hydrocortisone), diuretics (furosemide, spironolactone, chlorothiazide), gastric medications (omeprazole, Carobel, Gaviscon), or opioids (morphine, fentanyl, clonidine) at the time of the study.

A total of 205 eligible recording epochs were selected from 105 infants, comprising 4738.4 hours of recordings in total (median = 24.0 hours per epoch, range = 6.0–23.9 hours). The basic demographic characteristics of all included infants and recording epochs are presented in Table 2.2.

Variables	**Per infant (N = 112)**
Gestational age at birth (weeks)	29.3 (23.7 – 34.0)
Sex - Female	41.9%
Apgar score 5 min after birth	9 (2 – 10)

Variables	**Per recording epoch (N = 237)**
Recording length (hours)	24.0 (6.2 – 24.0)
Postmenstrual age at study (weeks)	32.9 (24.9 – 36.9)

Table 2.2: Basic demographic information for all infants and recording sessions included in Dataset 2. Value expressed as median (range) or percentage.

2.2.3 Dataset 3 - POPPI data

Dataset 3 was used in Chapter 5. This dataset was previously collected for a study on the analgesic efficacy and safety of morphine [108]. The dataset consisted of 30 infants: 15 received morphine during the study, and 15 received placebo. Each infant had a single test occasion at approximately 48 hours. Table 2.3 presents the demographic details of this dataset.

	Morphine (n=15)	Placebo (n=15)
Gestational age at birth (weeks)	28.1 (26.3–30.1)	28.6 (27.9–29.7)
Birthweight (g)	1107 (329)	1173 (350)
Mode of delivery		
Spontaneous vaginal delivery	8 (53%)	5 (33%)
Caesarean section	7 (47%)	10 (67%)
Male sex	12 (80%)	8 (53%)
Female sex	3 (20%)	7 (47%)
Postmenstrual age at study (weeks)	35.0 (34.3–35.4)	34.9 (34.3–36.3)
Weight at study (g)	2049 (426)	2127 (331)
Level of care at study		
Intensive care unit	1 (7%)	1 (7%)
High-dependency unit	5 (33%)	9 (60%)
Low-dependency unit	9 (60%)	5 (33%)
Ventilation mode at study		
Self-ventilating	9 (60%)	8 (53%)
Low-flow oxygen therapy	2 (13%)	1 (7%)
High-flow oxygen therapy	4 (27%)	6 (40%)

Table 2.3: Demographic information for all infants. Value expressed as median (range) or percentage.

2.3 Data pre-processing

Pre-processing and analyses were performed in MATLAB (ver. 2024a; MathWorks Inc., Natick, USA) and Python (ver. 3.11.5; Python Software Foundation, Wilmington, USA). IBIs were identified from the IP signals, by applying a previous validated algorithm [109]. The algorithm first filters the signal to reduce the noise introduced by movements and cardiac activity. It then calculates a threshold equivalent to 0.4 times the signal's standard deviation over the preceding 15 breaths. Finally, the algorithm eliminates false positive apnoeas using a machine learning classifier.

Because the original algorithm was designed specifically to classify pauses longer than 15 seconds, an additional machine learning classifier was applied to distinguish true and false positive IBIs between 5 and 15 seconds. This classifier was developed independently by colleagues within our research group and is described in the Supplementary Methods of Chen et al. [2]. Briefly, candidate pauses were manually labelled by two independent researchers as either true pauses in breathing or artefacts (e.g. shallow breathing or poor electrode contact). A quadratic support vector machine (SVM) was then trained using features derived from the amplitude and variability of the IP signal immediately before, during, and after each candidate pause. The classifier achieved a balanced accuracy of 0.93 on the training set and 0.90 on an independent test set [2]. The output of this classifier was used to exclude false-positive short IBIs (5-15 seconds) before subsequent analyses.

2.4 Operational definition of bradycardia in preterm infants

The clinical definition of bradycardia events in preterm infants is usually a minimum heart rate below 100–110 beats per minute. However, older infants may sometimes have baseline heart rates close to this range without experiencing physiological problems. To assess the influence of developmental maturity on heart rate responses to apnoea, relationships between PMA and multiple heart rate metrics were examined using Dataset 1. For analyses and figures presented in this section, each data point represents the median value across all apnoeas within an individual recording, ensuring that recordings with a large number of events did not disproportionately influence the results.

Baseline heart rate demonstrated a significant negative correlation with PMA (Pearson correlation coefficient r = -0.27, $p < 0.0001$, Figure 2.2A), indicating that older infants exhibited lower resting heart rates. Similarly, the minimum heart rate observed from the onset of an apnoea until 10 s after its termination was also negatively correlated with PMA (r = -0.22, $p < 0.0001$, Figure 2.2B). In contrast, the percentage change in heart rate from baseline to the minimum value after the start of the apnoea was not associated with PMA (r = -0.01, p = 0.84, Figure 2.2C).

These findings highlight substantial maturity-related variation in absolute heart rate values. In particular, baseline heart rate values for late preterm infants were already close to commonly used absolute thresholds for bradycardia. As a consequence, defining bradycardia solely on the basis of a minimum heart rate would potentially classify older infants as experiencing bradycar-

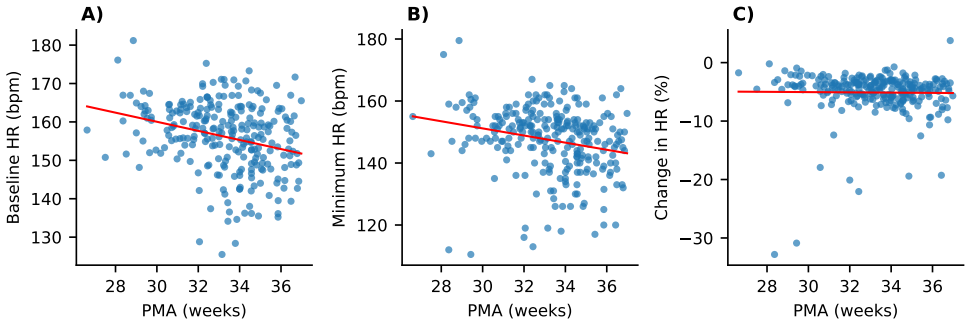

Figure 2.2: Relationship between PMA and heart rate characteristics during and after the start of the apnoea. Scatter plots show associations between PMA and (A) baseline heart rate, (B) minimum heart rate observed from the onset of apnoea until 10 s after apnoea termination, and (C) percentage change in heart rate relative to baseline. Each point represents the median value across all apnoeas within an individual recording. Solid red lines indicate the best-fit linear regression.

dia, even when the relative deviation from baseline was small and unlikely to reflect clinically meaningful instability. To mitigate this bias and to ensure comparability across developmental stages, bradycardia was therefore defined in this study as a relative decrease in heart rate exceeding 30% from baseline for at least 5 consecutive seconds.

Chapter 3

Apnoea duration and changes in cardiorespiratory and cerebrovascular responses

3.1 Introduction

Apnoea in preterm infants has been shown to induce physiological changes and is commonly associated with bradycardia and oxygen desaturations [110–112]. However, there is substantial variability both within and between neonates in their cardiorespiratory and cerebrovascular responses to an apnoea of a given duration [7]. Such heterogeneity suggests that apnoea duration alone may not adequately capture the physiological impact of respiratory pauses, and that individual demographic and development differences play an important role.

Understanding how different infants respond to apnoea of varying duration is essential for informing more individualised clinical management in the Neonatal Unit. In particular, improved characterisation of these responses may support the development of personalised alarm thresholds and intervention strategies, rather than relying on uniform duration-based criteria. Although studies have investigated the physiological consequences of apnoea in preterm neonates, the research results remain fragmented, and the methods vary greatly. The degree of the relationship between the duration of apnoea and changes in cardiorespiratory and cere-

brovascular responses has not been systematically synthesised.

Therefore, we conducted a systematic review and meta-analysis to comprehensively examine the existing evidence on the relationship between apnoea duration and associated physiological changes, including heart rate, oxygen saturation, cerebral oxygenation, and cerebral blood flow, in preterm neonates. In addition, this review aimed to identify and evaluate factors that may modulate these relationships, such as maturational stage, comorbid pathology, pharmacological treatment, and apnoea burden. Specifically, we sought to address the following questions:

1. What are the relationships between the duration of pauses in breathing/apnoea and changes in heart rate, blood oxygen saturation, cerebral oxygenation, and cerebral blood volume in hospitalised premature neonates?
2. How do these relationships vary across neonates with different PMA, pathology (specifically sepsis and necrotising enterocolitis [NEC]), medication (specifically methylxanthines and opioids), and any other potential modulating factors identified in the included papers?
3. How are these relationships modulated by the frequency/clustering of pauses in breathing?

The main results of this chapter have been published in Chen, Y., Zandvoort, C.S., Baxter, L. et al. Apnoea duration and changes in cardiorespiratory and cerebrovascular responses in preterm neonates: a systematic review and meta-analysis. Pediatr Res (2025) [1].

3.2 Methods

This systematic review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [113]. The review was started on November 2023, and the protocol for this systematic review was registered in PROSPERO (CRD42024534164) on April 2024. The registration was completed following initial searches and a pilot process of the study selection, but prior to the formal screening stage, in line with PROSPERO's guidelines for registering systematic reviews.

3.2.1 Eligibility criteria

All primary empirical and peer-reviewed studies reporting a relationship between apnoea or respiratory pause duration and at least one of the outcomes (change in heart rate, blood oxygen saturation, cerebral oxygenation and cerebral blood volume) in hospitalised human neonates with PMA $<$ 37 weeks were included. We excluded papers with the wrong population (non-human species; pre-natal humans; full-term neonates (PMA $\geq$ 37 weeks) and older (e.g., children, and adult subjects); neonates with neurological abnormalities (e.g. seizures) or congenital abnormalities; and non-hospitalised neonates. We also excluded papers with the incorrect study characteristics (secondary literature (e.g. reviews, book chapters); non-empirical literature (e.g. opinions, commentaries, perspectives); non-peer-reviewed grey literature (e.g. conference abstracts, meeting reports, theses); and study protocols). There were no restrictions based on publication language or date.

For studies that included neonates with both PMA < 37 weeks and ≥ 37 weeks, inclusion was permitted only if all neonates were born preterm. Where possible, data from neonates with PMA ≥ 37 weeks were excluded from the analysis.

3.2.2 Information sources and search strategy

The literature search for this review was conducted across multiple databases, including MEDLINE (Ovid), Embase (Ovid), PsycINFO (Ovid), and the Cochrane Central Register of Controlled Trials (Cochrane Library, Wiley). These databases were selected to ensure comprehensive coverage of relevant studies. The search strategy was developed in collaboration with a librarian from the Bodleian Health Care Libraries. A combination of subject headings and controlled keywords was used; Full search strategies for each database are provided in Appendix A.

The initial search was conducted on 23 November 2023. As the review spanned over a year, the search was updated on 10 December 2024 to ensure that the results were up-to-date. Additionally, backward citation tracking was performed on all included papers using the Citationchaser package [114] as a supplementary source of information.

3.2.3 Selection process

The Bodleian Health Care librarian recommended using Covidence (Melbourne, Australia) as the systematic review management tool. Covidence was used to manage records and data throughout the review process. References were imported into the platform, and automatic

deduplication was performed based on matching title, publication year, volume, and authors. However, variations in reference formatting - such as differences in author name formats or the presence of additional symbols (e.g. brackets) in titles - sometimes prevented the system from detecting duplicates. These undetected duplicates were manually identified and removed during the screening process.

In total, seven reviewers and one arbitrator participated in the selection process. The selection followed a two-stage process:

1. A title and abstract screening against the inclusion criteria to identify potentially relevant papers.
2. A full-text screening of all the papers identified as possibly relevant for inclusion from the initial screening above.

In the title and abstract screening stage, a pilot block of 20 randomly selected articles was first independently screened by all reviewers to ensure consistency. Each reviewer independently voted on each article and disagreements were discussed collectively until a consensus was reached on the inclusion and exclusion criteria.

Following the piloting, the remaining articles were randomly divided into five blocks, with two independent reviewers assigned to each block. I served as one of the reviewers for all blocks to maintain consistency throughout the screening process (i.e. I screened all articles). Articles were excluded at this stage if their abstracts did not describe any form of relationship between apnoea and changes in heart rate, oxygen saturation, cerebral oxygenation, and/or cerebral blood volume.

Articles that passed the title and abstract screening then underwent full-text review. Each full-text article was assessed independently by the same two reviewers. During both screening stages, non-English articles were reviewed by fluent speakers of that language, if possible, with an additional reviewer using Google Translate to validate the results. This procedure applied to all papers in Dutch, German, Italian, and French. One article in Bulgarian and one in Danish were screened using Google Translate only, due to the absence of fluent speakers in the review team; both were excluded at the title and abstract stage. In total, 11 non-English abstracts were identified during the screening process.

All screening was conducted in duplicate and independently. Any inconsistency between the votes were resolved through discussion between the two reviewers. If agreement could not be reached, an arbitrator was consulted. The same arbitrator provided oversight across all papers and at both stages of screening.

3.2.4 Data collection process and data items

I carried out all the data extraction process. The following variables were extracted: author's names; year of publication; country and region of publication (according to World Health Organisation categorisation [115]); hospital(s) where data was collected, study period; study design; sample size; neonate's age (gestational age and postmenstrual age), including median and range where available; type of respiratory support (the use of mechanical or non-mechanical devices to assist or maintain breathing, such as CPAP, NIPPV, mechanical ventilation, high-flow nasal cannula, and other forms of ventilator support) received by the neonates (if any); supplemental oxygen (the administration of additional oxygen, regardless of whether breathing

support is used) received by the neonates (if any); the pathological conditions of the neonates (specifically sepsis and NEC); type of medication used (if any); description of apnoea/pause in breathing (i.e., duration, method of measurement, any resuscitation provided, information on the frequency of pause/clustering of pauses); description of the changes in outcome variables (i.e., absolute change, baseline and minimum/maximum values if provided, method of measurement); duration of the recordings and time of day; signal quality measures (if given).

We also actively sought evidence for any other potential modulators of the relationship between apnoea duration and physiological instability investigated within the studies included in the review. When relevant data were only available in figure format, numeric values were extracted using PlotDigitizer software. All figures processed by this software in this review were either 2D bar charts or line charts plotted on linear scales. For each figure, the image was uploaded to the software, two reference points on the y-axis were marked, and their corresponding values were entered. The points of interest were then selected, and the software calculated their values using linear interpolation. This process was performed twice for each figure to ensure consistency and accuracy of the extracted data.

3.2.5 Study risk of bias assessment (quality assessment)

Assessment of the risk of bias in the included studies was carried out by using the checklist for systematic reviews and research syntheses provided by the Joanna Briggs Institute (JBI) Critical Appraisal tools. Two reviewers evaluated the risk of bias independently. Any disagreement between the two reviewers was settled through discussion.

To derive an overall risk of bias rating [116, 117], we focused exclusively on items within the JBI checklists related to internal validity. Each item in the relevant JBI checklists was classified as either 'critical' (pertaining to internal validity) or 'non-critical' (pertaining to other constructs). Critical items were mapped to established domains of bias, including selection bias, confounding, measurement bias, temporal bias, and attrition bias. Studies were then rated based on their responses to critical items, following a predefined rule set: low risk of bias was defined as no 'no' responses on critical items; moderate risk of bias was defined as one 'no' response on a critical item; high risk of bias was defined as two or more 'no' responses on critical items. Responses of 'unclear' or 'not applicable' were not treated as indicators of bias, ensuring a conservative and transparent assessment approach.

3.2.6 Data synthesis

The outcome of the database searches and study selection process was presented in a PRISMA flowchart. The data synthesis was performed using a comprehensive narrative approach to ensure a rigorous and informative analysis. Narrative and qualitative synthesis involved the tabulation of results where feasible.

Quantitative data from the selected studies regarding the relationship between apnoea duration and physiological changes in preterm neonates were extracted and pooled for meta-analysis where available. The dependent variable were the mean percentage changes in vital signs. If a study provided the values for percentage change in vital signs directly, these were used for the meta-analysis. When values at the beginning and end of apnoea were provided, the difference was calculated and divided by the starting value to determine the percentage change. The

independent variable was the apnoea durations. If a study provided the mean apnoea durations for each pooled point, these were used directly. When only a range of apnoea durations was provided, the midpoint of the range was used as the X-value (e.g., if the range of apnoea duration was 10 – 20 seconds, the X-value would be 15 seconds for that point). If multiple studies presented the same results using the same dataset, then we only included the results once in the meta-analysis.

Pooled data were analysed using Pearson correlation coefficients, linear regression models, and linear mixed-effects models. For the linear regression models, apnoea duration was the predictor, and changes in vital signs were the response. The square root of the number of apnoea events associated with each data point was used as the weight in the model. The coefficient of determination (R^2) was used to assess how much of the variation in vital signs could be explained by apnoea duration. The formula to calculate R^2 was:

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SS}_{\mathrm{RES}}}{\mathrm{SS}_{\mathrm{tot}}},$$

where $\mathrm{SS}_{\mathrm{RES}}$ is the sum of squares of residuals, and $\mathrm{SS}_{\mathrm{tot}}$ is the total sum of squares.

For linear mixed-effects models, the study ID number was used as a random effect, and the duration of apnoea as a fixed effect. The impact of apnoea type (whether the reported apnoea is central apnoea, mixed apnoea, obstructive apnoea, or a mixture of two or more types), PMA, and the use of theophylline (a medication that is commonly used for apnoea treatment) on the relationship between apnoea duration and physiological changes was analysed, where applicable, by adding interaction terms between these factors and apnoea duration as fixed effects in

the model.

For a post hoc analysis of how PMA affects the relationship between changes in oxygen saturation (SpO_2) and apnoea duration, we used a linear regression model with the formula:

$$\text{change in } SpO_2 = \beta_0 + \beta_1 \times \text{Duration} + \beta_2 \times \text{PMA} + \beta_3 \times (\text{Duration} \times \text{PMA}) + \sigma$$

Using this model, we predicted the apnoea duration at which oxygen saturation is likely to drop by more than 10% for neonates of different PMAs. The 10% threshold is commonly used to define neonatal desaturation [118].

The correlation coefficients from different papers reporting the same outcome were standardised and converted into a common effect size metric, Fisher's Z score, which was combined to provide overall correlation factors. The regression lines from the individual studies were plotted alongside the line obtained from the meta-analysis for visual comparison of the variation across studies.

3.3 Results

A total of 4,362 studies were identified, after the removal of 8,567 duplicates (Figure 3.1). Following the title and abstract screening, 4,190 studies were excluded, and a further 129 studies were excluded after the full-text review. One study [119] was excluded due to unavailability of the full text after database searches and inter-library requests. Overall, 42 studies were included in this review, involving a total of 1,483 neonates with 2,399 recording sessions (i.e. a

period during which a patient's vital signs are recorded).

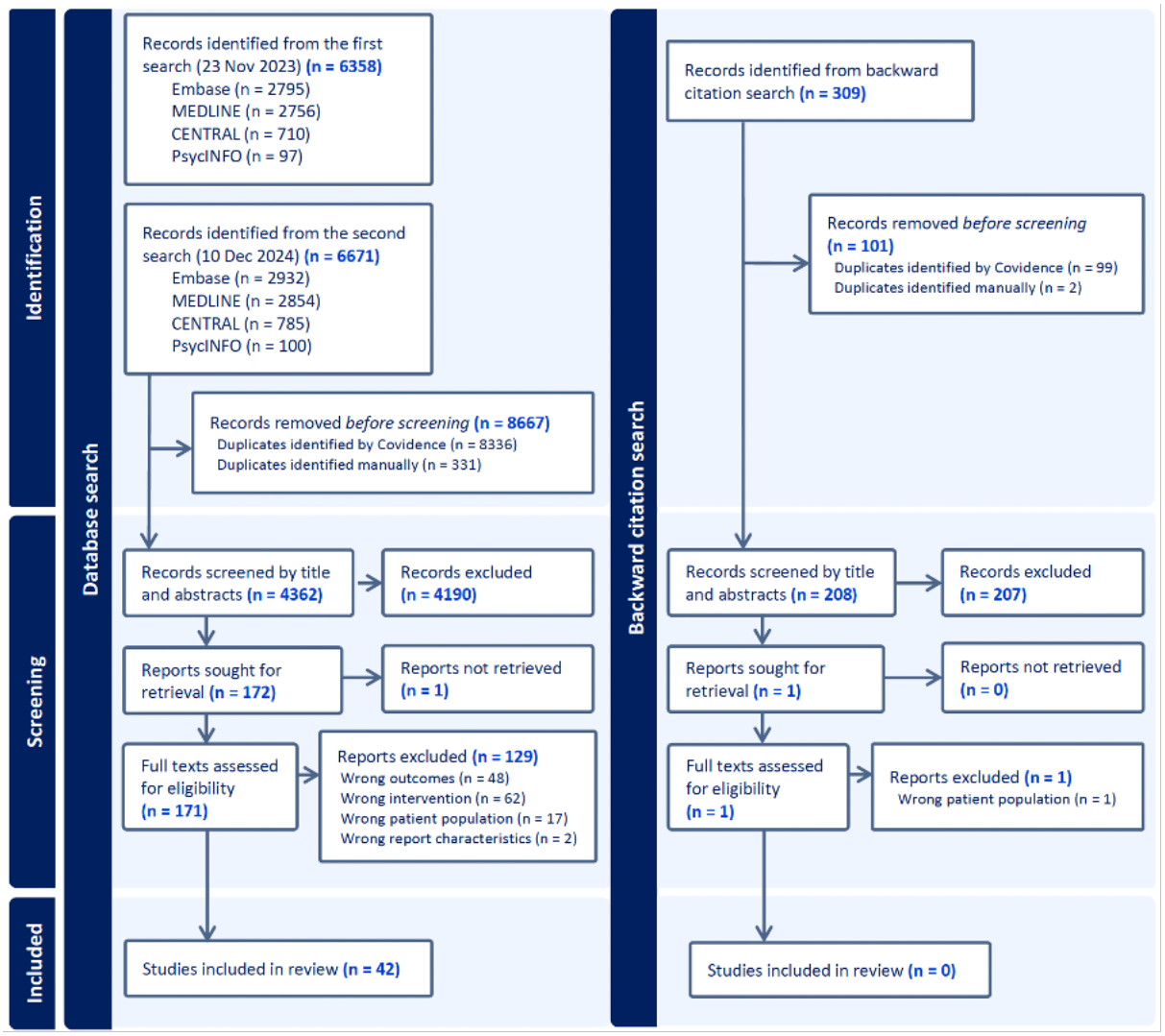

Figure 3.1: Prisma flow chart showing the study selection process

The median number of neonates per study was 24 (range: 1–335; IQR: 14–31.75), and the median number of recording sessions per study was 31 (range: 1–386; IQR: 18–78). Histograms showing the distribution of the number of neonates and recording sessions per study are presented in Figure 3.2. Mean PMA was reported or can be inferred in 12 studies, with an average of 35.3 weeks. In 17 studies, PMA was reported using alternative summary metrics, such as

the median or range. In the remaining 13 studies, PMA was not reported numerically.

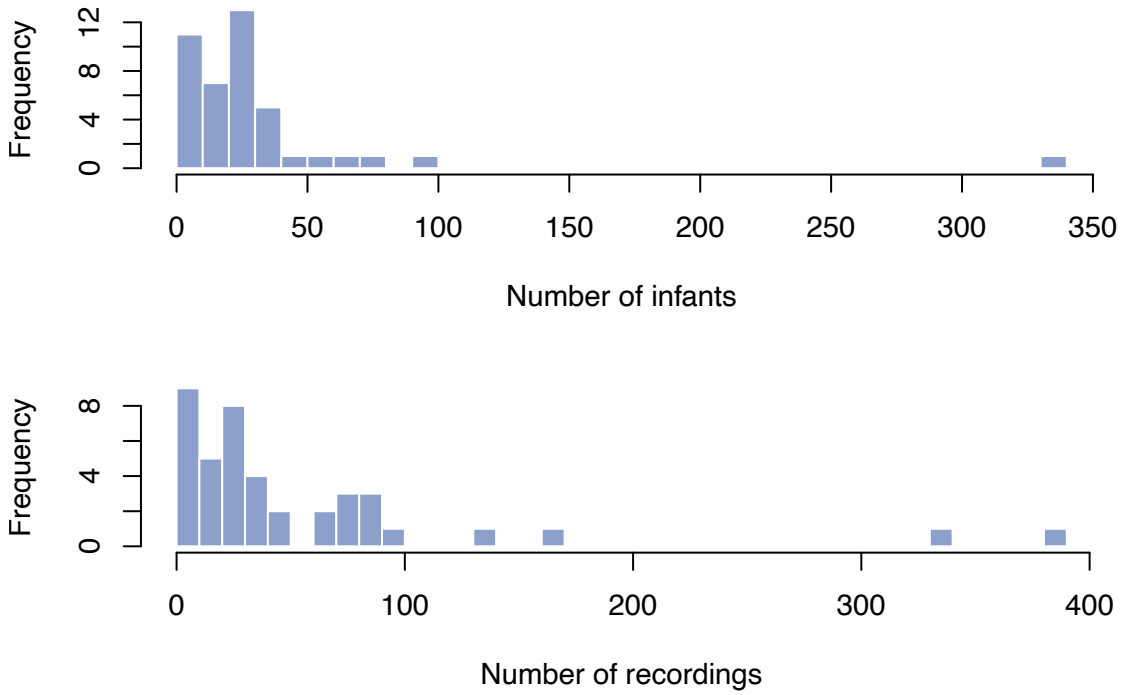

Figure 3.2: Histograms of the number of neonates and recordings in the included studies.

Twenty-three of the included studies (54.8%) were cross-sectional, 16 (38.1%) were cohort studies, 2 (4.8%) were case-control studies, and 1 (2.4%) was a case report. The year of publication ranged from 1969 to 2022, with a median of 1992. Most of the studies were from Europe (n = 21, 50%) and the Americas (n = 17, 40.5%). The remainder of studies were from the Western Pacific (n = 4, 9.5%). No studies were published from the South-East Asian Region, Eastern Mediterranean or Africa (regions classified using the World Health Organisation categorisation [115]). The demographic summary of the included studies can be found in Figure 3.3.

The included studies reported the relationship between apnoea duration and one or more phys-

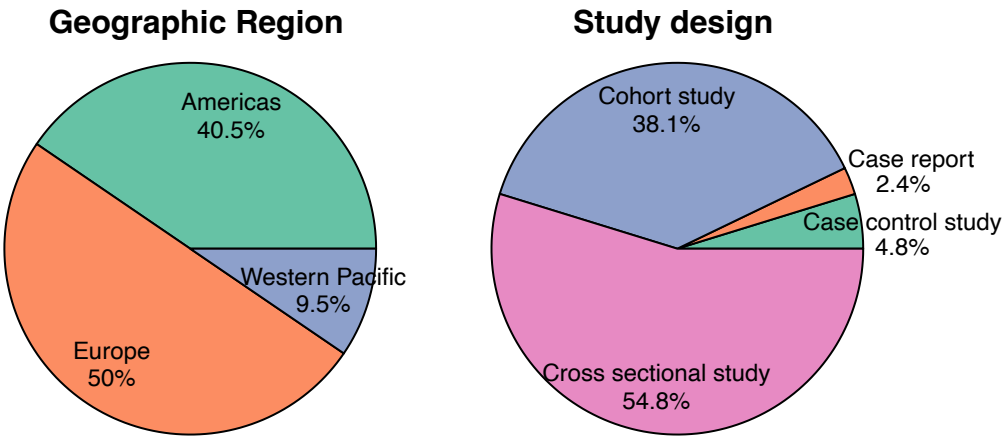

Figure 3.3: Geographic regions and study designs of the included studies.

iological outcome measures: change in heart rate (n = 32, 76.2%), change in oxygen saturation (n = 25, 60.0%), change in cerebral oxygenation (n = 3, 7.1%), and change in cerebral blood volume (n = 1, 2.4%). Details of the studies reporting each outcome are presented in Table 3.1.

Table 3.1: Demographic details for the studies associated with each outcome measures. *: Only the metrics that has corresponding apnoea duration reported were counted, some studies reported multiple metrics. #: Some of these studies did not specify the exact number of infants receiving the relevant treatment

Outcome Measure	**Heart rate**	**Oxygen Saturation**	**Cerebral Oxygenation**	**Cerebral Blood Volume**
No. of studies	32	25	2	3
No. of patients	1,044	1,122	97	114
No. of recording sessions	>1,792 (one study only stated that multiple recording was available for each infant, without stating the exact number)	1,875	475	386
Postmenstrual age (PMA) at study [weeks]	10 studies: averaged mean PMA = 35.0 weeks (237 neonates). 10 studies: PMA reported in alternative metrics such as median or range. 12 studies: PMA not specified numerically.	13 studies: averaged mean PMA = 35.6 weeks (378 neonates). 7 studies: PMA reported in alternative metrics such as median or range. 5 studies: PMA not specified numerically.	1 study: averaged mean PMA = 34.2 weeks (39 neonates). 1 study: PMA not specified numerically.	2 studies: averaged mean PMA = 34.9 weeks (56 neonates). 1 study: PMA not specified numerically.
Respiratory stimulants at the time of study	8 studies: no respiratory stimulant. 15 studies: at least 102 neonates# received respiratory stimulants. 9 studies: not specified.	7 studies: no respiratory stimulant. 14 studies: at least 130 neonates# received respiratory stimulants. 4 studies: not specified.	2 studies: 37 infants received respiratory stimulants.	3 studies: 47 infants received respiratory stimulants.

Respiratory support at the time of study	10 studies: no respiratory support. 7 studies: 287 neonates on respiratory support. 15 studies: not specified.	7 studies: no respiratory support. 7 studies: at least 320 neonates# on respiratory support at the time of study. 11 studies: not specified.	1 study: no respiratory support. 1 study: not specified.	1 study: no respiratory support. 2 studies: not specified.
Supplemental oxygen at the time of study	12 studies: no supplemental oxygen. 6 studies: at least 75 neonates# received supplemental oxygen. 14 studies: not specified	11 studies: no supplemental oxygen. 5 studies: at least 42 neonates# received supplemental oxygen. 9 studies: not specified.	1 study: no supplemental oxygen (58 infants). 1 study: not specified.	1 study: no supplemental oxygen (58 infants). 1 study: 9 infants received supplemental oxygen. 1 study: not specified.
Measuring techniques	ECG (n = 26). Plethysmography (n = 4). Impedance (n = 6). Thermistor (n = 3).	Pulse oximetry (n = 24)	NIRS (n = 2)	NIRS (n = 3)
Metrics used	Relative change in HR (n = 6). Absolute change in HR (n = 3). Minimum HR value (n = 1). Bradycardia incidences (n = 6). Correlation coefficients (n = 6). Linear regression results (n = 2). Other metrics only (n = 8).	Relative change in SpO_2 (n = 5). Minimum SpO_2 (n = 1). Desaturation instances (n = 2). Correlation coefficients (n = 9). Linear regression (n = 3). Duration of apnoea with and without desaturation (n = 2). Other metrics only (n = 7).	Change in cerebral Hb difference (cHbD) (n = 2).	Change in total cerebral blood volume (CBV) (n = 1). Amplitude of total Hb (tHb) (n = 1). Change in concentration changes of total cerebral haemoglobin (cHbtot) (n = 1).

The definition of apnoea, desaturation, and bradycardia events investigated in the different studies varied as summarised in Table 3.2. The techniques used to measure apnoea also varied, including polysomnography, acoustic monitoring, inductive plethysmography, thermistor, and pneumotachograph (Table 3.1, Tables 1-4).

Parameter	Variations
Definition of apnoea	Cessation of breathing longer than: $\geq$3–5 seconds (n = 6 studies); $\geq$10 seconds (n = 28 studies); $\geq$15–20 seconds (n = 7 studies). 9 studies required associated desaturation and/or bradycardia to define apnoea.
Definition of desaturation	Oxygen saturation $<$90% in 7 studies, $<$85% in 4 studies, and $<$80% in 6 studies. 5 studies defined desaturation as a $\geq$10% drop from baseline.
Definition of bradycardia	Heart rate $<$90 bpm (n = 5 studies), or $<$100 bpm (n = 15 studies). 3 studies defined bradycardia as a $\geq$20% drop from baseline. 3 studies graded severity (mild vs. severe).

Table 3.2: Definition of apnoea, desaturation and bradycardia used across the studies.

3.3.1 Relationship between apnoea duration and change in heart rate

A total of 31 articles investigated the change in heart rate compared with apnoea duration (Table 3.8), involving 1008 neonates and 1734 recording sessions in total. Most studies (n = 26, 83.9%) measured heart rate using the electrocardiogram (ECG), while some used different monitors (Table 3.1). The relationship between heart rate changes and apnoea duration was examined using various methods, including correlation coefficients, linear regression models, and scatter plot diagrams. Most studies demonstrated that a longer apnoea duration is correlated with a greater reduction in heart rate.

Six studies [7, 110, 111, 120, 121] reported correlation coefficients, with all identifying a

significant correlation between apnoea duration and metrics relating to decreased heart rate. The correlation coefficients were calculated based on different heart rate metrics: Minimum heart rate: n = 2 studies (magnitude of r ranged between 0.10 and 0.77), Relative change in heart rate: n = 1 studies (r = 0.09, p = 0.001), Bradycardia incidence: n = 1 study (magnitude of r for different apnoea types ranged between 0.51 and 0.65), Area under the heart rate curve: n = 1 study (r = 0.81, $p < 0.001$), Time of onset of bradycardia: n = 1 study (r = 0.73, $p < 0.0001$).

Six studies [110, 120, 122–125] presented the mean change in heart rate during apnoea for different apnoea duration subgroups or reported the information where the mean change can be inferred. These studies were included in the meta-analysis. The average of the mean PMA of the neonates across the six studies was 32.84 $\pm$ 1.56 (mean $\pm$ SD). The pooled data across all six studies did not show a significant relationship between apnoea duration and heart rate change (Figure 3.4, n = 6 studies; slope = -0.27, R^2 = -0.05; r = -0.31, p = 0.062). However, data from Finer et al. (1992) [123] showed an opposite trend to the other four studies, demonstrating a smaller change in heart rate for longer apnoeas. This might be due to the study's unique definition of apnoea – the authors only recorded apnoeas associated with bradycardia or hypoxemia, excluding shorter pauses with less significant decrease in heart rate. Removing this study from further meta-analysis, the pooled data from the other five studies showed a negative relationship between apnoea duration and percentage change in heart rate during apnoea (Figure 3.5, n = 5 studies; slope = -1.55, R^2 = 0.36; r = -0.61, p = 0.0019).

Six studies [111, 121, 123, 126–128] reported the percentage of apnoea episodes associated with a bradycardia event. Gabriel et al. (1976) (b) [126] was not included in the meta-analysis

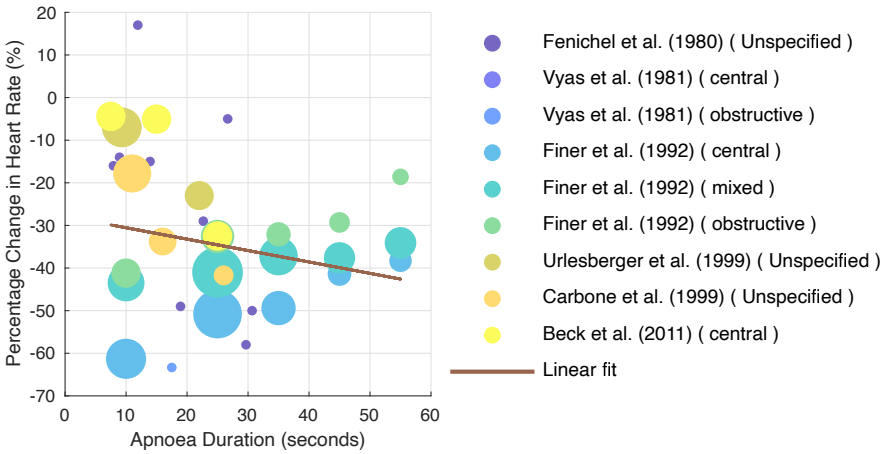

Figure 3.4: The percentage change in heart rate during apnoea compared with apnoea duration, before removing the data from Finer et al. (1992) [123]. The size of the points indicates their weight in producing the linear best fit, calculated as the square root of the number of apnoea episodes associated with each point.

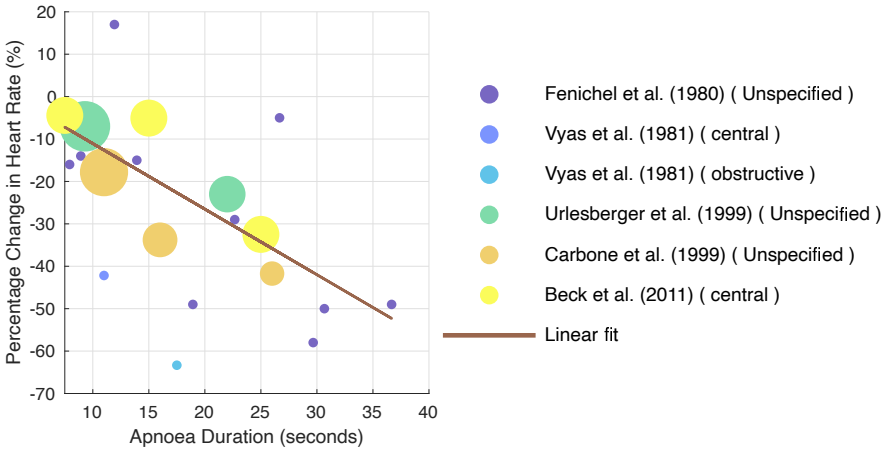

Figure 3.5: The percentage change in heart rate during apnoea compared with apnoea duration, after removing the data from Finer et al. (1992) [123]. The size of the points indicates their weight in producing the linear best fit, calculated as the square root of the number of apnoea episodes associated with each point.

as the data and results were the same as Gabriel et al. (1976) (a) [121]. The remaining 5 studies [111, 121, 123, 127, 128] were pooled in a separate meta-analysis; four of them [121, 123, 127, 128] used heart rate below 100 bpm as the definition of bradycardia whilst Henderson-Smart et al. (1986) [111] defined bradycardia as a fall in heart rate of more than 30% below the baseline before the apnoea event. Generally, longer apnoeas were associated with higher rates of bradycardia (Figure 3.6, n = 5 studies; slope = 1.71, R^2 = -0.05; r = 0.44, p = 0.0052). The study from Finer et al (1992) [123] was again removed from the meta-analysis due to its unique definition of apnoea. The pooled data from the other four studies showed a positive correlation between apnoea duration and the percentage of apnoeas associated with bradycardia (Figure 3.7, n = 4 studies, slope = 2.31, R^2 = 0.57; r = 0.81, $p < 0.0001$).

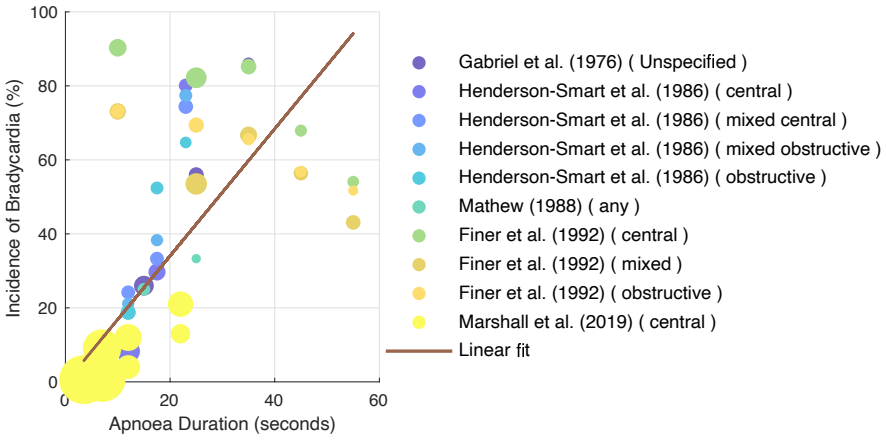

Figure 3.6: The percentage of apnoea events accompanied by bradycardia against apnoea duration, before removing the data from Finer et al. (1992) [123]. The size of the points indicates their weight in producing the linear best fit, calculated as the square root of the number of apnoea episodes associated with each point.

Three studies [8, 124, 125] reported the relationship between apnoea duration and absolute change in heart rate and all three studies found that longer apnoea durations were associated with greater reductions in heart rate during the events. However, as the original data points and correlation coefficients were not available, the results could not be pooled for meta-analysis.

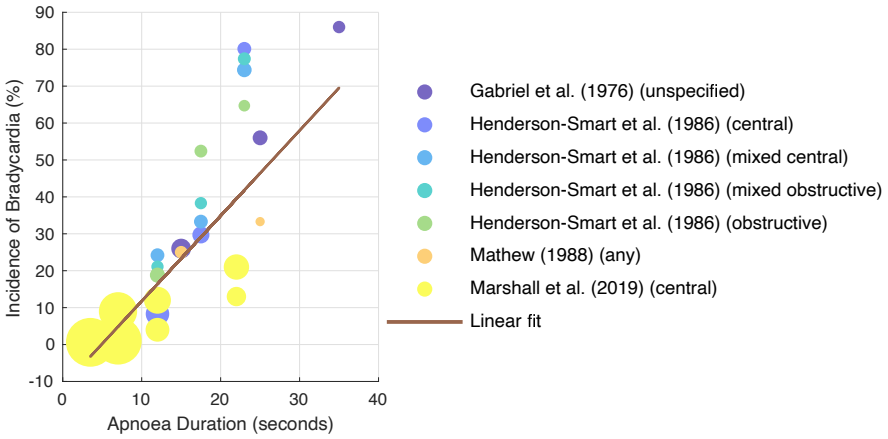

Figure 3.7: The percentage of apnoea events accompanied by bradycardia against apnoea duration, after removing the data from Finer et al. (1992) [123]. The size of the points indicates their weight in producing the linear best fit, calculated as the square root of the number of apnoea episodes associated with each point.

Upton et al. (1991) [7] showed that minimal heart rate = 74.22 - 0.14 $\times$ duration of apnoea ($r = 0.11$, $p < 0.0001$). Waggner et al. (1989) [129] modelled the decrease in heart rate as a function of apnoea duration and showed an average accuracy of 73%.

Another 11 studies [44, 118, 130–138] used a range of heart rate metrics (e.g., outcomes from multivariate linear analyses assessing the dependence of normalised percentage changes in heart rate on apnoea duration). These studies could not be pooled for meta-analysis due to the limited availability of data for each metric or because some only described the relationship without providing numeric values (Table 3.8). The majority (10 out of 11) reported a statistically significant association between apnoea duration and change in heart rate, generally indicating that longer apnoeas were associated with greater reductions in heart rate. The direction of the reported statistical associations varied depending on how heart rate changes were quantified. In contrast, Curzi-Dascalova et al. (1989) [135], who analysed 602 pauses all under 12 seconds, found that pauses lasting 10 to 12 seconds (N = 11 pauses) did not result in greater

cardiac deceleration compared to shorter pauses under 10 seconds.

3.3.2 Relationship between apnoea duration and change in oxygen saturation

A total of 25 articles investigated the relationship between change in oxygen saturation with apnoea duration (Table 3.9), involving 1,199 neonates and 1,970 recording sessions in total. All the studies used a pulse oximeter to measure oxygen saturation. As with the studies investigating changes in heart rate, a wide variety of oxygen saturation metrics and statistical analysis techniques were used (Table 5).

Nine studies [7, 123, 139–145] reported correlations coefficients between apnoea duration and different oxygen saturation metrics (change or drop in oxygen saturation: n = 4 studies, min oxygen saturation: n = 2 studies, duration of desaturations: n = 1 study, rate of desaturations: n = 1 study, area under the oxygen saturation curve: n = 1 study). Seven of these studies [7, 123, 139–142, 145] reported significant correlation coefficients, all indicating that longer apnoea durations were associated with greater decreases in oxygen saturation or longer desaturations. In contrast, Poets et al. (1995) [143] reported no significant correlation between the duration of the apnoeic pause and the duration of desaturation (r = 0.3; $P > 0.05$, Spearman's rank correlation test). Similarly, Adams et al. (1997) [144] found no significant correlation between the length of respiratory pauses and the rate of desaturations (r = -0.20). Both of these studies recorded all desaturation or hypoxaemic events and then analysed the apnoea episodes associated with them, whereas most other studies conducted the analysis in the opposite direction – starting with apnoea episodes and examining the associated desaturation events. The 4 studies [7, 123, 141, 142] that reported correlation coefficients based on the percentage decrease in

oxygen saturation were pooled together. The pooled correlation coefficient was 0.37 (n = 4 studies, $p < 0.0001$, 95% confidence interval: [0.34, 0.39]).

Five studies [8, 123, 124, 140, 146] presented the mean change or drop in oxygen saturation during apnoea for different apnoea duration subgroups. The average of the mean PMA of participants in the five studies is 32.68 ± 2.48 weeks. In the study by Urlesberger et al. (1999) [120], none of the participants received respiratory support or supplemental oxygen during the study period. In the study by Finer et al. (1992) [123], supplemental oxygen was administered only to infants whose baseline oxygen saturation fell below 90%-92%. For the remaining three studies, information regarding the use of respiratory support or supplemental oxygen was not reported. Meta-analysis pooling the data from these studies showed a negative relationship between apnoea duration and change in oxygen saturation (Figure 3.8, n = 5 studies, slope = -0.52, $R^2 = 0.95$; r = - 0.98, $p < 0.0001$).

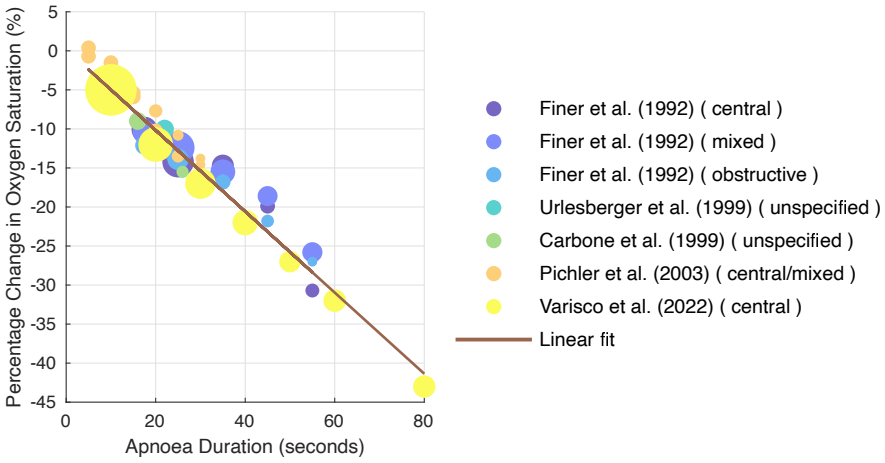

Figure 3.8: The percentage change in oxygen saturation during apnoea compared with apnoea duration. The size of the points indicates their weight in producing the linear best fit, calculated as the square root of the number of apnoea episodes associated with each point.

Three studies [7, 141, 142] used linear regression models to relate apnoea duration and change

in oxygen saturation. The regression line slope from the pooled data (-0.51) were within the range of these models (Figure 3.9, -0.34 for Muttitt et al. (1988) [141], -0.43 for Upton et al. (1991) [7], -0.67 for Upton et al. (1992) (B) [142]).

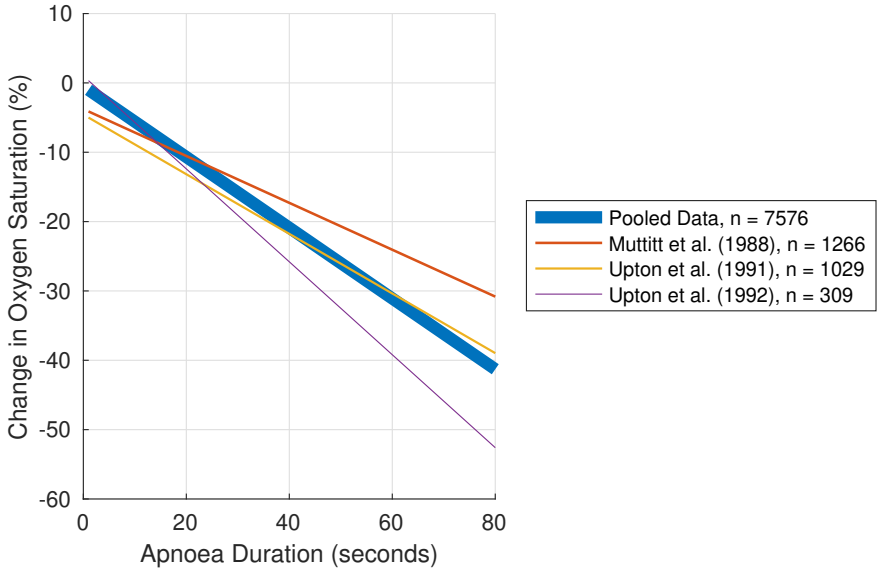

Figure 3.9: Linear regression lines between change in oxygen saturation and apnoea duration. The three thin lines (red, yellow, purple) represent the regression lines reported by individual studies. The thick blue line represents the regression lines of the data pooled from Gabriel et al. (1976) (A) [121], Henderson-Smart et al. (1986) [111], Mathew (1988) [127], and Marshall et al. (2019) [128]. The number n and the thickness of the lines represents the number of apnoea events associated with each line.

Two studies [118, 147] reported the relationship between apnoea duration and desaturation incidence. Spear et al. (1992) [118] found that both the longest duration of desaturation episodes with oxygen saturation below 80% and the total duration of desaturation episodes with oxygen saturation below 90% were greater in recordings containing prolonged apnoeas compared with those without prolonged apnoeas. Abdulhamid et al. (1992) [147] reported the percentage of apnoea episodes associated with desaturations according to apnoea type and corresponding duration ranges. However, the use of duration ranges prevented the data from being pooled.

There were eight other studies [44, 45, 128, 132–134, 148, 149] (Table 3.9) which used a variety of oxygen saturation metrics that were not pooled for meta-analyses, such as minimum oxygen saturation during apnoea episodes (n = 1 study), area under the curve (n = 1 study), lowest recorded oxygen saturation (n = 1 studies), desaturation incidences (n = 2 studies) and duration of apnoeas with and without desaturation (n = 2 studies). The data could not be pooled for additional meta-analyses due to limited data availability for each metric. In addition, some studies reported the relationships without numeric values that can be pooled. Most of these studies demonstrated significant correlations between apnoea duration and oxygen saturation, except Curzi-Dascalova et al. (2000) [148] found there were no significant differences between the amplitude of oxygen saturation changes observed following apnoea versus control periods.

3.3.3 Relationship between apnoea duration and change in cerebral oxygenation

Two studies [8, 120] examined the relationship between apnoea duration and cerebral oxygenation (Table 3.10). Changes in cerebral oxygenation were measured using Near Infra-Red Spectroscopy (NIRS). Urlesberger et al. (1999) [120] and Pichler et al. (2003) [8] both presented the mean change in cerebral haemoglobin difference (cHbD) during apnoea for different apnoea duration groups. Meta-analysis of these two studies demonstrated that cHbD decreased more during apnoeas with longer durations (Figure 2D, n = 2 studies, slope = -0.35, R^2 = 0.83; r = -0.92, $p < 0.0001$).

3.3.4 Relationship between apnoea duration and change in cerebral blood volume

Three studies [8, 120, 139] reported the relationship between apnoea duration and cerebral blood volume (Table 3.11). The changes in cerebral blood volume were measured using NIRS. Jenni et al. (1996) [139] reported the correlation coefficients between the amplitude of total haemoglobin concentration (tHb) and duration of apnoea as 0.55 for central apnoea, -0.005 for obstructive apnoea, and 0.7 for mixed apnoea. Urlesberger et al. (1999) [120] found there was no correlation between apnoea and change in concentration changes of total cerebral haemoglobin (cHbtot) for both apnoea duration $\geq$ 15 seconds (R^2 = 0.07 and apnoea duration of 5-14 seconds (R^2 = 0.01). The tHb and cHbtot are used as surrogate markers of cerebral blood volume. The study from Pichler et al. (2003) [8] had a population comprised of two subgroups: a bradycardia group (bradycardia was defined as a heart rate decrease to below 80 beats per minute, n = 20) and a non-bradycardia group (n = 19). Episodes of apnoea with associated bradycardia were matched with episodes of apnoea without bradycardia based on the duration and PMA of the neonates. The cerebral blood volume decreased significantly with increased apnoea duration in the bradycardia group, but the relationship was not significant in the non-bradycardia group.

3.3.5 Factors which modulate the physiological response to apnoea

While most studies showed that longer apnoea durations correspond to greater physiological changes, there is considerable variation, particularly in heart rate changes (Figure 3.5 - 3.10).

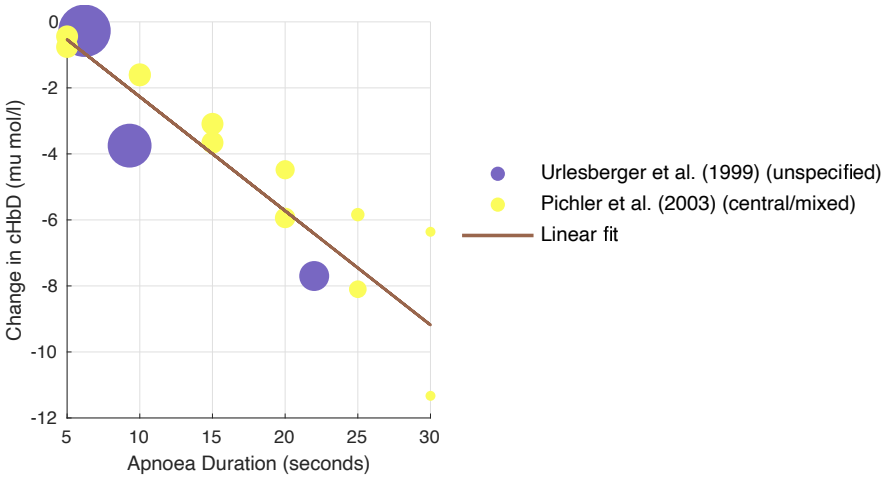

Figure 3.10: The change in cerebral haemoglobin difference (cHbD) during apnoea against apnoea duration. The size of the points indicates their weight in producing the linear best fit, calculated as the square root of the number of apnoea episodes associated with each point.

Studies included in this review have investigated whether apnoea type and the use of theophylline can affect the relationship between apnoea duration and changes in heart rate, oxygen saturation, cerebral oxygenation and/or cerebral blood volume. Using meta-analysis, we also explored whether PMA modulates the relationship.

Apnoea type

Several studies [139, 141, 150] showed that central and obstructive apnoeas could lead to different physiological responses. Muttitt et al. (1988) [141] identified a significant correlation between apnoea duration and heart rate decrease in central apnoeas (computer-diagnosed: $r = 0.19$, $p < 0.0001$; nurse-diagnosed: $r = 0.20$, $p < 0.0001$), but not in obstructive or mixed apnoea groups. Suichies et al. (1989) [150] reported significant positive correlations between the percentage of apnoeas associated with bradycardias and apnoea duration in central apnoeas ($r = 0.65$, $p < 0.01$) and mixed apnoeas ($r = 56$, $p < 0.05$), but the correlation was not signifi-

cant for obstructive apnoeas. Similarly, Jenni et al. (1996) found that the minimum heart rate during apnoea was negatively correlated with apnoea duration in central apnoeas (r = -0.77, $p < 0.01$) and mixed apnoeas (r = -0.69, $p < 0.01$) only, but again, no significant correlation was observed in obstructive apnoeas. However, with regard to oxygen saturation, Jenni et al. (1996) [139] observed a significant correlation between apnoea duration and oxygen saturation for all apnoea types (central: r = -0.49, $p < 0.01$; obstructive: r = -0.49, $p < 0.05$; mixed: r = -0.64, $p < 0.01$).

Studies shown in Figure 3.5 included four data points associated with central apnoea, one data point associated with obstructive apnoea, and 18 data points involving unspecified apnoea type. Meta-analysis on apnoea type was not performed for these studies due to limited data availability. The meta-analyses of the studies presented in Figures 3.7 and 3.8 indicated that apnoea type was not a significant modulating factor in the relationship between bradycardia incidence rate and apnoea duration (Table 3.3), but played a significant role in modulating the relationship between changes in oxygen saturation and apnoea duration (Table 3.4).

Theophylline

There is limited evidence for theophylline as a modulating factor. Muttitt et al. (1988) [141] found that the correlation between apnoea duration and decrease in heart rate was significant during theophylline treatment (r = 0.16, $p < 0.0001$) but not in observations without treatment. Theophylline did not significantly affect the decrease in oxygen saturation in this study. However, the correlation between apnoea duration and decrease in oxygen saturation was strongest during theophylline treatment (r = 0.43, $p < 0.0001$). Similarly, Upton et al. (1991) [7] found

Name	Estimate	SE	t-stats	DF	p Value
(Intercept)	-19.00	19.23	-0.99	12	0.34
Apnoea type (central)	13.26	19.48	0.68	12	0.51
Apnoea type (obstructive)	-12.92	37.08	-0.35	12	0.73
Apnoea type (mixed central)	-17.93	36.48	-0.49	12	0.63
Apnoea type (mixed obstructive)	-25.70	41.55	-0.62	12	0.55
Apnoea type (any)	31.55	61.24	0.52	12	0.62
Apnoea duration	3.00	0.86	3.49	12	0.0045**
Apnoea type (central):Apnoea duration	-1.56	0.91	-1.72	12	0.11
Apnoea type (obstructive):Apnoea duration	1.41	2.09	0.68	12	0.51
Apnoea type (mixed central):Apnoea duration	1.64	1.89	0.87	12	0.40
Apnoea type (mixed obstructive):Apnoea duration	2.17	2.18	1	12	0.34
Apnoea type (any):Apnoea duration	-2.17	3.39	-0.64	12	0.53

Table 3.3: Results of the statistical analysis investigating if apnoea type modulates the relationship between bradycardia incidence and apnoea duration. Fixed effects coefficients from the linear mixed-effects model with the formula: "bradycardia incidence 1 + ID + apnoea type * apnoea duration". The dataset used in this model included 24 data points from 4 studies. The apnoea types were classified into 6 categories: central apnoea (n = 10), obstructive apnoea (n = 3), mixed central apnoea (n = 3), mixed obstructive apnoea (n = 3), unspecified apnoea (n = 3), and any apnoea (n = 2). * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$. SE = standard error. DF = degree of freedom.

Name	Estimate	SE	t-stats	DF	p Value
(Intercept)	-0.20	0.47	-0.42	29	0.68
Apnoea type (mixed)	-2.57	1.36	-1.90	29	0.068
Apnoea type (obstructive)	-4.49	1.91	-2.35	29	0.026*
Apnoea type (unspecified)	0.18	1.55	0.12	29	0.91
Apnoea type (central/mixed)	3.55	1.23	2.89	29	0.0071**
Apnoea duration	-0.53	0.01	-39.34	29	<0.0001***
Apnoea type (mixed):Apnoea duration	0.14	0.04	3.67	29	0.00098***
Apnoea type (obstructive):Apnoea duration	0.15	0.06	2.55	29	0.016*
Apnoea type (unspecified):Apnoea duration	0.01	0.10	0.09	29	0.93
Apnoea type (central/mixed):Apnoea duration	-0.07	0.07	-1.12	29	0.27

Table 3.4: Results of the statistical analysis investigating if apnoea type modulates the relationship between oxygen saturation and apnoea duration. Fixed effects coefficients from the linear mixed-effects model with the formula: "oxygen saturation fall 1 + ID + apnoea type * apnoea duration". The dataset used in this model included 39 data points from 5 studies. The apnoea types were classified into 5 categories: central apnoea (n = 12), obstructive apnoea (n = 5), mixed apnoea (n = 5), central/mixed apnoea (n = 12), and unspecified apnoea (n = 5). * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$. SE = standard error. DF = degree of freedom.

that theophylline did not reduce the slope of the relationship between apnoea duration and reduction in oxygen saturation (treated: r = 0.45, $p < 0.0001$; untreated: r = 0.24, $p < 0.0001$).

Postmenstrual age

While none of the included studies analysed the effect of age on neonates' responses to apnoea, the meta-analysis of the studies shown in Figure 3.8 revealed that PMA significantly modulated the relationship between changes in oxygen saturation and apnoea duration (Table 3.5). A post hoc analysis showed that for neonates with a PMA of 30, 32, 34, and 36 weeks, the predicted apnoea durations associated with a drop in oxygen saturation greater than 10% were 18.9, 20.0, 21.5, and 23.4 seconds, respectively (i.e. younger neonates have a significant drop in oxygen saturation with shorter apnoeas). In the mixed-effects model analysis comparing percentage change in heart rate with apnoea duration, PMA was not a significant factor (Table 3.6). PMA was not incorporated into the other mixed-effects models due to a lack of information.

Apnoea frequency

None of the included studies reported data that enabled investigation of how apnoea frequency or clustering modulates neonatal physiological responses to apnoea of a given duration. Consequently, the third aim of this chapter could not be addressed due to insufficient available data.

Name	Estimate	SE	t-stats	DF	p Value
(Intercept)	22.44	15.78	1.42	27	0.17
PMA	-0.78	0.54	-1.45	27	0.16
Apnoea type (mixed)	-0.75	1.67	-0.45	27	0.65
Apnoea type (obstructive)	-2.68	2.07	-1.29	27	0.21
Apnoea type (unspecified)	5.89	4.06	1.45	27	0.16
Apnoea type (central/mixed)	7.76	2.99	2.60	27	0.015*
Apnoea duration	-1.68	0.50	-3.34	27	0.0025**
PMA:Apnoea duration	0.04	0.02	2.29	27	0.030*
Apnoea type (mixed):Apnoea duration	0.06	0.05	1.17	27	0.25
Apnoea type (obstructive):Apnoea duration	0.07	0.07	1.03	27	0.31
Apnoea type (unspecified):Apnoea duration	-0.27	0.15	-1.80	27	0.083
Apnoea type (central/mixed):Apnoea duration	-0.28	0.11	-2.58	27	0.016*

Table 3.5: Results of the statistical analysis investigating if PMA and apnoea type modulates the relationship between oxygen saturation and apnoea duration. Fixed effects coefficients from the linear mixed-effects model with the formula: "oxygen saturation fall 1 + ID + apnoea type * apnoea duration + PMA * apnoea duration". The dataset used in this model included 39 data points from 5 studies. The apnoea types were classified into 5 categories: central apnoea (n = 12), obstructive apnoea (n = 5), mixed apnoea (n = 5), central/mixed apnoea (n = 12), and unspecified apnoea (n = 5). * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$. SE = standard error. DF = degree of freedom.

Name	Estimate	SE	t-stats	DF	p Value
(Intercept)	6.86	55.09	0.12	19	0.90
PMA	-0.13	1.52	-0.08	19	0.93
Apnoea duration	-2.70	2.48	-1.09	19	0.29
PMA:Apnoea duration	0.03	0.07	0.46	19	0.65

Table 3.6: Results of the statistical analysis investigating if PMA modulates the relationship between heart rate drops and apnoea duration. Fixed effects coefficients from the linear mixed-effects model with the formula: "heart rate fall 1 + ID + apnoea type * apnoea duration + PMA * apnoea duration". The dataset used in this model included 23 data points from 5 studies. * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$. SE = standard error. DF = degree of freedom.

3.3.6 Quality assessment

A summary of the outcomes of the quality assessments can be found in Tables 3.8, 3.9, 3.10 and 3.11, and the detailed results from the JBI checklists are provided in Tables 3.12, 3.13, 3.14, 3.15 and 3.16. None of the papers satisfied all the requirements in the checklists. The results of overall risk of bias assessments can be found in Table 3.7. Five studies (12.2%) had a low overall risk of bias, 21 studies (51.2%) had a moderate overall risk of bias, and 15 studies (36.6%) had a high overall risk of bias. Among the 15 studies included in the meta-analyses in Figures 3.5 - 3.10 and the combined correlation factors in Section 3.3.2, 7 had a high risk of bias, 6 had a moderate risk, and 2 had a low risk.

Study	Checklist	'No' Responses	Overall Risk of Bias
Gabriel et al. (1976)	Cohort	2	High
Storrs (1977)	Cohort	3	High
Vyas et al. (1981)	Cohort	2	High
Southall et al. (1983)	Cohort	3	High
Henderson-Smart et al. (1986)	Cohort	2	High
Mathew (1988)	Cohort	2	High

Waggener et al. (1989)	Cohort	2	High
Upton et al. (1991)	Cohort	2	High
Spear et al. (1992)	Cohort	2	High
Upton et al. (1992) (A)	Cohort	0	Low
Upton et al. (1992) (B)	Cohort	1	Moderate
Urlesberger et al. (1999)	Cohort	4	High
Pichler et al. (2003)	Cohort	2	High
Fathabadi et al. (2017)	Cohort	3	High
Marshall et al. (2019)	Cohort	1	Moderate
Poets & Southall (1991)	Case-Control	1	Moderate
Elder et al. (2011)	Case-Control	2	High
Daily et al. (1969)	Cross-Sectional	1	Moderate
Fenichel et al. (1980)	Cross-Sectional	2	High
Werthammer et al. (1983)	Cross-Sectional	1	Moderate
Muttitt et al. (1988)	Cross-Sectional	0	Low
Curzi-Dascalova et al. (1989)	Cross-Sectional	0	Low
Suichies et al. (1989)	Cross-Sectional	1	Moderate
Hodgman et al. (1990)	Cross-Sectional	1	Moderate
Mathew et al. (1991)	Cross-Sectional	1	Moderate
Abdulhamid et al. (1992)	Cross-Sectional	1	Moderate
Finer et al. (1992)	Cross-Sectional	1	Moderate
Poets et al. (1993)	Cross-Sectional	1	Moderate
Poets et al. (1995)	Cross-Sectional	1	Moderate
Jenni et al. (1996)	Cross-Sectional	1	Moderate
Adams et al. (1997)	Cross-Sectional	1	Moderate
Carbone et al. (1999)	Cross-Sectional	1	Moderate
Curzi-Dascalova et al. (2000)	Cross-Sectional	0	Low
Nock et al. (2004)	Cross-Sectional	1	Moderate
Cardot et al. (2007)	Cross-Sectional	1	Moderate
Tourneux et al. (2008)	Cross-Sectional	1	Moderate
Beck et.al (2011)	Cross-Sectional	1	Moderate
Mohr et al. (2015)	Cross-Sectional	1	Moderate
Seppa-Moilanen et al. (2021)	Cross-Sectional	0	Low
Varisco et al. (2022)	Cross-Sectional	1	Moderate

Falsaperla et al. (2019)	Case Report	2	High

Table 3.7: Overall risk of bias assessment. For each study, the table reports the number of critical items rated as 'no' by the assessors.

3.4 Discussion

This systematic review and meta-analysis investigated how the duration of apnoea, or pauses in breathing, correlates with changes in heart rate, oxygen saturation, cerebral oxygenation and cerebral blood volume in preterm neonates. Overall, the meta-analyses demonstrated that the minimum value and the extent of decrease in heart rate, oxygen saturation, and cerebral oxygenation during apnoea were significantly correlated with the duration of apnoea. This aligns with the outcomes of most studies included in this review. A meta-analysis on cerebral blood volume could not be performed, as the three relevant studies used different metrics. Change in cerebral blood volume was correlated with apnoea duration in the group of neonates with bradycardia in one study, but not in the non-bradycardia group. The other two studies reported surrogate markers for cerebral blood volume, showing inconsistent results – one reported a significant relationship, while the other did not. Due to this variability, further validation is needed, and no definitive conclusion can be drawn from this systematic review regarding the relationship between apnoea duration and changes in cerebral blood volume.

In addition to summarising individual studies, conducting meta-analyses can quantitatively synthesise the results of heterogeneous cohorts. This could enhance statistical power and improve the accuracy of the investigated relationship between apnoea duration and physiological

responses. Individual studies in this field are often limited by small sample sizes, different definitions of apnoea, and variations in monitoring techniques. When studies are considered in isolation, these factors may lead to inconsistent patterns. By pooling the data and results using appropriate effect estimation, this meta-analysis provides a more reliable assessment than any individual study.

Despite the strong overall associations observed, there was considerable variability in physiological responses to apnoea (Figures 3.5, 3.7, 3.8, and 3.10). The relatively lower R^2 values for heart rate changes compared to oxygen saturation suggest greater inter- and intra-individual variability in heart rate responses. In the included studies, physiological changes were typically expressed as percentage deviations from baseline values measured prior to apnoea onset. Baseline oxygen saturation was generally high and relatively consistent across infants, as clinical intervention is usually initiated when saturation falls below 90% for a sustained period. In contrast, baseline heart rate varies substantially between neonates due to differences in maturity, autonomic regulation, and clinical status, which may partly explain the greater variability observed in heart rate responses. Additional sources of heterogeneity such as differences in measurement techniques and ventilatory support, are also likely contributors.

Variation was observed not only across studies but also across individual apnoeic events within the same infant, as reported in several included papers, including Upton et al. (1991) [7]. Recurrent and pronounced physiological disturbances associated with apnoea may have important clinical implications, both acutely and long term. Previous work has linked frequent or severe desaturation and bradycardia episodes with adverse outcomes such as retinopathy of prematurity [10] and later cognitive impairment [11, 38]. Identifying factors that modulate

the relationship between apnoea duration and physiological instability may therefore help distinguish neonates who are particularly vulnerable to large physiological instabilities following even short apnoeas from those who remain relatively stable.

Several modulating factors were identified in this review, including apnoea type, the use of theophylline, and PMA. The effect of theophylline on the relationship between apnoea duration and heart rate was reported in only one study [141]. In that study, a statistically significant correlation was observed during theophylline treatment and in the combined observations, but not in observations without treatment. However, these findings should be interpreted cautiously. The reported correlations were weak (r = 0.093 in the combined observations and r = 0.16 during treatment), and no formal statistical comparison between treatment conditions was conducted. Importantly, a statistically significant result in one condition but not another does not necessarily indicate a true difference. Furthermore, as theophylline is used to treat apnoea of prematurity, differences between treated and untreated observations should not be interpreted as reflecting distinct underlying causes of apnoea. Although theophylline was associated with a reduction in the mean heart rate decrease during apnoea, the available evidence is insufficient to establish whether it modifies the relationship between apnoea duration and associated physiological responses.

Meta-analysis demonstrated that PMA significantly modulates the relationship between apnoea duration and oxygen saturation, with younger infants experiencing greater reductions in oxygen saturation for a given apnoea duration compared with older infants. This finding is physiologically plausible given the progressive maturation of the respiratory and cardiovascular systems with increasing PMA. As PMA increases, functional residual capacity rises [151], providing a

larger pulmonary oxygen reservoir and delaying desaturation during periods of absent ventilation. Concurrently, the transition from predominantly fetal haemoglobin to adult haemoglobin [152] improves tissue oxygen delivery. Additional maturational changes, including improved lung compliance, enhanced surfactant production, increased airway stability, and stronger respiratory muscles, further support more efficient gas exchange. Collectively, these factors likely explain why older preterm infants tolerate longer apnoeas before exhibiting clinically significant desaturation.

These findings highlight the potential importance of age-stratified apnoea alarm thresholds. However, given that much of the existing evidence derives from relatively small and dated studies, further work is required to determine whether these relationships persist in contemporary neonatal cohorts. This systematic review therefore establishes the rationale for the subsequent chapters of this thesis. In the next chapter, data from a large, modern Oxford cohort will be used to evaluate the influence of PMA on the relationship between apnoea duration and changes in heart rate and oxygen saturation, and to assess whether the findings from this review can be replicated under current clinical practice.

Several studies suggested that central apnoeas may be more strongly associated with physiological instability than obstructive events; however, this evidence remains limited. Few studies systematically distinguished apnoea types, and many provided incomplete data. Moreover, accurate classification of apnoea type requires nasal airflow measurements, which are not routinely recorded in many neonatal monitoring systems. As a result, further validation of this potential modulating effect may be challenging in future observational studies.

The effect of apnoea frequency and clustering on physiological responses could not be assessed in this review, as insufficient studies reported relevant data. This represents an important gap, as clustered apnoeas may impact the way babies respond to an apnoea. The next chapter will address this limitation by examining how recent apnoea history, including apnoea rate and time since the most recent event, modulates the relationship between apnoea duration and changes in heart rate and oxygen saturation.

Investigation of additional modulating factors was limited by incomplete reporting and the small number of eligible studies. As a result, it was not possible to examine the influence of variables such as sex, gestational age at birth, ventilation mode, weight z-score, infection status, or medications other than theophylline. These factors are of particular interest, as sex differences in cardiorespiratory control have been reported in preterm infants [153, 154], ventilation mode may directly influence respiratory mechanics and physiological responses [155, 156], weight z-score reflects relative maturity status and is associated with bradycardia and desaturation incidences [157], and infection is known to affect physiological stability [158]. Similarly, baseline heart rate and oxygen saturation may reflect overall vulnerability to physiological stress. The next chapter will explore the impact of these factors using a large contemporary dataset and incorporate those found to be significant into a model aimed at predicting apnoea-related cardiorespiratory instability.

Inconsistencies in measurement techniques and ventilatory support across studies likely contributed to the observed heterogeneity. Heart rate was measured using modalities including ECG and plethysmography, which differ in temporal resolution and susceptibility to artefact. Definitions and detection thresholds for apnoea, bradycardia, and desaturation varied across

studies. The duration of post-apnoea observation windows used to quantify physiological responses also varied. Additionally, differences in ventilatory support, such as spontaneous breathing, CPAP, or mechanical ventilation, may have influenced both physiological responses and their detection. These methodological variations limit comparability across studies and highlight the need for standardised measurement protocols and improved reporting.

This review was further limited by the small sample sizes of many included studies and by the age of the evidence base, with over half of the studies published more than 30 years ago. Measurement techniques and clinical practices have evolved substantially since then. In contrast, the Oxford dataset analysed in the next chapter comprises 20,626 hours of vital sign recordings collected between 2017 and 2025, offering a more representative and clinically relevant resource for contemporary neonatal care.

Only five studies demonstrated a low overall risk of bias, with the remainder exhibiting moderate or high risk. Common issues included unclear definitions of apnoea and bradycardia, limited reporting of confounders, and insufficient participant characterisation. Notably, fewer than half of studies reported infant sex. Improved standardisation, transparent reporting, and adequately powered prospective designs are needed to strengthen the evidence base and enable more refined meta-analyses in the future.

Despite these limitations, the consistent associations observed across multiple outcomes provide a robust foundation for further investigation. By identifying PMA as a key modulator of apnoea-related desaturation, this review supports the rationale for individualised monitoring strategies in preterm neonates. The next chapter builds directly on these findings by repli-

cating the observed relationships in a contemporary cohort, examining additional modulating factors, and predicting apnoea-related cardiorespiratory instability with the aim of improving personalised healthcare in the neonatal units.

Table 3.8: Summary of included studies reporting the relationship between apnoea duration and change in heart rate. ECG: Electrocardiogram. SD: standard deviation.

Study	Number of neonates	Postmenstrual age at study (weeks)	Definition or inclusion criteria of apnoea/pause in breathing	Method of apnoea/pause measurement	Description of apnoea/pause in breathing	Method of outcome measurement	Main results	Quality assessment
Daily et al. 1969 [6]	22	Premature	Apnoeic episodes of 20 or more seconds.	Impedance plethysmography (impedance pneumography)	A total of 540 apnoeic episodes were observed in 13 infants. 400 episodes occurred in 8 of 14 infants who were otherwise clinically well. 29 episodes occurred in 4 of the 7 infants recovering from respiratory distress syndrome. Of the 540 episodes, 428 exceeded 30 seconds in duration.	ECG	For apnoea >45 seconds, heart rate fell steadily; in episodes of shorter duration, there was a tendency for the heart rate to rise in the first 5 seconds and thereafter to fall.	Inclusion criteria was not stated. No confounding factors identified. No statistical analysis performed.
Gabriel et al. 1976 (A) [121]	8	Range: 31 - 35.	Respiratory arrest: pauses >3 seconds. Apnoeic spells: respiratory arrest of more than 10 seconds duration.	Polysomnography	493 apnoeic spells and 2450 respiratory pauses.	ECG	Pauses <10 s: no correlation between the duration of the respiratory arrest and the incidence of bradycardia. Pauses >= 10 seconds: the incidence of cardiac slowing as well as the percentage of bradycardia increased with longer duration of apnoeas.	Method of exposure and outcome measurements were unclear. No statistical analysis performed.
Gabriel et al. 1976 (B) [126]	8	Range: 31 - 35.	Breathing pauses are categorised into two groups: pauses lasting 3–10 seconds and apnoeas which exceeded 10 seconds.	Polysomnography	n = 2,450 pauses of 3-10 seconds, n = 493 apnoeas	ECG	Most of the breathing pauses (especially apnoeas) are accompanied by bradycardia. The number of bradycardias increased with increasing apnoea duration.	Method of exposure and outcome measurements were unclear. No statistical analysis performed.

Storrs 1977 [136]	9	Range: 27.3 – 34.4.	Apnoea: cessation of spontaneous respiratory effort for >5 seconds coinciding with cardiovascular changes.	Plethysmography	Total of 47 apnoeas	ECG	The heart rate fell by a mean (±SD) of 43% ± 13%, on average 8.6 seconds ± 6.3 (range 2-33 s) after the onset of apnoea, rising on average 4.8 seconds ± 4.4 (range 2-16 s) after the resumption of breathing. The maximum fall in heart rate is correlated to the apnoea duration.	Method of exposure measurement was unclear. Statistical analysis method was also unclear.
Fenichel et al. 1980 [122]	11 newborns, only 6 newborns eligible for this review, the remaining 5 had conceptional age >= 37 weeks.	32; 32; 36; 33; 33; 35.	Not given.	Thermocouple or piezoelectric transducer.	A total of 16 nonconvulsive apnoeic spells were recorded in the eligible patients. Nine of them were between 4-19 seconds, and 7 were >= 20 seconds.	ECG	Alterations in heart rate differed between the two groups (apnoea duration <20 s and >= 20 s). During apnoea lasting less than 20 seconds, heart rate was found to increase, decrease, or remain unchanged. During apnoea lasting 20 seconds or longer, heart rate usually fell by 40% or more within 8 seconds of the onset of apnoea.	The inclusion criteria and the population details were not clearly stated. The measurement of exposure was unclear. The measurement of outcome was unreliable. Confounding factors were not identified. No statistical analysis performed.

Vyas et al. 1981 [110]	7	Mean: 34.1.	Apnoea: cessation of tidal flow for >5 seconds.	Pneumotachograph	Total of 172 episodes of apnoea. Number (percentage): obstructive - 86 (50%); central - 48 (27.9%); mixed - 38 (22.1%). Mean apnoea duration (sec): obstructive - 9.93; central - 8.3; mixed - 10.9.	ECG	The lowest heart rate, the mean duration of heart rate below 100 bpm, and the incidences of bradycardia all differed among the three apnoea type groups. Central apnoea tended to have the shortest duration and the smallest changes in heart rate, whereas obstructive and mixed apnoeas were associated with longer durations and greater changes in heart rate.	No statistical analysis performed.
Werthammer et al. 1983 [137]	8	Premature.	Apnoea: cessation of respiration for >15 seconds associated with a decrease in heart rate. Classed as mild when heart rate decreased by 20 beats per minute below the average of the previous two minutes, and as severe when the heart rate fell below 100 beats per minute.	Acoustic monitoring	Total of 26 apnoeas. 17 mild, 9 severe. Cessation of breathing (seconds) for the 17 mild episodes: mean 18.2, SD 4.3, range 15-30. Information not given for severe episodes.	ECG	During apnoeic spells that became severe, a mean 27.5 ± 9.7 (SD) seconds elapsed after cessation of respiration before the heart rate fell to less than 100 beats per minute, which is longer than the average duration of mild episodes.	The inclusion criteria and the population details were not clearly stated. Confounding factors were not identified. No statistical analysis performed.

Southall et al. 1983 [130]	14 (13 eligible for this review, subject 10 excluded due to infant's age)	Mean: 31.8, range: 28 - 25.	Apnoeic episodes >= 20 seconds and >= 30 seconds.	Impedance respirometer, under-mattress pressure pad, segmented air-filled mattress, pressure capsule system.	A total of 203 apnoeic episodes >= 20 seconds, and 120 episodes >= 30 seconds were identified. Duration of apnoea >= 30 seconds (s): mean: 46, range: 30 - 216.	ECG	Of a total of 203 episodes of apnoea, 145 (71%) were accompanied by bradycardia of 80 bpm. Duration of apnoea is correlated with the duration of accompanying bradycardia.	Confounding factors were not identified. No statistical analysis performed.
Henderson-Smart et al. 1986 [111]	28	Range 28 - 44.	All events of 10 seconds or more in duration were noted.	Mercury in rubber strain gauge, thermistor.	There were 1520 apnoeas of over 10 seconds in duration, 1002 with a duration of 10-14 seconds, 311 with duration 15-20 seconds and 207 of more than 20 seconds.	ECG	The major factor influencing the incidence of bradycardia was duration of apnoea. For each apnoea type there was a significant increase in the incidence of bradycardia as duration of apnoea increased ($p<0.001$ in each case). Time to the onset of the fall in oxygen saturation and to the onset of bradycardia became longer with increasing apnoea duration.	The strategies to deal with confounding factors were not stated.

Muttitt et al. 1988 [141]	27	Preterm.	The apnoea alarm was set to 15 seconds. This alarm notifies staff of a cessation of respiratory input after 10 seconds by way of an intermittent beep which turns into a continuous alarm after 15 seconds.	Impedance respirometry	A total of 1266 apnoeas were documented in the 27 infants. 28% of the apnoeas were less than 20 seconds in duration, 48% were between 20 and 30 seconds, 21% were between 30 and 50 seconds, and 3% were greater than 50 seconds. The longest lasting recorded apnoea was 104.1 seconds.	ECG	A significant correlation was found between apnoea duration and decrease in heart rate (r = .093, P = .001).	The inclusion criteria and the population details were not stated. The measurement of outcome was unclear.
Mathew 1988 [127]	24	Mean: 35.0, standard deviation: 2.0.	Apnoea: the absence of nasal airflow. Prolonged apnoea: apnoea of 20 seconds or longer duration. Short apnoea: apnoea between 10-19 seconds.	Plethysmography	No episodes of apnoea longer than 10 seconds or bradycardia occurred in 9 infants. 39 episodes of apnoea were observed in the remaining 15 infants. 36 episodes were short, 3 prolonged.	ECG	9/36 short apnoeas and 2/3 prolonged apnoeas were associated with bradycardia. Prolonged apnoeas are more likely to be accompanied by bradcardia.	The strategies to deal with confounding factors were not stated. The strategies to address incomplete follow up were not utilised.
Curzi-Dascalova et al. 1989 [135]	23 infants recruited: 7 eligible for this review, 16 excluded as had post-conceptional age >37 weeks).	35-36 week subgroup included in this review	Apnoea: central respiratory pauses >3 seconds.	Polygraphic recordings	602 pauses in total. None of them lasted more than 12 s.	ECG	Pauses lasting 10 to 12 seconds were relatively rare compared to shorter pauses and did not result in greater cardiac deceleration than shorter pauses.	The inclusion criteria and the population details were not stated. The population details and study settings were unclear. The measurement of exposure was unclear. No statistical analysis performed.

Suichies et al. 1989 [150]	18	Preterm	Apnoeic spell: the cessation of breathing for a minimum of 6 seconds.	Thoracic impedance, thermistor	A total of 212 apnoeic spells were analysed. The mean duration of an apnoeic spell was 10.3, 10.2 and 14.3 s for central, obstructive and mixed apnoeic spells, respectively.	ECG	A positive correlation was observed between apnoea duration and the percentage of episodes accompanied by bradycardia. The correlation coefficients varied by apnoea type: central apnoea (r = 0.65, p <0.01), obstructive apnoea (r = 0.51), and mixed apnoea (r = 0.56, p <0.05).	The inclusion criteria and the population details were not stated. The population details and study settings were unclear. The measurement of exposure was unclear. Confounding factors were not identified.
Waggener et al. 1989 [129]	14	Premature	Apnoea: cessation of airflow at mask for a minimum of 3 seconds.	Pneumotachograph, magnetometers.	182 apnoeas >=3 s in duration, 38 apnoeas >= 10 in duration.	ECG	The decrease in heart rate during apnoea was modelled as a function of apnoea duration using multiple linear regression and dichotomous logistic regression. The contingency table based on logistic regression showed that 87 out of 99 events with a decrease of <10 bpm and 24 out of 41 events with a decrease of >=10 bpm were successfully predicted, resulting in an average accuracy of 73%.	Confounding factors were not identified. No statistical analysis performed.

Hodgman et al. 1990 [131]	66	Range: 32 - 36.	The duration of an apnoeic episode was measured from the end of the inspiration before the pause to the beginning of the next inspiration.	Impedance pneumography.	Approximately half of the infants (53%) had no apnoeic episodes of 15 seconds or longer. 25 (38%) infants had one or more 15- to 19-second apnoea, and in 10 infants (15%), the duration was 20 seconds or longer. Four of these 10 infants (6% of the total population) had two or more episodes of apnoea of this duration. There was a total of 94 individual apnoeic episodes of 15 seconds or longer. Twenty of the 94 episodes lasted 20 seconds or longer, and 5 of these were between 30 and 40 seconds in duration.	ECG	A TEB (transient episodes of bradycardia) was defined as a drop in heart rate below 90 beats per minute. While the majority of TEB did not accompany a long apnoeic episode, all but 5 of the apnoeic episodes of 15 seconds or longer were accompanied by a TEB.	The measurement of exposure was unclear. Confounding factors were not identified. Methods of statistical analysis were unclear.
Mathew et al. 1991 [132]	10	Preterm.	Apnoeic episodes: absence of airflow for 5 seconds or greater. Apnoeas lasting >= 15 seconds were termed long apnoeas, whereas those lasting <15 seconds were termed short apnoeas.	Plethysmography, nasal Screen Flowmeter	A total of 352 apnoeic events. Out of the 202 episodes preceded by arousal, most of the apnoeas were short: 181 were less than 15 s, and only 21 were >15 s.	ECG	Hypoxia or bradycardia developed during 24% of long apnoea (>= 15 s) and only 7% of short apnoea (<15 s).	The inclusion criteria and the population details were not stated. Confounding factors were not identified. No statistical analysis performed.

Upton et al. 1991 [7]	27	Median: 32, range: 26 - 36.	Apnoea: cessation of breathing movement for at least 10 seconds, or a shorter period if associated with bradycardia of 90 beats per minute or less.	Thoracic impedance, plethysmography.	A total of 1029 episodes were analysed. Duration of apnoea (s): median 10 (range 0-80) for infant receiving theophylline (45 studies, 535 episodes); median 11 (range 0-31) for infants not receiving theophylline (44 studies, 494 episodes).	ECG	Correlation between minimum heart rate during episodes of bradycardia and apnoea is extremely poor. Minimal heart rate=74.22-0.14*duration of apnoea (r=0.11, $p<0.0001$).	The measurement of exposure was unclear. The strategies to deal with confounding factors were not stated.
Spear et al. 1992 [118]	26	All patients - mean: 34, SD: 2. Patients with prolonged apnoeas - mean 33, SD: 1. Patients with no prolonged apnoeas - mean: 35, SD: 2. p = 0.001.	Prolonged apnoea: >20 seconds, short apnoea: 6 - 20 seconds.	Thoracic impedance, nasal air flow (by thermistor).	Of the 78 traces, 20 had prolonged apnoea and 58 had no prolonged apnoea. Traces had a mean of 16.1 episodes of central apnoea, 3.04 episodes of obstructive apnoea, and 6.1% of periodic breathing.	ECG	The number of bradycardias >5 seconds were greater in the traces with prolonged apnoea comparing to short apnoea.	The two groups were not similar in study weight, post-conceptional age and postnatal age. No statistical analysis performed.

Finer et al. 1992 [123]	47	Premature	Apnoea was included in this study if it lasted at least 15 seconds and was associated with bradycardia, hypoxaemia, or both.	Impedance respirometry	2082 episodes of apnoeas in total.	ECG	As apnoea duration increased, the percentage fall in heart rate decreased, this trend being significantly greater for central and obstructive apnoea than for mixed events. As the duration of central apnoea became longer, untreated infants had a significantly lesser fall in heart rate; apnoea between 40 and 50 seconds in duration was associated with a mean fall in heart rate of 38%, compared with 51% during apnoea lasting 20 to 30 seconds ($p = <0.0001$).	The inclusion criteria and the population details were not stated. The details of the study subjects and the setting were unclear (study period was not stated). The measurement of exposure was unclear. Confounding factors were not identified.
Upton et al. 1992 (A) [159]	27 (24 have available flow recordings).	Median: 32, range: 26 - 36.	Apnoeic episodes of at least 5 seconds duration, as detected on the flow trace, were analysed for comparison.	Thoracic impedance, plethysmography.	From the initial polygraphic recording: 86% of the bradycardic episodes were associated with apnoea. During airway flow measurements: apnoea was present on the flow trace in all 41 episodes in which bradycardia occurred. Median apnoea duration was 14 seconds (range 2-51).	ECG	The duration of bradycardia, expressed by the number of bradycardic beats, was positively correlated with apnoea duration in seconds. There was also a significant positive correlation between apnoea duration and time of onset of bradycardia.	The measurements of exposure and outcome were unclear.

Poets et al. 1993 [44]	80	Mean: 36.3, SD: 2.3.	Apnoeic pauses of 4 seconds or longer associated with bradycardia. An association was considered present if commenced within ±15 seconds of the onset of the bradycardia.	Volume expansion capsule, plethysmography.	143 (83%) of the 172 bradycardias associated with interpretable airflow and breathing movement signals were accompanied by an apnoeic pause.	ECG	The apnoeic pauses associated with bradycardia had a median duration of 10 seconds, with range between 4 and 38 seconds.	The inclusion criteria and the population details were not stated. The details of the study subjects and the setting were unclear (study period was not stated). The measurements of exposure were unclear. Confounding factors were not identified. No statistical analysis performed.
Jenni et al.1996 [139]	17	Preterm	Apnoea: cessation of respiratory effort for 10 seconds or more.	Thoracic impedance, thermistor.	A total of 130 apnoeic episodes were recorded in 14 infants. The median duration of those episodes was 20 s, the 5th and 95th percentiles were 12 and 54, respectively. Each infant had between 4 and 20 episodes of apnoea (mean 9.3) per study period).	ECG	The duration of apnoea was correlated with the lowest heart rate: central apnoea: r = -0.772 ($p < 0.01$), obstructive apnoea: r = 0.104, and mixed apnoea: r = -0.695 ($p < 0.01$).	The inclusion criteria and the population details were not stated. The measurements of exposure were unclear. Confounding factors were not identified.

Urlesberger et al. 1999 [120]	58	Preterm	Apnoeas were divided into 2 groups according to their duration: Group 1 included apnoeas with a duration between 5-14 seconds, and Group 2 included apnoeas >= 15 seconds. If there were changes in oxygen saturation to a degree of >= 3%, and/or in heart rate to a degree of >= 5% of baseline values in either direction, the episode was placed in Subgroup b, otherwise the episode was placed in Subgroup a.	2 peizo respiratory effort sensors.	A total number of 1345 apnoeas in Group 1 and 74 apnoeas in Group 2 were recorded. In Group 1, 587 apnoeas (44%) fulfilled criteria for further analysis, 382 of them were in Subgroup 1a, and 205 were in 1b. In Group 2, 60 apnoeas (81%) fulfilled criteria for further analysis, all of them were placed into Subgroup 2b.	ECG	Group 2b has a longer mean duration of apnoeas and a greater decrease in heart rate comparing to Group 1b.	The exposures were not measured similarly to assign people to both exposed and unexposed groups. The measurements of exposure were unclear. Confounding factors were not identified. Participants were not free of the outcome at the start of the study. Lack of information on follow-up time.

Carbone et al. 1999 [124]	90 infants recruited, multi-channel recording data were complete in 32 infants.	Mean: 36.9.	Episodes of apnoea of >= 10 seconds in duration.	Thermistor.	236 episodes of apnoea analysed. The long apnoea group consisted of 14 apnoeic events of >=20 seconds in duration, the medium apnoea group consisted of 50 apnoeic events of 15 to 19 seconds in duration, and the short apnoea group consisted of 172 apnoeic events of 10 to 14 seconds in duration.	Monitor	Longer apnoeas were significantly correlated with a lower heart rate nadir, a longer duration of heart rate below two-thirds of the baseline, a greater drop in heart rate, and a faster rate of heart rate decline.	The inclusion criteria and the population details were not stated. The details of the study subjects and the setting were unclear (study period was not stated). The measurements of exposure were unclear. Confounding factors were not identified. Methods of statistical analysis were unclear.
Curzi-Dascalova et al. 2000 [148]	5	Mean: 34, range: 33.6 - 34.7.	Apnoea: a 20% or greater decrease in the amplitude of abdominal respiratory movements (detected 2 cm above the umbilicus) or/and airflow.	Thermistors.	492 apnoeas longer than 5 s were detected, of those only 27 were longer than 10 s.	ECG	Multivariate linear analysis showed that the normalised percentage change in HR was dependent on apnoea duration ($p < 0.0001$); regression analysis models explained 24.7% of the changes found, and coefficient of correlation was r = 0.50.	The details of the study subjects and the setting were unclear (study period was not stated). The measurements of exposure were unclear.

Pichler et al. 2003 [8]	Bradycardia group: n = 20; Non-Bradycardia group: n = 19.	Bradycardia group - mean: 34.2, SD: 2.8. Non-Bradycardia group - mean: 35.0, SD: 2.4. p value >0.05	Episodes of apnoea lasting longer than 15 s with bradycardia (bradycardia group) were analysed and compared to episodes of apnoea lasting longer than 15 s without bradycardia (non-bradycardia group).	Piezo-respiratory effort sensors	Twenty-six episodes of apnoea with bradycardia and 26 without bradycardia were observed. The mean duration of apnoea episodes was 25.9 ± 6.0 s in the bradycardia group and 25.0 ± 8.6 s in the non-bradycardia group (p >0.05).	ECG	Longer apnoea durations are associated with greater drops in heart rate in both the bradycardia and non-bradycardia groups.	Unclear whether the two groups were similar and recruited from the same population. Confounding factors were not identified.
Nock et al. 2004 [133]	17	Mean: 35, SD: 1.	Number of apnoeas >=15 seconds and >=20 seconds were recorded. Those lasting >=20 seconds are a subset of the episodes >=15 seconds in length.	Plethysmography.	The subjects had 0 to 46 apnoeic episodes >=15 seconds and 0 to 14 episodes >=20 seconds.	ECG	Heart rate was 119 ± 17 (94-163) bpm during apnoeas >=15 seconds and 110 ± 11 (91-126) bpm during apnoeas >=20 seconds.	The details of the study subjects and the setting were unclear (study period was not stated). The measurements of exposure were unclear. Confounding factors were not identified.
Beck et al. 2011 [125]	10	Mean: 32.1, range: 29.6 - 36.4.	Central apnoeas: periods where the electrical activity of the diaphragm was a flat tracing for more than 5 seconds.	Oesophageal electromyography.	188 central apnoeas identified. Apnoeas were classified by their duration (5 - 10 s, 10 - 15 s, and >15 s).	Not given	Longer apnoea durations were associated with greater drops in heart rate during the events.	The details of the study subjects and the setting were unclear (study period was not stated). The strategies to deal with confounding factors were not stated.

Mohr et al. 2015 [145]	335	Preterm	ABD: Apnoea events accompanied by bradycardia (heart rate <100 bpm) and oxygen desaturation (oxygen saturation <80%). ABD events with apnoea lasting more than 10 seconds were recorded and categorised by duration. Very long apnoea (VLA) was defined as breathing cessation >60 seconds with bradycardia and oxygen desaturation.	Chest impedance-derived respiratory rate by central network server.	86 VLA occurred in 29 infants. The number of events per infant ranged from 1 to 19. In all infants there were 100 events of a duration 10-20 seconds, 100 events of 30-40 seconds, 46 events of 60-80 seconds, and 40 events of 80 seconds.	ECG	The areas under the HR curves increase approximately linearly with the duration of the apnoea. The sample correlation coefficient is 0.81 for heart rate ($p < 0.001$).	The inclusion criteria and the population details were not stated. Confounding factors were not identified.
Marshall et al. 2019 [128]	30	Median: 30, interquartile range: 28-33.	Individual respiratory pauses between 5 and 29 seconds (inclusive) were identified. Events with an FiO_2 (fraction of inspired oxygen) adjustment (defined as a change in FiO_2 of at least 0.01) occurring during the event window were excluded.	Abdominal capsule.	In total, 17677 isolated and 9623 clustered respiratory pause events were identified from a total of 55767 respiratory pauses greater than or equal to 5 seconds. After exclusion of 572 isolated and 443 clustered events because of FiO_2 adjustments within the event window, there were 17105 isolated and 9180 clustered events available for analysis.	Monitor	The mean trajectories of heart rate over the event window demonstrate more prominent instability after longer duration pauses, both for pauses that were isolated and clustered. The proportion of events followed by physiological instability increased with longer pause duration, and was higher for clustered events for all pause durations.	The exposure and outcomes were measured in valid and reliable ways. Confounding factor (clustering of apnoea) and the strategy to deal with the factor were stated.

Falsaperla et al. 2019 [138]	1	29.6	Apnoea: cessation of airflow lasting more than 20 seconds, or cessation of airflow associated with bradycardia (20% below the baseline heart rate), or cessation of airflow for less than 20 seconds with oxygen desaturation less than 80%.	Polygraphic recordings	Number of apnoeic episodes: 8 per day. Duration of episodes: 20-28 secs.	Polygraphic recordings	After about 10 s of central apnoea an abrupt decrease of the child's heart rate (more than 50% variation, from 160 bpm to 65 bpm) was recorded. The drop in the heart rate was followed by a new onset of thoracic efforts, associated with mixed central apnoea at the final stage, which lasted about 10 s, and then the episode resolved with a clear end-point.	The diagnostic tests and the results were not clearly described. The intervention and the post-intervention clinical condition were unclear. The adverse events were not identified.
Seppa-Moilanen et al. 2021 [134]	21	Median: 35.7, interquartile range: 35.0 - 36.3	Apnoea: central respiratory pauses longer than two breathing cycles and lasting 4 seconds or more. All obstructive breaths were noted regardless of the duration of apnoea. AOP: an apnoea with a heart rate decrease to $<$100 beats/min, or a drop in oxygen saturation to under 80%, or it lasted longer than 20 s.	Nasal airflow by a pressure sensor, abdominal band.	Total of 3993 apnoeas in the recordings.	Monitor	For mixed apnoea, events accompanied by arousal were associated with longer durations and higher heart rates. In AOP, events with arousal had longer durations and lower heart rates. For mixed apnoea AOP, events with arousal exhibited longer durations but similar heart rates.	The details of the study subjects and the setting were unclear (study period was not stated).

Table 3.9: Summary of included studies reporting the relationship between apnoea duration and change in oxygen saturation. SD: standard deviation.

Study	Number of neonates	Postmenstrual age at study (weeks)	Definition or inclusion criteria of apnoea/pause in breathing	Method of apnoea/pause measurement	Description of apnoea/pause in breathing	Method of outcome measurement	Main results	Quality assessment
Muttitt et al. 1988 [141]	27	Preterm.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	A significant correlation was found between apnoea duration and decrease in oxygen saturation in the total population of infants with apnoea (r = .388, P <.001). Regression equation: percentage decrease in oxygen saturation = 3.778 + 0.338 * duration of apnoea.	See Table 3.8.
Poets et al. 1991 [45]	31 (16 cases, 15 control)	Cases - mean: 40, range: 36 - 47. Control - mean: 40, range: 35 - 45.	Apnoeic pauses: a cessation in breathing movements for at least 4 seconds. Periodic breathing: a sequence of at least three apnoeic pauses, each separated by not more than 20 breaths.	Graseby pressure capsule, respitrace, thermistor.	Apnoea duration with episodes (s): 1. Cases: periodic - median: 7.8, range: 6.2 - 19.2; non-periodic - median: 11.9, range: 7.3 - 21.0. 2. Controls: periodic - median: 5.4, range: 4.0 - 5.6; non-periodic - median: 6.0, range: 4.0 - 8.5.	Pulse oximetry	The mean duration of apnoeic pauses associated with a hypoxaemic episode was significantly shorter during periodic breathing.	Unclear whether the two groups were comparable. Unclear whether the same criteria used for identification of cases and controls. Confounding factors were not identified.
Mathew et al. 1991 [132]	10	Preterm.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	Hypoxia or bradycardia developed during 24% of long apnoea (>= 15 s) and only 7% of short apnoea (<15 s).	See Table 3.8.

Upton et al. 1991 [7]	27	Median: 32, range: 26 - 36.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	There was a positive correlation between the reduction in oxygen saturation and the duration of apnoea. Reduction in oxygen saturation = 4.57+0.43*duration of apnoea (r=0.41, p<0.0001).	See Table 3.8.
Spear et al. 1992 [118]	26	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	Total number of desaturations <90%, desaturations <80% and <90% in association with apnoea and/or bradycardia, longest duration of desaturation <80%, and total duration of desaturation <90% during sleep were all greater in the tracings with prolonged apnoea.	See Table 3.8.

Finer et al. 1992 [123]	47	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	As apnoea duration increased, the fall in oxygen saturation also increased (r = 0.26; p = 0.0001); this relationship was significant for all apnoea types and did not differ significantly between apnoea types or between treatment groups. There was a significant association between the pre-apnoea mean oxygen saturation and apnoea duration (r =-0.14; p <0.0001).	See Table 3.8.
Abdulhamid et al. 1992 [147]	91 enrolled, 25 eligible.	Mean: 40.5, range: 34 - 88. 25 patients <37.	Central apnoea >= 20 seconds; central apnoea <20 seconds associated with bradycardia <80 bpm for >5 seconds in preterm infants .	Thoracic impedance, nasal / oral airflow.	A total of 184 episodes were recorded.	Pulse oximetry	The duration of obstructive apnoea ranged from 11 to 27 seconds, with 118 episodes (94%) associated with desaturation levels in the 71–84% range. Mixed apnoea episodes lasted between 10 and 45 seconds, with 25 episodes (47%) linked to desaturation levels in the 60–84% range.	The inclusion criteria and the population details were not stated. The details of the study subjects and the setting were unclear (study period was not stated). Confounding factors were not identified. Methods of statistical analysis were unclear.

Upton et al. 1992 (B) [142]	24	Median: 32, range: 26 - 36.	Apnoea: cessation of airflow for 5 seconds or longer.	Pneumotachograph, plethysmography, thoracic impedance.	309 apnoeas. Episodes occurred in 60 of the 83 studies, with a median two apnoeas per study (range 0-19). Mean apnoea duration recorded was 8.3 seconds (range 5-51 seconds).	Pulse oximetry	Positive correlation between oxygen saturation drops during apnoea and apnoea duration (oxygen saturation drop = (apnoea duration * 0.67) - 1.0, r = 0.75, p <0.0001).	Unclear whether the participants were free of outcomes at the start of the study.
Poets et al. 1993 [44]	80	Mean: 36.3, SD: 2.3.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	The interval between onset of apnoeic pause and onset of desaturation has a median of 5.2 seconds, ranged between 0 and 15 seconds.	See Table 3.8.
Poets et al. 1995 [143]	96	Median: 34, range: 29 - 37.	Apnoeic pause of 20 seconds or longer associated with desaturations were recorded. An association was considered present if the apnoeic pause began within 12 seconds of the desaturation.	Volume expansion capsule.	23 desaturations in 7 infants were associated with prolonged apnoeic pauses (>= 20 seconds). The duration of these pauses ranged from 21 to 120 seconds (median, 45 seconds).	Pulse oximetry	There was no significant correlation between the duration of the apnoeic pause and that of the desaturation (r = 0.3; P >.05, Spearman's rank correlation test).	The details of the study subjects and the setting were unclear (study period was not stated). The measurements of exposure were unclear. The strategies to deal with confounding factors were not stated.
Jenni et al. 1996 [139]	17	Preterm	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	The duration of apnoea was correlated with the lowest oxygen saturation as follows: central apnoea: r = -0.486 (p <0.01), obstructive apnoea: r = -0.491(p <0.005), and mixed apnoea: r = -0.637 (p <0.01).	See Table 3.8.

Adams et al. 1997 [144]	30	Mean: 32.8, SD: 2.3.	The duration of all respiratory pauses that led to a fall in oxygen saturation of <80% for >4 seconds.	Plethysmography, thermistor.	156 pauses were recorded (40 apnoeas >15s, 22 pauses with lower EELV (end-expiratory lung volume), 94 pauses with unchanged EELV)	Pulse oximetry	There was no significant correlation between the length of the respiratory pauses and the rate of desaturations (r = -0.198)	The inclusion criteria and the population details were not stated. Confounding factors were not identified. The method of statistical analysis was unclear.
Urlesberger et al. 1999 [120]	58	Preterm	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	Group 2b has a longer mean duration of apnoeas and a greater decrease in oxygen saturation comparing to Group 1b.	See Table 3.8.
Carbone et al. 1999 [124]	90 infants recruited, data were complete in 32 infants.	Mean: 36.9.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	Longer apnoeas were significantly correlated with a lower saturation nadir, a longer duration of saturation <85%, a greater drop in saturation and a faster rate of saturation decline.	See Table 3.8.
Curzi-Dascalova et al. 2000 [148]	5	Mean: 34, range: 33.6 - 34.7.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	There were no significant differences between the amplitude of SaO_2 changes observed following apnoea versus control periods. Linear mixed model analysis failed to demonstrate any significant dependence of the normalised changes in oxygen saturation on apnoea duration.	See Table 3.8.

Pichler et al. 2003 [8]	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	Longer apnoea durations are associated with greater drops in arterial oxygen saturation in both the bradycardia and non-bradycardia groups.	See Table 3.8.
Nock et al. 2004 [133]	17	Mean: 35, SD: 1.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximetry	Oxygen saturation during spontaneous apnoeas >= 15 seconds fell to 86 ± 9%, mean ± SD (range, 62%-99%). During apnoeic episodes >= 20 seconds, saturation fell to 81 ± 11% (58%-92%).	See Table 3.8.
Cardot et al. 2007 [160]	36	Premature	respiratory pause lasting >3 seconds.	Thoracic impedance, pneumotachograph.	Mean apnoea duration was 5.5 ± 1.3 seconds.	Pulse oximetry	The duration of apnoeas with desaturation (6.3 ± 2.0 seconds) was longer compared to the duration of apnoeas without desaturation (5.2 ± 1.0 seconds).	The inclusion criteria and the population details were not stated. The details of the study subjects and the setting were unclear (study period was not stated). The measurements of exposure were unclear. The strategies to deal with confounding factors were not stated.

Tourneux et al. 2008 [161]	29	Late preterm.	Central apnoeic events were defined as respiratory pauses lasting more than 3 seconds.	Plethysmography.	About 1900 apnoeic events were analysed. All lasted less than 20 s.	Pulse oximetry	Apnoeas with desaturation had longer durations compared to apnoeas without desaturation.	The inclusion criteria and the population details were not stated. The details of the study subjects and the setting were unclear (study period was not stated). The strategies to deal with confounding factors were not stated.
Elder et al. 2011 [140]	Cases (n = 16). Controls (n = 18).	Cases - median: 38, range: 35 - 45. Controls - median: 37, range: 35 - 44.	Apnoea: cessation of nasal/oral airflow for >= 5 seconds or a pause of <5 seconds if there was a drop in oxygen saturation to <95% or an associated 3% desaturation.	Thermistor.	Mean apnoea length (s): cases 6.9; controls 6.4. Longest apnoea (s): cases 17.4; controls 12.3. (Log-transformed, values expressed as geometric mean.)	Pulse oximetry	For all infants, there was a significant correlation between the lowest recorded oxygen saturation in active sleep for each infant and the longest apnoea duration recorded for each infant (r = -0.530, p = 0.002). This remained significant when group status was controlled for (r = -0.436, p = 0.01).	The groups were not comparable other than the presence of disease in cases or the absence of disease in controls. The measurements of exposure were unclear. Confounding factors were not identified. No statistical analysis performed.
Mohr et al. 2015 [145]	335	Preterm	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximeter	The area under the oxygen saturation curve increased approximately linearly with the duration of the apnoea (correlation coefficient: 0.88, p <0.001).	See Table 3.8.

Fathabadi et al. 2017 [149]	20	Preterm	Apnoea: a pause in respiration of 4 seconds or greater.	Respiratory motion.	Apnoea occurred concomitant with the hypoxic event in 555 (45 %) cases, including 495 (40 %) in which apnoea preceded the event and 165 events in which it occurred concurrently/subsequently (105 hypoxic events had both preceding and concurrent/subsequent apnoea).	Pulse oximetry	Longer apnoea durations were associated with longer hypoxia durations and a smaller oxygen saturation slope.	Confounding factors were not identified. Unclear whether the participants were free of the outcome at the start of the study. The measurements of outcomes were unclear.
Marshall et al. 2019 [128]	30	Median: 30, interquartile range: 28-33.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oximeter	In pooled data, the mean trajectories of oxygen saturation over the event window demonstrate more prominent instability after longer duration pauses, both for pauses that were isolated and clustered. The proportion of events followed by physiological instability increased with longer pause duration, and was higher for clustered events for all pause durations.	See Table 3.8.

Seppa-Moilanen et al. 2021 [134]	21	Median: 35.7, inter-quartile range: 35.0 - 36.3	See Table 3.8.	See Table 3.8.	See Table 3.8.	Pulse oxymeter	For central apnoea, events accompanied by arousal were associated with similar durations and higher oxygen saturation. For mixed apnoea, AOP and mixed apnoea AOP, events accompanied by arousal were associated with longer durations and higher oxygen saturation.	See Table 3.8.
Varisco et al. 2022 [146]	26	Mean: 29.2.	CASEs: Count of central apnoea-suspected events. CASEs were then divided and grouped according to the length of their cessation of breathing L. The following lengths L were considered: 10s <= L <20s, 20s <= L <30s, 30s <= L <40s, 40s <= L <50s, 50s <= L <60s , 60s <= L <80s , L >= 80s.	Chest impedance.	A total number of 4756 CASEs were recorded. Number of CASEs for 10s <= L <20s: 3140; 20s <= L <30s: 655; 30s <= L <40s: 404; 40s <= L <50s: 202; 50s <= L <60s: 106; 60s <= L <80s: 127; L >= 80s: 122.	Pulse oxymeter	Longer apnoea durations were associated with greater decrease in oxygen saturation. The decrease was 5%, 12%, 17%, 22%, 27%, 32% and 43% for apnoea lengths of 10, 20, 30, 40, 50, 60 and 80 s, respectively.	The details of the study subjects and the setting were unclear (study period was not stated). The strategies to deal with confounding factors were not stated. No statistical analysis performed.

Table 3.10: Summary of included studies reporting the relationship between apnoea duration and change in cerebral oxygenation. cHbD: cerebral haemoglobin oxygenation index.

Study	Number of neonates	Postmenstrual age at study (weeks)	Definition or inclusion criteria of apnoea/pause in breathing	Method of apnoea/pause measurement	Description of apnoea/pause in breathing	Method of outcome measurement	Main results	Quality assessment
Urlesberger et al. 1999 [120]	58	Preterm	See Table 3.8.	2 piezo respiratory effort sensors.	See Table 3.8.	Near-Infrared Spectroscopy	There was poor correlation between apnoea length $\geq$ 15 sec and delta cHbD (R2 = 0.27). There was no correlation between an apnoea length of 5-14 sec and delta cHbD (R2 = 0.06).	See Table 3.8.
Pichler et al. 2003 [8]	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Near-Infrared Spectroscopy	Longer apnoea durations are associated with greater drops in cHbD in both the bradycardia and non-bradycardia groups.	See Table 3.8.

Table 3.11: Summary of included studies reporting the relationship between apnoea duration and change in cerebral blood volume. tHb: total haemoglobin concentration. cHbtot: concentration changes of total cerebral haemoglobin.

Study	Number of neonates	Postmenstrual age at study (weeks)	Definition or inclusion criteria of apnoea/pause in breathing	Method of apnoea/pause measurement	Description of apnoea/pause in breathing	Method of outcome measurement	Main results	Quality assessment
Jenni et al. 1996 [139]	17	Preterm	See Table 3.8.	See Table 3.8.	See Table 3.8.	Near-Infrared Spectroscopy	The duration of apnoea was correlated with the amplitude of tHb as follows: central apnoea: r = 0.549 ($p < 0.01$), obstructive apnoea: r = -0.005, and mixed apnoea: r = 0.7.	See Table 3.8.
Urlesberger et al. 1999 [120]	58	Preterm	See Table 3.8.	2 piezo respiratory effort sensors.	See Table 3.8.	Near-Infrared Spectroscopy	There was no correlation between apnoea length $\geq$ 15 sec and delta cHbtot (R2 = 0.07). There was no correlation between an apnoea length of 5-14 sec and delta cHbtot (R2 = 0.01).	See Table 3.8.
Pichler et al. 2003 [8]	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	See Table 3.8.	Near-Infrared Spectroscopy	Longer apnoea durations are associated with greater drops in cerebral blood volume (CBV) in the bradycardia group, but not in the non-bradycardia group.	See Table 3.8.

Table 3.12: JBI Critical Appraisal Checklist for the included cross-sectional studies (Part A).

Variables assessed	**Daily et al. (1969)**	**Fenichel et al. (1980)**	**Werthammer et al. (1983)**	**Muttitt et al. (1988)**	**Curzi-Dascalova et al. (1989)**	**Suichies et al. (1989)**	**Hodgman et al. (1990)**	**Mathew et al. (1991)**	**Abdulhamid et al. (1992)**	**Finer et al. (1992)**	**Poets et al. (1993)**	**Poets et al. (1995)**
Were the criteria for inclusion in the sample clearly defined?	No	No	No	No	No	No	Yes	No	No	No	No	Yes
Were the study subjects and the setting described in detail?	Yes	No	No	Yes	Unclear	Unclear	Yes	Yes	Unclear	Unclear	Unclear	Unclear
Was the exposure measured in a valid and reliable way?	Yes	Unclear	Yes	Yes	Unclear	Unclear	Unclear	Yes	Yes	Unclear	Unclear	Unclear
Were objective, standard criteria used for measurement of the condition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were confounding factors identified?	No	No	No	Yes	Yes	No	No	No	No	No	No	Yes
Were strategies to deal with confounding factors stated?	N/A	N/A	N/A	Yes	Yes	N/A	N/A	N/A	N/A	N/A	N/A	No
Were the outcomes measured in a valid and reliable way?	Yes	No	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was appropriate statistical analysis used?	N/A	N/A	N/A	Yes	N/A	Yes	Unclear	N/A	Unclear	Yes	N/A	Yes

Table 3.13: JBI Critical Appraisal Checklist for the included cross-sectional studies (Part B).

Variables assessed	**Jenni et al. (1996)**	**Adams et al. (1997)**	**Carbone et al. (1999)**	**Curzi-Dascalova et al. (2000)**	**Nock et al. (2004)**	**Cardot et al. (2007)**	**Tourneux et al. (2008)**	**Beck et.al (2011)**	**Mohr et al. (2015)**	**Seppa-Moilanen et al. (2021)**	**Varisco et al. (2022)**
Were the criteria for inclusion in the sample clearly defined?	No	No	No	Yes	Yes	No	No	Yes	No	Yes	Yes
Were the study subjects and the setting described in detail?	Yes	Yes	No	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	Unclear
Was the exposure measured in a valid and reliable way?	Unclear	Yes	Unclear	Unclear	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes
Were objective, standard criteria used for measurement of the condition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were confounding factors identified?	No	No	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Were strategies to deal with confounding factors stated?	N/A	N/A	N/A	Yes	N/A	No	N/A	No	N/A	Yes	No
Were the outcomes measured in a valid and reliable way?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was appropriate statistical analysis used?	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A

Table 3.14: JBI Critical Appraisal Checklist for the included cohort studies.

Variables assessed	Gabriel et al. (1976)	Storrs (1977)	Vyas et al. (1981)	Southall et al. (1983)	Henderson-Smart et al. (1986)	Mathew (1988)	Waggener et al. (1989)	Upton et al. (1991)	Spear et al. (1992)	Upton et al. (1992) (A)	Upton et al. (1992) (B)	Urlesberger et al. (1999)	Pichler et al. (2003)	Fathabadi et al. (2017)	Marshall et al. (2019)
Were the two groups similar and recruited from the same population?	N/A	N/A	N/A	Yes	N/A	N/A	N/A	N/A	No	N/A	N/A	Yes	Unclear	N/A	N/A
Were the exposures measured similarly to assign people to both exposed and unexposed groups?	N/A	N/A	N/A	Yes	N/A	N/A	N/A	N/A	Yes	N/A	N/A	No	Yes	N/A	N/A
Was the exposure measured in a valid and reliable way?	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes
Were confounding factors identified?	No	No	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes
Were strategies to deal with confounding factors stated?	N/A	No	Yes	N/A	No	N/A	N/A	No	Yes	Yes	Yes	N/A	N/A	N/A	Yes
Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?	Unclear	Unclear	No	No	Unclear	Yes	Unclear	Yes	Unclear	Unclear	Unclear	No	Unclear	Unclear	Unclear

Were the outcomes measured in a valid and reliable way?	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	No	Yes
Was the follow up time reported and sufficient to be long enough for outcomes to occur?	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Was follow up complete, and if not, were the reasons to loss to follow up described and explored?	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	N/A	Unclear	Unclear	Unclear	Unclear	Unclear
Were strategies to address incomplete follow up utilized?	No	No	No	No	No	No	No	No	No	N/A	No	No	No	No	No
Was appropriate statistical analysis used?	N/A	Unclear	N/A	N/A	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.15: JBI Critical Appraisal Checklist for the included case control studies.

Variables assessed	**Poets & Southall (1991)**	**Elder et al. (2011)**
Were the groups comparable other than the presence of disease in cases or the absence of disease in controls?	Unclear	No
Were cases and controls matched appropriately?	Yes	Yes
Were the same criteria used for identification of cases and controls?	Unclear	Yes
Was exposure measured in a standard, valid and reliable way?	Yes	Unclear
Was exposure measured in the same way for cases and controls?	Yes	Yes
Were confounding factors identified?	No	No
Were strategies to deal with confounding factors stated?	N/A	N/A
Were outcomes assessed in a standard, valid and reliable way for cases and controls?	Yes	Yes
Was the exposure period of interest long enough to be meaningful?	Yes	Yes
Was appropriate statistical analysis used?	Yes	N/A

Table 3.16: JBI Critical Appraisal Checklist for the included case report study.

Variables assessed	**Falsaperla et al. (2019)**
Were patient's demographic characteristics clearly described?	Yes
Was the patient's history clearly described and presented as a timeline?	Yes
Was the current clinical condition of the patient on presentation clearly described?	Yes
Were diagnostic tests or assessment methods and the results clearly described?	No
Was the intervention(s) or treatment procedure(s) clearly described?	Unclear
Was the post-intervention clinical condition clearly described?	Unclear
Were adverse events (harms) or unanticipated events identified and described?	No
Does the case report provide takeaway lessons?	Yes

Chapter 4

Risk of apnoea-related cardiorespiratory instability

4.1 Introduction

The systematic review and meta-analysis in Chapter 3 revealed significant relationships between apnoea duration and changes in heart rate and oxygen saturation, as well as notable variance in these relationships, indicating that shorter apnoeas can sometimes lead to severe outcomes. For example, PMA has been demonstrated to be a significant modulating factor of how babies respond to apnoeas, with younger neonates more likely to exhibit oxygen desaturation from short apnoeas. It also highlighted a gap in the investigation of factors that modulate these relationships. Therefore, it is important to identify additional modulating factors to help determine which infants are more likely to experience significant cardiorespiratory instability (bradycardia and/or desaturation) during or after an apnoea.

For instance, caffeine is commonly used as a treatment for apnoea and may alter the physiological response to apnoea events [15, 16]. Similarly, infection can affect an infant's overall physiological condition. Different ventilation modes can also support oxygen saturation levels, potentially preventing the occurrence of desaturations. Therefore, I aimed to explore how these variables influence infants' responses to apnoea. These features could also serve as predictors

for forecasting the degree of cardiorespiratory instability following apnoea episodes of specific durations in individual infants.

The aim of this study was to

1. Reproduce the results from the systematic review using the data collected at the John Radcliffe Hospital, Oxford.
2. Explore which factors either influence the risk of cardiorespiratory instability from apnoea and/or modulate the relationship between cardiorespiratory instability and apnoea duration.
3. Determine whether clinical, demographic and dynamic (i.e., changing physiological parameters which vary within an individual) factors can predict which apnoeas will result in cardiorespiratory instability.

Successful completion of this study can help determine which infants are most at risk of significant cardiorespiratory changes from apnoea, ultimately leading to the development of predictive models and from these personalised alarm limits.

The main results from this chapter are currently under review at *Pediatric Research*, and a minor revision decision has been received. A preprint of this manuscript is available on medRxiv [2].

4.2 Methods

4.2.1 Study design and data collection

The 181,511 eligible apnoeas from Dataset 1 (N = 257 recording sessions in 146 infants) as described in Section 2.2 were used to detect apnoeas and cardiorespiratory instability associated with it. In this study, apnoea is defined as a pause in breathing greater than 5 seconds. In another word, IBIs greater than 5 seconds after filtering using the method described in Section 2.3 are included as apnoea events. An apnoea is defined to be associated with cardiorespiratory instability if it is associated with bradycardia and/or desaturation.

The baseline period was defined as the 90 seconds preceding the start of each apnoea episode, and the post-event period was defined as the interval from the onset of an apnoea episode to 10 seconds after its end. It is possible to have other apnoea episodes in the 90 seconds baseline periods, and the time since last apnoea was also included as a feature in the analyses. These time frames were chosen to capture delayed cardiorespiratory effects, as changes in heart rate and oxygen saturation may lag behind apnoea due to physiological response times and pulse oximeter delay, and are typically fully captured within 10 seconds after the end of the apnoea [146]. The window was not extended further to minimise potential confounding and to ensure that observed changes were associated with the apnoea.

Relevant demographic and clinical information for each infant was extracted from medical records for each recording session. The demographic and clinical indicators investigated as potential modulating factors are defined as follows:

- Sex: the biological sex at birth.
- PMA: postmenstrual age in weeks at the time of recording.
- GA: gestational age in weeks at the time of birth.
- Weight z-score: the infant's weight z-score at the time of recording, relative to their PMA, calculated according to the guideline published by the RCPCH [162], using the UK-WHO preterm weight data as reference [163].
- Ventilation: the mode of ventilation the infant was receiving at the time of recording.
- Apgar (5 min): Apgar score at five minutes after birth.
- Caffeine: whether the infant was receiving caffeine medication at the time of recording.

The dynamic features considered as modulators are defined as follows:

- HR baseline: the mean heart rate during the baseline period (90 seconds preceding the apnoea).
- SpO_2 baseline: the mean oxygen saturation during the baseline period (90 seconds preceding the apnoea).
- Time since last apnoea: the interval between the end of the most recent apnoea and the start of the apnoea of interest in seconds.
- 5-min apnoea rate: the number of apnoeas within the last 5 minutes.
- Duration: the duration of the apnoea in seconds.

4.2.2 Data analysis

Resampling and capping

As the number of apnoeas per infant and per recording had a very wide range, to ensure the results were not overly influenced by infants with higher numbers of apnoeas, I generated 500 resampled datasets by randomly sampling apnoea events within each recording, capping the number at 200 per recording. This approach reduced the influence of recordings with large numbers of events while allowing estimation of parameter variability. All the statistical and machine learning analysis were performed on resampled datasets and aggregated using the methods below. A flowchart of the analysis steps and modelling procedure is presented in Figure 4.1.

Descriptive analysis of physiological responses by apnoea duration

For visualisation purposes, the associated physiological responses were compared with apnoea duration using sliding windows of 5 seconds' duration with 70% overlap. Windows with fewer than 20 data points, or derived from fewer than three unique recordings, were excluded from the plots. Within each window, summary statistics were estimated on 500 resampled datasets. For each resampled dataset and duration window, the mean change in heart rate, mean minimum oxygen saturation, and the proportion of apnoeas associated with bradycardia and/or desaturation (apnoea with bradycardia/desaturation) were calculated. Across replicates, the distribution of each statistic was used to derive the mean estimate and corresponding 95% con-

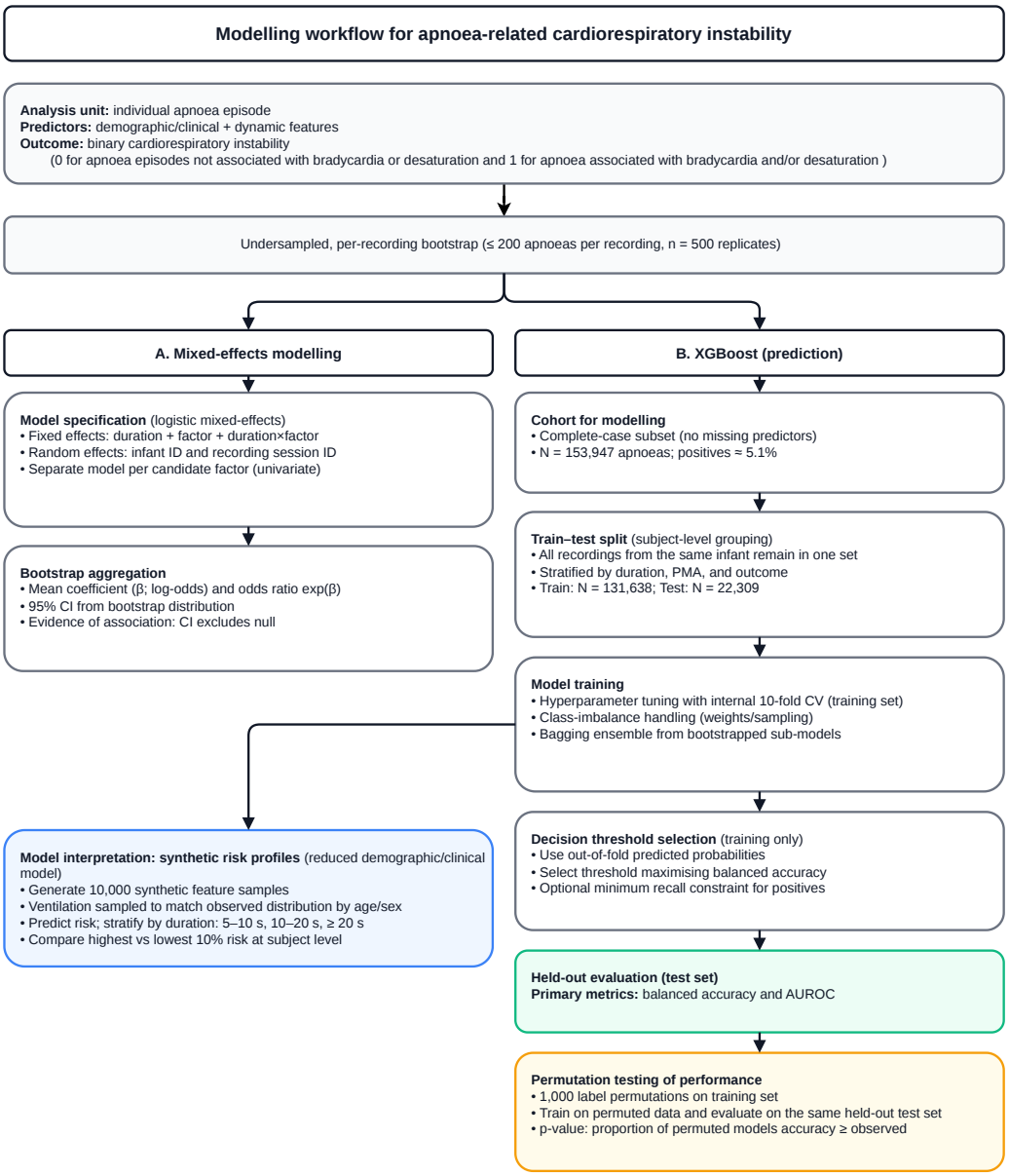

Figure 4.1: Flowchart of the modelling procedure. CV = cross-validation, AUROC = area under the receiver operating characteristic curve.

fidence intervals. In addition, Pearson's correlation coefficients were computed to quantify the linear association between apnoea duration and each physiological outcome. The correlation coefficients were then compared with the results of the previously conducted systematic review [1].

Mixed-effects modelling of factors associated with apnoea-related instability

This analysis aimed to distinguish between overall effects of each factor on the risk of cardiorespiratory instability (i.e., the likelihood of an apnoea of any duration being associated with bradycardia and/or desaturation), and interaction effects indicating whether a factor modulates the relationship between apnoea duration and instability. In other words, I examined both general risk differences between infants and whether specific characteristics alter the duration–response association.

Generalised linear mixed-effect models [164, 165] with a logistic link function were employed on the resampled datasets to analyse the influence of various factors on the relationship between apnoea duration and physiological responses. The interaction terms between apnoea duration and each factor of interest (sex, PMA, GA, weight z-score, ventilation, Apgar score at 5 min, heart rate baseline, oxygen saturation baseline, time since last apnoea, and the number of apnoeas in the past 5 minutes) were included as fixed effects. Recording session ID and infant ID were included as random effects to account for repeated measures within infants. The response variable was binary, defined as 0 for apnoea episodes not associated with bradycardia or desaturation, and 1 for apnoea episodes associated with either bradycardia, desaturation, or

both. The formulae of the model were:

$$\mathrm{y_i} \sim \text{Bernoulli}(\mathrm{p_i}) \tag{4.1}$$

$$\text{logit}(\mathrm{p_i}) = \beta_0 + \sum_{\mathrm{k}=1}^{\mathrm{n}} \beta_{\mathrm{k}} \mathrm{X_{ki}} + \mathrm{b_i} \tag{4.2}$$

where:

- $\mathrm{y_i}$ is the binary response variable for observation i,
- $\mathrm{p_i} = \Pr(\mathrm{y_i} = 1)$ is the probability of a positive outcome for observation i,
- β_0 is the fixed intercept,
- β_{k} is the coefficient for the k-th fixed effect predictor $\mathrm{X_{ki}}$, for $\mathrm{k} = 1, \ldots, \mathrm{n}$, where n is the total number of fixed effect predictors,
- $\mathrm{b_i}$ is the random effect associated with group i, $\mathrm{b_i} \sim \mathcal{N}(0, \sigma_{\mathrm{b}}^2)$,
- $\text{logit}(\cdot)$ is the logit link function defined as $\text{logit}(\mathrm{p}) = \log\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right)$.

The model coefficients and associated p-values from all replicates were extracted and combined across resampled datasets. For each fixed-effect term, I calculated the mean coefficient (representing the average effect size). The mean coefficient corresponded to the log-odds ratio, and its exponential transformation was used to obtain the odds ratio. I also derived the 95% confidence intervals (CIs) for each coefficient and determined whether the aggregated coefficients were significant by assessing whether the 95% CI crossed the null value.

In addition, figures were generated using the resampled datasets in a similar manner to that described in Section 4.2.2, but stratified by each feature. Stratification thresholds were selected

to reflect either clinically meaningful cut-offs or values close to the median of the respective feature.

Machine learning prediction of apnoea-related instability

I constructed binary classification models using the eXtreme Gradient Boosting (XGBoost) algorithm to predict apnoea-related bradycardia and/or desaturation events from clinical, demographic, and dynamic features. XGBoost is a gradient-boosted decision tree algorithm that has demonstrated strong predictive performance for structured tabular data and is well suited to modelling complex nonlinear relationships and interactions among predictors while incorporating regularisation to reduce overfitting. In addition, XGBoost provides measures of feature importance that facilitate interpretation of the factors contributing to model predictions [166].

Model development was restricted to a complete-case subset with no missing feature values (N = 153,947 apnoeas), and apnoea episodes with missing features were excluded (N = 27,564). Three binary classification models were constructed: Model 1 predicted apnoea-related bradycardia events, Model 2 predicted apnoea-related desaturation events, and Model 3 predicted events associated with either bradycardia and/or desaturation. For each model, the dataset was partitioned at the subject level into a training set and a held-out test set, stratified by apnoea duration, PMA, and the response variable. All recording sessions from the same infant were assigned to the same set. The number of apnoeas in each class is summarised in Table 4.1.

Sub-models were trained on 100 independently resampled versions of the training data and combined via averaging. Model training involved tuning of hyperparameters via random search

	Model 1	Model 2	Model 3
Class 0 definition	No bradycardia	No desaturation	No bradycardia/desaturation
Class 1 definition	Any bradycardia	Any desaturation	Any bradycardia/desaturation
Class 0 size (N)	151,671 (97.9%)	148,994 (96.2%)	146,869 (94.8%)
Class 1 size (N)	3,191 (2.1%)	5,868 (3.8%)	7,993 (5.2%)
Train set size (N)	108,104 (69.8%)	105,514 (68.1%)	121,059 (78.2%)
Test set size (N)	46,758 (30.2%)	49,348 (31.9%)	33,803 (21.8%)

Table 4.1: Summary of the three binary classification models.

with 500 iterations, using internal 10-fold cross-validation on the training set. Table 4.2 lists the parameters and instances that were explored.

Parameter	Instances
Number of estimators (`n_estimators`)	200, 400, 800, 1000
Maximum tree depth (`max_depth`)	3, 5, 7, 10, 12
Learning rate (`learning_rate`)	0.001, 0.01, 0.05, 0.1, 0.2
Subsample ratio (`subsample`)	0.5, 0.7, 0.85, 1.0
Column subsample ratio (`colsample_bytree`)	0.4, 0.6, 0.8, 1.0
Minimum child weight (`min_child_weight`)	1, 3, 5, 7
Minimum loss reduction (`gamma`)	0, 0.1, 0.5, 1.0
L1 regularisation (`reg_alpha`)	0, 0.01, 0.1, 1.0
L2 regularisation (`reg_lambda`)	0.5, 1.0, 2.0, 5.0
Class weight scaling (`scale_pos_weight`)	1, 10, 20

Table 4.2: The parameters and instances explored during the hyperparameter optimisation.

Out-of-fold (OOF) predicted probabilities were generated on the training set. Classification thresholds were selected by evaluating all candidate thresholds on the OOF predictions and choosing the threshold that maximised balanced accuracy. An ensemble model was constructed from the models trained on all resampled versions of the training data. Model performance was

then evaluated on the independent test set using balanced accuracy and area under the receiver operating characteristic curve (AUROC).

Strategies to deal with unbalanced data during machine learning

As the datasets were highly imbalanced with respect to the response variables, I applied several approaches to the training set.

1. Undersampling: randomly removing data from the majority classes to match the number of samples in the minority class.

2. SMOTE (Synthetic Minority Over-sampling Technique): this is an oversampling method. It uses the k-nearest neighbours algorithm to find the k closest samples in the minority class, randomly selects one of them, and creates a new point between the two existing points. This process is repeated until the number of samples in the class reaches the desired level, which in this case was when the number of data points in each class became equal [167, 168].

3. Combination of undersampling and SMOTE: I used SMOTE to increase the size of the minority class moderately, and applied undersampling to the majority class to achieve balance.

4. Assigning weights to the model: instead of removing or generating data points, I used the original imbalanced dataset and assigned weights to each class during training. Weights were computed based on the bootstrap sample for each tree grown and were inversely proportional to the class frequencies in the input data.

After evaluating the performance of the different approaches, the models performed best with the fourth approach (assigning class weights), and all machine learning results presented in this chapter are based on this method.

Permutation testing of model performance

To evaluate whether the XGBoost classification model performs significantly better than chance, permutation tests were conducted for every model [169, 170]. This method involves generating a null distribution of predictive performance by training the classifier on data (1,000 permutations) in which the class labels in the training set were randomly permuted. For each permutation, a new XGBoost model was trained and evaluated on the same test set. The p-value was then calculated as the proportion of permuted models that achieved an accuracy equal to or greater than that of the original model. The evaluation was done on the fixed, held-out test set, which was not used in any stage of model training or tuning.

Model-interpretation analysis of predicted risk profiles using synthetic data

To characterise the combined feature profiles associated with higher and lower predicted risk of apnoea-related bradycardia/desaturation I used the reduced demographic/clinical model to identify the types of infants classified as relatively high or low risk within each apnoea-duration band. This analysis served as a model-interpretation step, aiming to clarify the clinical meaning of the multivariable predictions. It is profile-based rather than variable-based and therefore complements, rather than replaces, the mixed-effects analyses of individual factors and inter-

actions. Moreover, the purpose was descriptive and interpretative, not causal, and should not be considered a new independent test of the contribution of individual variables.

Synthetic data were generated by randomly sampling each feature from a uniform distribution 10,000 times. Ventilation mode was sampled categorically, with the probability of each mode set to match the observed distribution of ventilation modes among infants of the corresponding age and sex in the real dataset. The optimised model which only used demographic and clinical variables was applied to the synthetic dataset to generate predicted risk scores for apnoea with bradycardia/desaturation for each observation. Predicted events were stratified by apnoea duration into three categories: 5–10 seconds, 10–20 seconds, and $\geq$20 seconds. Within each duration category, comparisons were conducted between high- and low-risk groups, defined as the highest and lowest 10% of predicted risk, respectively. To avoid pseudo-replication arising from multiple observations per recording, all risk-group comparisons were performed at the subject level. Specifically, subjects were ranked according to their mean predicted risk within each duration category, and those in the highest and lowest 10 percentiles were compared. The comparisons were done on demographic or clinical variables only. For continuous variables, group differences were assessed using Welch's two-sample t-test, while categorical variables were compared using Fisher's exact test or Pearson's χ^2 test, as appropriate. All statistical tests were two-sided, with significance evaluated using conventional thresholds ($p < 0.05$, $p < 0.01$, $p < 0.001$).

4.3 Results

A total of 181,511 apnoeas (pauses in breathing for at least 5 seconds) were included in the analysis, from 257 recording sessions in 146 infants. An average of 287 apnoeas (median, interquartile range: 40–914) occurred per recording, and the average number of apnoeas per hour was 6.8 (median, interquartile range: 3.7–13.0).The details of all included apnoea episodes can be found in Table 4.3.

Variables	**All apnoeas (N = 183,053)**
Apnoea duration (s)	6.8 (5.0 – 117.3)
HR baseline (bpm)	158.2 (100.0 – 200.0)
SpO_2 baseline (%)	96.3 (85.0 – 100.0)
Number of apnoeas in last 5 minutes	2 (0 – 24)
Time since last apnoea (seconds)	84.9 (2.0 – 954614.2)
Minimum heart rate post-event (bpm)	150.0 (29.0 – 209.0)
Min oxygen saturation post-event (%)	94.2 (1.9 – 100.0)
Change in heart rate post-event (%)	-4.3 (-82.4 – 60.0)
Change in oxygen saturation post-event (%)	-1.6 (-98.0 – 17.5)

Table 4.3: Details of all included apnoea episodes. Value expressed as median (range) or percentage.

4.3.1 Do the results of the systematic review hold in a contemporary data set?

Overall, heart rate and oxygen saturation significantly decreased with increasing apnoea duration (heart rate: r = -0.84, $p < 0.0001$; oxygen saturation: r = -0.95, $p < 0.0001$; Figure 4.2A-B). Similarly, the proportion of apnoeas with bradycardia/desaturation significantly in-

creased with apnoea duration (r = 0.80, p $<$ 0.0001, Figure 4.2C). This result aligns with the finding from Chapter 3, where I found that heart rate and oxygen saturation dropped more with increasing apnoea duration.

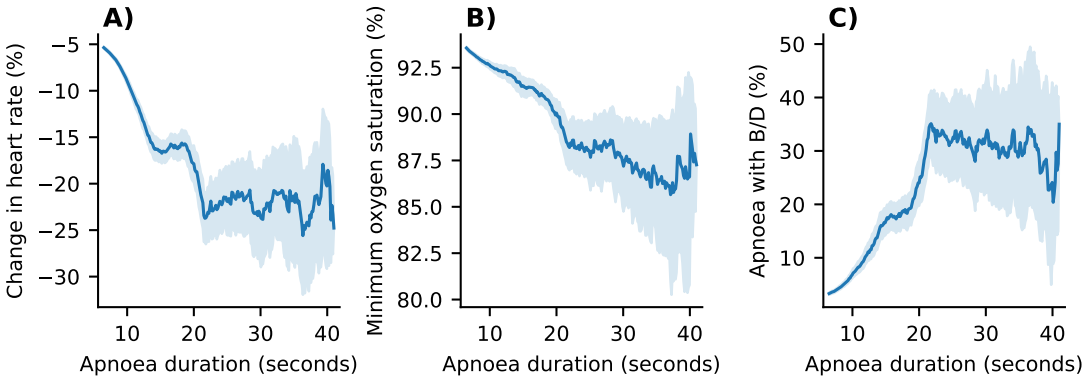

Figure 4.2: Physiological responses compared with apnoea duration. (A) Change in heart rate (%) relative to baseline compared with apnoea duration. (B) Minimum oxygen saturation (%) observed during apnoeas compared with apnoea duration. (C) Proportion (%) of apnoeas associated with bradycardia and/or desaturation (B/D) compared with apnoea duration. Apnoeas were grouped by duration using overlapping windows (70% overlap). Solid lines represent the mean estimates across bootstrap replicates, and shaded regions indicate the corresponding 95% confidence intervals.

Nevertheless, there was wide variation in cardiorespiratory changes for apnoeas of a given duration; even with short pauses in breathing of 5-10 seconds, bradycardia/desaturation occurred in 3.6% (a total of 5617 short pauses in breathing from 95 infants, on average 6.34 per day in these infants). Conversely, for 61.2% of apnoeas (N = 2214) longer than or equal to 20 seconds no bradycardia or desaturation occurred.

Mixed-effect models were first applied following the same approach used in the systematic review, incorporating interaction terms between PMA and apnoea duration as fixed effects and the entire dataset. While PMA alone was not found to be a significant predictor, the interaction between PMA and apnoea duration was statistically significant in relation to both heart rate and oxygen saturation changes. Specifically, for heart rate change, the effect estimates were:

Duration = 0.18 (p = 0.16), PMA = 3.37 $\times 10^{-5}$ (p = 1.00), and PMA * Duration = -0.04 ($p < 0.0001$). For oxygen saturation change, the estimates were: duration = -0.03 (p = 0.61), PMA = 0.05 (p = 0.17), and PMA * duration = 0.88 $\times 10^{-3}$ ($p < 0.0001$).

4.3.2 Factors modulating the apnoea-associated cardiorespiratory instability

To better understand the variability observed in these relationships, additional factors were incorporated into the linear mixed-effects models, with bootstrapping methods applied as described in Section 4.2, and response variable defined as whether or not the apnoea was associated with bradycardia and/or desaturation.

Demographic and clinical factors increase risk of cardiorespiratory instability

PMA at study, GA at birth, mode of ventilation, weight z-score and sex all had a significant effect overall (i.e., irrespective of apnoea duration) on the likelihood of apnoea-related bradycardia and/or desaturation (mixed effects models, Table 4.4, Figure 4.3 columns A and C). Infants with a younger PMA, younger GA, receiving high flow therapy at the time of study, smaller babies and those who are male were more likely to have substantial cardiorespiratory instability (Figure 4.3, columns A and C). Apgar score at birth did not have a significant impact on the likelihood of bradycardia/desaturation (Table 4.4, Figure 4.3).

I also considered which factors significantly modulated the relationship between apnoea duration and the proportion of apnoeas with substantial cardiorespiratory instability. PMA at study, weight z-score, sex, and caffeine were all significant modulating factors (Table 4.4, Figure 4.3

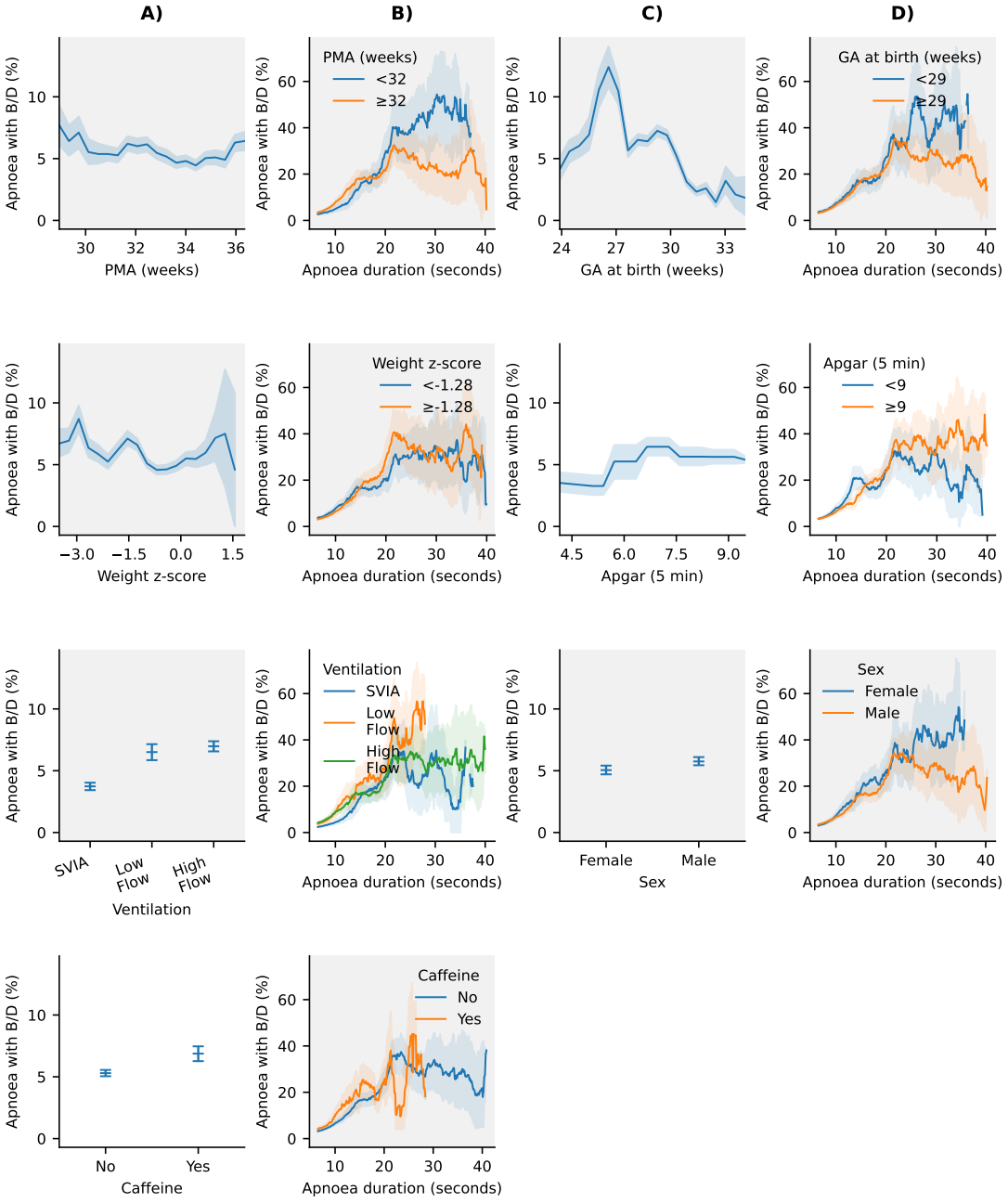

Figure 4.3: Influence of demographic and clinical factors on apnoea-related bradycardia and/or desaturations (B/D). Columns A) and C) illustrate the proportion of apnoeas associated with bradycardia/desaturation compared with demographic and clinical variables. Columns B) and D) show the relationship between apnoea duration and the proportion of apnoeas with bradycardia/desaturation, stratified by the same variables (indicating whether the factor modulates the relationship between apnoea duration and occurrence of bradycardia/desaturation, with stratification thresholds selected to reflect either clinically meaningful cut-offs or values close to the median of the respective feature). Apnoeas were grouped by duration using overlapping windows. Solid lines represent mean estimates across bootstrap replicates, and shaded regions indicate the corresponding 95% confidence intervals. Background grey shading indicates significant factors (see Table 4.4). For columns B) and D) continuous variables were split at either clinically meaningful thresholds or at their median values.

columns B and D). In contrast, mode of ventilation, Apgar score and time since the most recent apnoea were not significant modulating factors.

Interestingly, although male infants were more likely to experience bradycardia/desaturation overall, the relationship with apnoea duration differed between sexes. As shown in Figure 4.3 (column D), male infants generally had a higher proportion of apnoeas associated with bradycardia/desaturation at shorter durations, whereas female infants appeared to have a higher proportion for apnoeas longer than approximately 25 seconds. This suggests that the effect of sex on cardiorespiratory instability may vary with apnoea duration. However, this pattern should be interpreted cautiously, particularly for longer apnoeas where fewer events may contribute to the estimates.

Multiple dynamic features increase the risk of cardiorespiratory instability

Baseline heart rate, baseline oxygen saturation, and the number of apnoeas in the preceding five minutes all had a significant effect on the likelihood of apnoea-related bradycardia/desaturation (mixed-effects models; Table 4.4, Figure 4.4 columns A and C). Infants with lower baseline oxygen saturation values and a greater number of apnoea episodes within the preceding five minutes were more likely to experience substantial cardiorespiratory instability (Figure 4.4, columns A and C). These variables were also significant modulators of the relationship between apnoea duration and the proportion of apnoeas associated with bradycardia/desaturation (Table 4.4; Figure 4.4, columns B and D). In contrast, time since the last apnoea did not significantly affect the likelihood of bradycardia/desaturation and was not a significant modulating factor. This may be because the relationship between apnoea history and physiological responses is

not linear and can be influenced by multiple other factors, such as the duration and severity of the preceding apnoea.

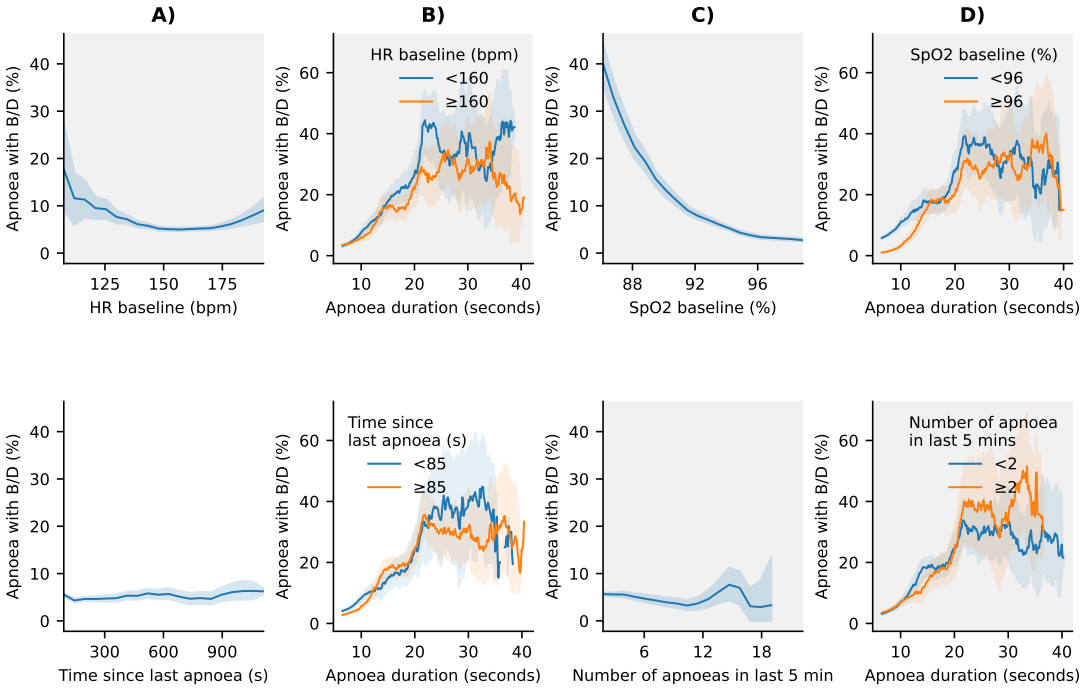

Figure 4.4: Influence of dynamic variables on apnoea-related bradycardia and/or desaturations (B/D). Columns A) and C) illustrate the proportion of apnoeas associated with bradycardia/desaturation compared with key dynamic variables. Columns B) and D) show the relationship between apnoea duration and the proportion of apnoeas with bradycardia/desaturation, stratified by the same variables (indicating whether the factor modulates the relationship between apnoea duration and occurrence of bradycardia/desaturation, with stratification thresholds selected to reflect either clinically meaningful cut-offs or values close to the median of the respective feature). Apnoeas were grouped by duration using overlapping windows. Solid lines represent mean estimates across bootstrap replicates, and shaded regions indicate the corresponding 95% confidence intervals. Background grey shading indicates significant factors (see Table 4.4). For columns B) and D) continuous variables were split according to their median value.

Term	**Significant fraction (%)**	**Mean odds ratio**	**Lower 95% CI**	**Higher 95% CI**	**Significance**
PMA at study	100.0	1.16	1.09	1.21	***
SpO_2 baseline	100.0	0.71	0.68	0.72	***
Ventilation	99.4	1.29	1.15	1.43	***
weight z-score	97.0	0.84	0.77	0.92	***
HR baseline	88.2	1.01	1.00	1.02	**
Sex	87.2	1.36	1.14	1.67	***
GA at birth	86.8	0.93	0.89	0.98	***
Number of apnoeas in last 5 min	81.4	0.94	0.90	0.98	*
Caffeine	9.6	1.10	0.88	1.38	
Apgar (5 min)	7.4	0.99	0.94	1.06	
Time since last apnoea	5.6	1.00	1.00	1.00	
Duration:PMA at study	100.0	0.99	0.98	0.99	***
Duration:HR baseline	100.0	1.00	1.00	1.00	***
Duration:SpO_2 baseline	100.0	1.01	1.01	1.01	***
Duration:weight z-score	96.6	1.01	1.01	1.02	***
Duration:Number of apnoeas in last 5 min	76.0	1.01	1.00	1.01	*
Duration:Sex	70.2	0.98	0.96	1.00	**
Duration:Caffeine	69.6	1.03	1.01	1.05	*
Duration:GA at birth	65.6	1.00	0.99	1.00	
Duration:Apgar (5min)	21.0	1.00	0.99	1.00	
Duration:Ventilation	15.8	1.00	0.99	1.01	
Duration:Time since last apnoea	13.6	1.00	1.00	1.00	

Table 4.4: Results from univariate mixed-effects models examining the effects of individual factors and their interactions with apnoea duration on the likelihood that an apnoea is associated with bradycardia and/or desaturation. Significant fraction indicates the percentage of 500 bootstrap replicates in which the effect was significant (see Section 4.2). CI = confidence interval. * significant at the 95% CI, ** significant at the 99% CI, *** significant at the 99.9% CI.

4.3.3 Cardiorespiratory instability from apnoea can be predicted using machine learning

As a first step towards personalised alarm limits, I investigated if it is possible to predict apnoeas associated with bradycardia/desaturation. I included all demographic, clinical and dynamic factors as inputs to the model. Clinically, it is not always possible to obtain dynamic features, and adjusting apnoea alarm thresholds based on changing physiological information can be impractical. Therefore, I also investigated the possibility of using only demographic and clinical variables included as features, as these are relatively stable within an infant. The results of all models described in Table 4.1 are presented in Table 4.5.

Model	**Train Accuracy**	**Test Accuracy**	**AUROC**
Model 1 with all features	83.2%	85.6%	0.91
Model 1 with demographic features only	84.5%	82.6%	0.88
Model 2 with all features	74.8%	74.8%	0.84
Model 2 with demographic features only	59.2%	56.9%	0.62
Model 3 with all features	75.6%	75.8%	0.83
Model 3 with demographic features only	62.8%	66.0%	0.72

Table 4.5: Performance of the optimised XGBoost models. Both train and test accuracies are balanced accuracies. Model 1: predicting apnoea-related bradycardia events, model 2: predicting apnoea-related desaturation events, model 3: predicting apnoea-related bradycardia and/or desaturation events. Definitions of the response variables and dataset sizes can be found in Table 4.1. AUROC = area under the receiver operating characteristic curve.

Predicting apnoea-related bradycardia events

Following random search optimisation, the best-performing models for predicting apnoea-related bradycardia events correctly identified 87.1% of apnoeas associated with bradycardia

and 84.6% of those not associated with bradycardia when all features were included (Figure 4.5A). When restricted to demographic and clinical features, the model correctly identified 82.9% of apnoeas associated with bradycardia and 82.3% of those not associated with bradycardia (Figure 4.5B). Both models performed significantly better than chance (permutation test, $p < 0.0001$).

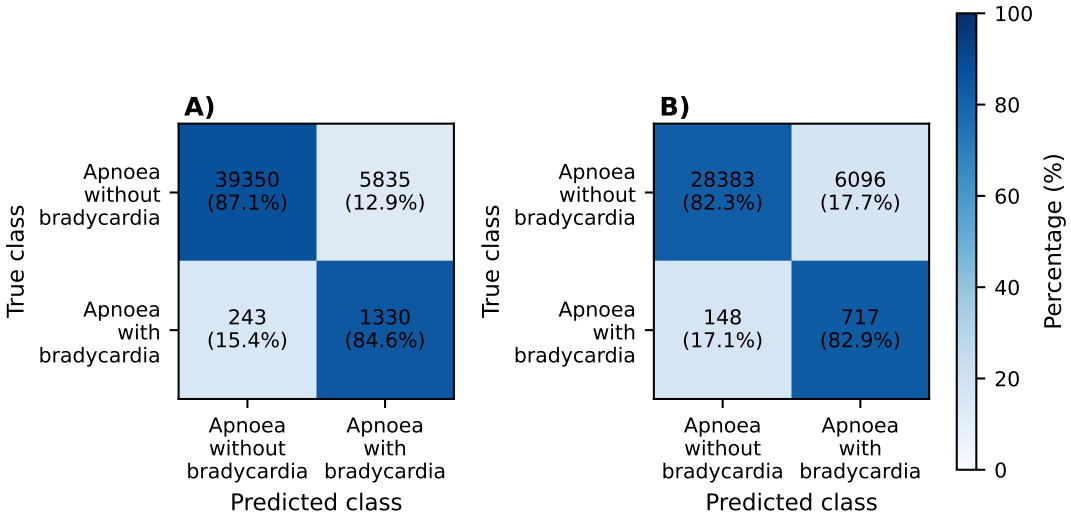

Figure 4.5: Confusion matrices of the XGBoost model predicting apnoea-related bradycardia events. Column A shows results from the optimised model trained using all features, whereas column B shows results from the optimised model trained using demographic and clinical features only. Numbers in each cell indicate the count of apnoea events in each category.

Predicting apnoea-related desaturation events

Similarly, the best-performing models for predicting apnoea-related desaturation events correctly identified 63.0% of apnoeas associated with desaturation and 86.5% of those not associated with desaturation when all features were included (Figure 4.6A). When restricted to demographic and clinical features, the model performance dropped, and it was only able to correctly identified 58.8% of apnoeas associated with desaturation and 55.0% of those not as-

sociated with desaturation (Figure 4.6B). Both models still performed significantly better than chance (permutation test, $p < 0.0001$).

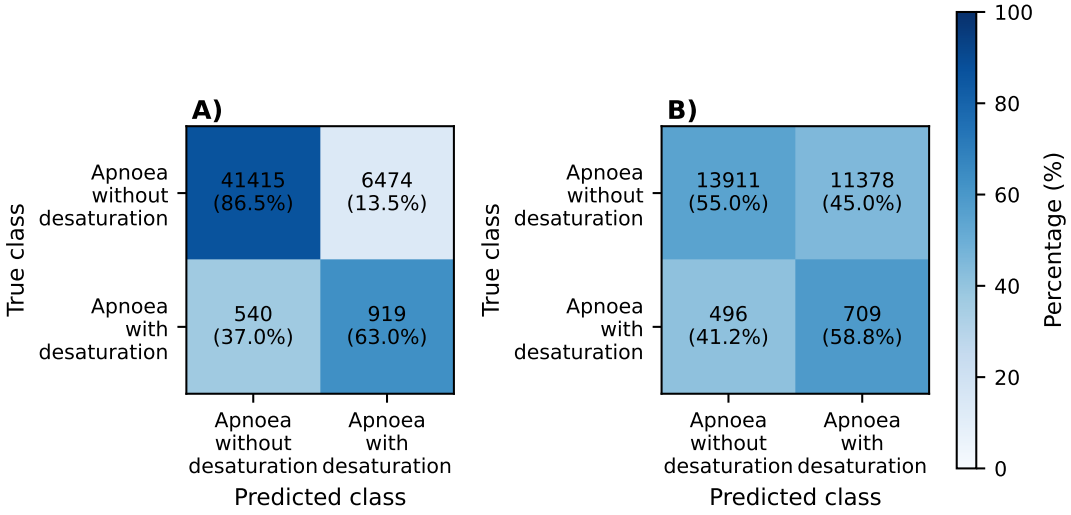

Figure 4.6: Confusion matrices of the XGBoost model predicting apnoea-related desaturation events. Column A shows results from the optimised model trained using all features, whereas column B shows results from the optimised model trained using demographic and clinical features only. Numbers in each cell indicate the count of apnoea events in each category.

Predicting apnoea-related bradycardia and/or desaturation events

The optimised models for predicting apnoea-related bradycardia and/or desaturation events performed significantly better than chance (permutation test, $p < 0.0001$), both when including only demographic and clinical features and when incorporating all features. The model including all features correctly identified 75.2% of apnoea events associated with bradycardia/desaturation and 76.4% of apnoea events not associated with bradycardia/desaturation (Figure 4.7A). The AUROC on the held-out test set was 0.83.

The model restricted to demographic and clinical features also performed significantly better than chance (permutation test, $p < 0.0001$), correctly identifying 66% of apnoea events associ-

ated with bradycardia/desaturation and 66% of those not associated with bradycardia/desaturation (Figure 4.7B).

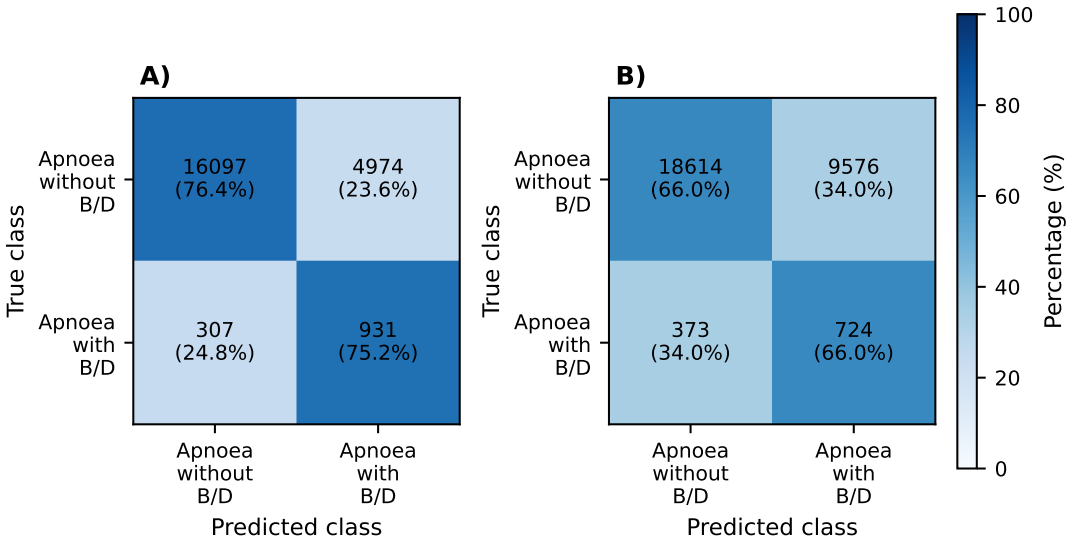

Figure 4.7: Confusion matrices of the XGBoost model predicting apnoea-related bradycardia and/or desaturation events. Column A shows results from the optimised model trained using all features, whereas column B shows results from the optimised model trained using demographic and clinical features only. Numbers in each cell indicate the count of apnoea events in each category. B/D denotes bradycardia and/or desaturation.

Risk assessment using synthetic dataset

To identify the combined clinical factors associated with high and low risk of cardiorespiratory instability during apnoea, Model 3 (predicting both bradycardia and/or desaturation) incorporating only demographic features was used to estimate the risk of apnoea-related bradycardia and/or desaturation across a synthetic dataset. Synthetic profiles in the highest and lowest 10% of predicted risk for apnoea associated with bradycardia and/or desaturation at durations between 5 and 10 seconds differed significantly in GA at birth, PMA at the time of the study, sex, Apgar score 5 minutes after the birth, and ventilation mode (Table 4.6). Males, younger gesta-

tional age, younger postmenstrual age, lower Apgar scores, or receiving high-flow respiratory support are associated with a higher risk of experiencing short apnoea episodes associated with bradycardia and/or desaturation. Pairwise heatmaps of these significant features were generated to illustrate how predicted risk varies across these variables (Figure 4.8).

Term	**High-risk mean**	**Low-risk mean**	**p value**	**Significance**
PMA at study	32.81	34.39	<0.0001	***
GA at birth	27.59	31.72	<0.0001	***
Weight z-score	-0.61	-0.66	0.65	
Apgar (5 min)	3.11	6.19	<0.0001	***
Ventilation==SVIA	0.35	0.75	<0.0001	***
Ventilation==Low Flow	0.10	0.04	0.00037	***
Ventilation==High Flow	0.55	0.21	<0.0001	***
Sex==female	0.45	0.57	0.0010	***
Sex==male	0.55	0.43	0.0010	***

Table 4.6: Comparison of model features between apnoeas with the highest and lowest 10% predicted risk of bradycardia and/or desaturation. Features were compared between high- and low-risk groups based on model-predicted risk for all apnoeas with duration between 5 and 10 seconds. Statistically significant differences are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Feature comparisons between high- and low-risk groups for longer apnoea durations are presented in Tables 4.7 - 4.8 and Figures 4.9 - 4.10.

Personalised apnoea alarm thresholds

Building on the classification model described above, a personalised predictive model was developed to estimate a suitable apnoea alarm threshold for individual infants. By inputting patient-specific characteristics into the model, the optimised XGBoost classifier is applied across a range of apnoea durations. The shortest duration at which a bradycardia or desat-

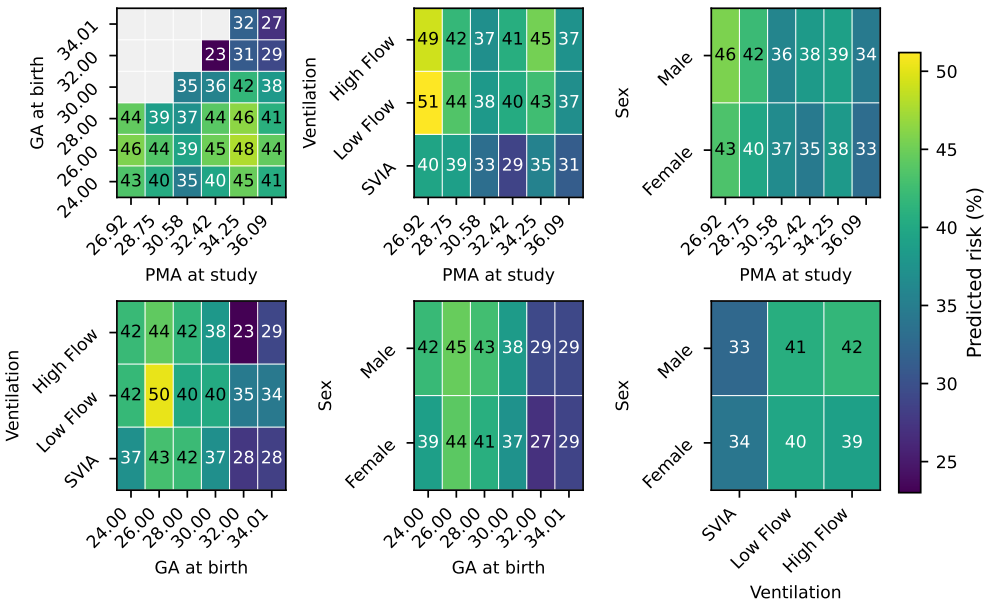

Figure 4.8: Pairwise heatmaps of the predicted risk of apnoea-related bradycardia and/or desaturation. Risk was predicted for all apnoeas with durations between 5 and 10 seconds using the optimised XGBoost model. Values shown in each heatmap cell (colour-coded and numerically annotated) represent the mean predicted risk for the corresponding feature combinations.

Term	High-risk mean	Low-risk mean	p value	Significance
PMA at study	33.58	34.31	<0.001	***
GA at birth	27.65	32.15	<0.001	***
Weight z-score	-0.25	-0.27	0.790	
Apgar 1 min	5.69	5.18	<0.001	***
Ventilation==SVIA	0.15	0.90	<0.001	***
Ventilation==Low Flow	0.18	0.03	<0.001	***
Ventilation==High Flow	0.67	0.07	<0.001	***
Sex==female	0.38	0.58	<0.001	***
Sex==male	0.62	0.42	<0.001	***

Table 4.7: Comparison of model features between apnoeas with the highest and lowest 10% predicted risk of bradycardia and/or desaturation. Features were compared between high- and low-risk groups based on model-predicted risk for all apnoeas with duration between 10 and 20 seconds. Statistically significant differences are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Term	High-risk mean	Low-risk mean	p value	Significance
PMA at study	32.41	34.78	<0.001	***
GA at birth	27.39	32.45	<0.001	***
Weight z-score	-0.48	-0.18	<0.001	***
Apgar 1 min	6.31	4.68	<0.001	***
Ventilation==SVIA	0.06	0.96	<0.001	***
Ventilation==Low Flow	0.13	0.02	<0.001	***
Ventilation==High Flow	0.81	0.02	<0.001	***
Sex==female	0.39	0.52	<0.001	***
Sex==male	0.61	0.48	<0.001	***

Table 4.8: Comparison of model features between apnoeas with the highest and lowest 10% predicted risk of bradycardia and/or desaturation. Features were compared between high- and low-risk groups based on model-predicted risk for all apnoeas with durations longer than 20 seconds. Statistically significant differences are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

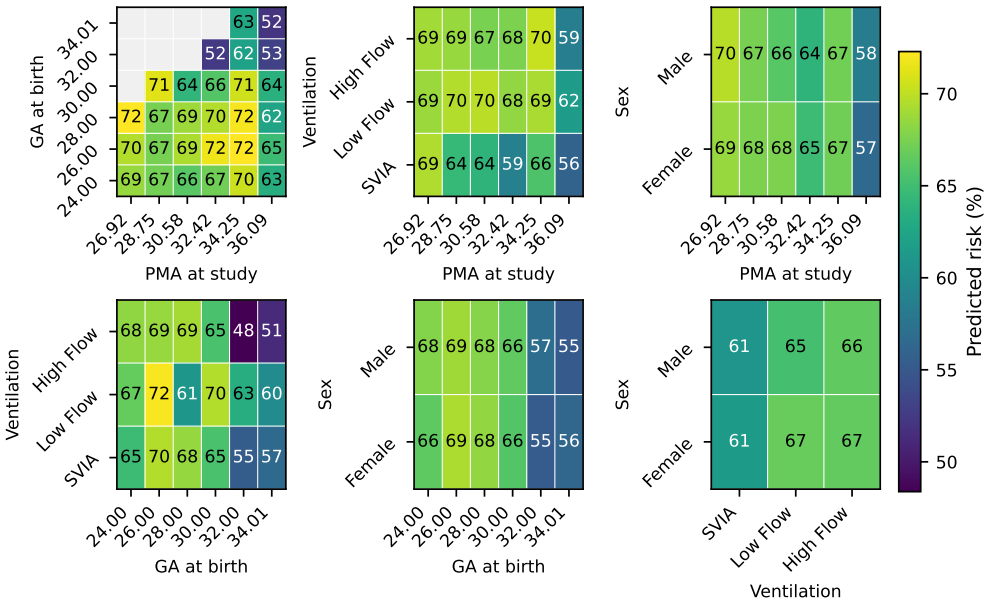

Figure 4.9: Pairwise heatmaps of the predicted risk of apnoea-related bradycardia and/or desaturation. Risk was predicted for all apnoeas with durations between 10 and 20 seconds using the optimised XGBoost model. Values shown in each heatmap cell (colour-coded and numerically annotated) represent the mean predicted risk for the corresponding feature combinations.

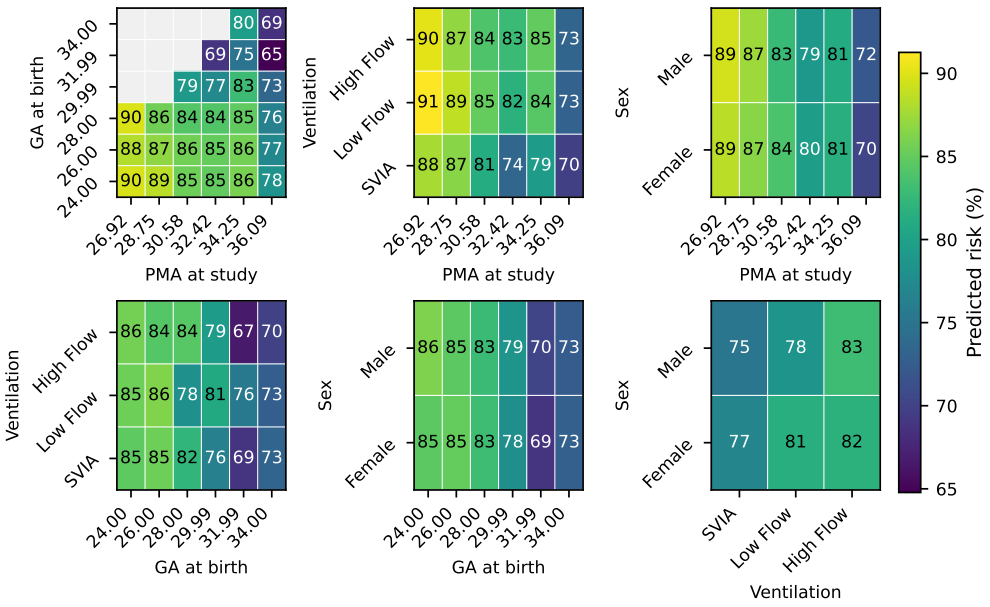

Figure 4.10: Pairwise heatmaps of the predicted risk of apnoea-related bradycardia and/or desaturation. Risk was predicted for all apnoeas with durations longer than 20 seconds using the optimised XGBoost model. Values shown in each heatmap cell (colour-coded and numerically annotated) represent the mean predicted risk for the corresponding feature combinations.

uration becomes likely is then recommended as the personalised apnoea alarm limit for that infant. To demonstrate the potential clinical utility of this approach, a prototype web application was created. This interface allows users to input an infant's demographic and clinical details and receive a suggested apnoea alarm threshold. An example usage of the application is presented in Figure 4.11, where the recommended apnoea threshold for the illustrated infant is 15 seconds.

4.4 Discussion

In this chapter, over 27,000 hours of vital signs recording data were analysed to investigate factors underlying cardiorespiratory instability during apnoea in preterm infants aged 26-36

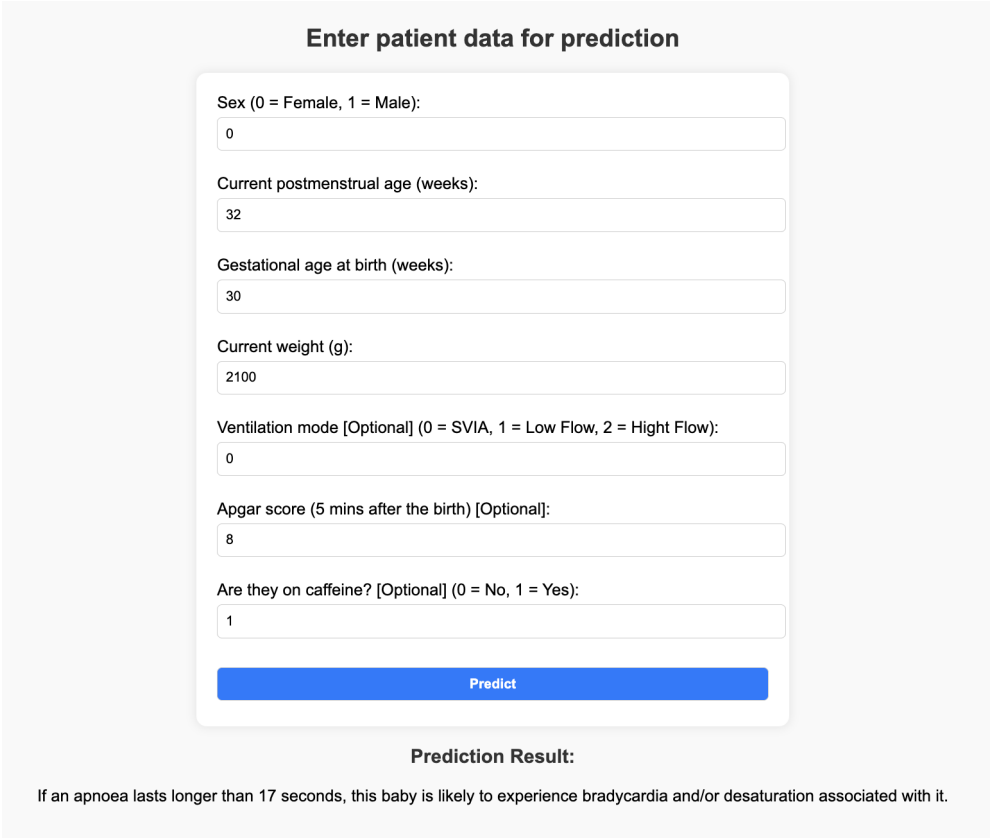

Figure 4.11: Example usage of the personalised apnoea alarm tool.

weeks' PMA. The findings of the systematic review presented in Chapter 3 were successfully reproduced, confirming that changes in heart rate and oxygen saturation are both associated with apnoea duration, but with substantial variability. The systematic review highlighted that relatively little research has explored factors modulating the relationship between apnoea duration and cardiorespiratory instability. In addition, many of the included studies were conducted before 2000, during periods when treatment regimens and respiratory support practices differed considerably from current standards. This chapter addresses these limitations by providing a rigorous evaluation of potential modulating factors within a contemporary cohort. The findings offer new insight into why some infants are at greater risk of cardiorespiratory instability during apnoea, while others may demonstrate relative resilience.

I identified multiple demographic, clinical and dynamic factors which influence the risk of bradycardia and/or desaturation. Moreover, apnoeas with bradycardia and/or desaturation could be predicted using machine learning, either predicting one of the two events or both. Whilst the models were more accurate when it included dynamic features, such as baseline heart rate and oxygen saturation, a model including clinical and demographic features alone could still predict risk of cardiorespiratory instability at a sufficient accuracy, except when predicting desaturations alone. A tool which can accurately predict an infant's risk of bradycardia/desaturation associated with apnoea duration could pave the way for development of individualised apnoea definitions and personalised alarm limits.

Although the predictive models achieved sensitivity and specificity significantly above chance, it remains unclear whether their performance would be significantly better than a simpler model using apnoea duration alone. Given the strong association between apnoea duration and car-

diorespiratory instability, duration itself may account for a substantial proportion of the predictive performance. Future studies should therefore compare the proposed models against an apnoea duration-only baseline to determine the additional predictive value of incorporating demographic, clinical and dynamic physiological features.

In current clinical practice, apnoea alarms are commonly defined using uniform thresholds across all infants and breathing pauses shorter than 10 seconds are often not classified as apnoea. However, our findings suggest that a subset of infants may experience severe cardiorespiratory instability following even relatively brief respiratory pauses. The instability reaches a plateau for apnoeas longer than approximately 20 seconds (Figure 2), which may be because clinical interventions prevent further declines in heart rate and/or oxygen saturation. These findings highlight the potential value of a personalised definition of apnoea incorporating demographic characteristics and dynamic factors, such as recent apnoea history.

A major contribution of this chapter is the development of a predictive framework capable of estimating an infant-specific apnoea duration threshold associated with a high probability of bradycardia and/or desaturation. By integrating demographic, clinical, and dynamic physiological information, the model moves beyond conventional fixed-duration definitions of apnoea and instead estimates each infant's physiological tolerance to respiratory pauses. To demonstrate its potential clinical application, a prototype web application was developed that enables clinicians to enter an infant's characteristics and obtain a personalised apnoea duration threshold.

Before such a framework could be implemented clinically, several stages of validation would be

required. First, the model should be externally validated using independent cohorts from multiple neonatal centres to assess its robustness across different patient populations, monitoring systems, and clinical practices. Second, prospective observational studies should determine whether personalised thresholds remain stable within individual infants over time and accurately identify clinically significant events. Finally, interventional studies would be required to establish whether incorporating personalised alarm thresholds into routine monitoring improves clinical outcomes, reduces alarm fatigue, and supports earlier intervention without increasing the risk of missed clinically important apnoeas. If successful, this could enable earlier detection of clinically significant apnoeas in more vulnerable infants, which could trigger earlier clinical assessment and management, and potentially help to mitigate the risk of long-term impacts of apnoea-associated hypoxaemia [25]. Understanding the risks of cardiorespiratory instability from apnoea in different infants could also be used when planning procedures where infants may be more likely to experience apnoea [171]. Whilst the personalised apnoea alarm model was not externally validated, this study nevertheless provides the foundations for such a tool.

Consistent with Chapter 3, PMA at the time of study was found to significantly influence the risk of cardiorespiratory instability in this study. In addition, gestational age (GA) at birth was also a significant factor, with younger infants at higher risk. The respiratory system and the body's oxygen carrying capacity matures with increasing PMA. With advancing PMA, functional residual capacity increases, surfactant production is enhanced, respiratory muscles become stronger and alveoli development is more mature. Moreover, the shift from predominantly foetal haemoglobin to adult haemoglobin may result in increased oxygen delivery to

tissues. These changes likely means that infants with older PMA are more resilient to longer duration apnoeas.

The age at which apnoeas stop is related to GA at birth, with infants born at younger gestational age experiencing apnoea until older PMA [41]. Our data shows that when infants of younger GA at birth experience apnoeas, they are more likely to be associated with bradycardia and desaturation, and that GA at birth also modulates the relationship between apnoea duration and cardiorespiratory instability. This could in part be related to relative immaturity of lung function, which is known to persist after the initial neonatal period and could impact on cardiorespiratory response to apnoea, compared to their higher GA at birth counterparts. Infants born at younger GA are also at increased risk of numerous pathologies, including anaemia of prematurity, intraventricular haemorrhage, infections, metabolic disturbances and temperature instability, all of which can both be pathological causes of apnoea but could also reduce infants' ability to tolerate apnoeas. Moreover, infants born at a lower GA are likely to require prolonged mechanical respiratory support and may subsequently develop bronchopulmonary dysplasia (BPD), which can impair pulmonary gas exchange and could increase extremely preterm infants' vulnerability to apnoea-associated desaturation.

Sex was also identified as a significant factor, with male infants showing a higher overall likelihood of apnoea-related bradycardia/desaturation. However, the stratified analysis suggested that this relationship may vary with apnoea duration. Although male infants generally showed greater instability for shorter apnoeas, female infants appeared to have a higher proportion of events associated with bradycardia/desaturation for apnoeas longer than approximately 25 seconds. This suggests that sex-related differences in physiological responses to apnoea may

depend on event duration, although the underlying mechanisms remain unclear. This observation should be interpreted cautiously, particularly given the potentially smaller number of prolonged apnoeic events.

Our data show that infants who experience clusters of apnoeas are more vulnerable to cardiorespiratory instability with a subsequent apnoea. Clinically, short recurrent apnoeas can be seen in periodic breathing, defined as at least 3 pauses in breathing (lasting $\geq$ 3 seconds each), separated by less than 20 seconds each [172]. They are relatively common, particularly in preterm infants, and are generally considered benign and self-resolving [173]. However, previous studies have shown that clusters of brief apnoeas as short as 3 seconds may cause significant bradycardia and desaturation, and even changes in cerebral oxygenation and blood volume [128, 174–176]. Our data adds to this work and suggests that having multiple (even brief) apnoeas within a 5-minute period should alert clinicians to the increased chance of a subsequent desaturation/bradycardia. Previous studies have also found associations between multiple brief apnoeas (e.g. in sequential apnoeas and periodic breathing) and poorer neurodevelopmental outcomes at 6 months and 2 years of life, reinforcing the potential clinical seriousness of these seemingly insignificant events [9, 174].

It has also been found that mode of respiratory support (SVIA, low flow and high flow oxygen) is associated with the likelihood of apnoea-associated bradycardia/desaturation. The mode of respiratory support an infant receives may be influenced by a number of factors, including the infant's overall clinical status and the presence of respiratory or other comorbidities (e.g. respiratory distress syndrome (RDS), BPD, pneumonia), some of which may also impair an infant's ability to tolerate apnoeas without cardiorespiratory compromise. The choice of respiratory

support is also often driven by local clinical guidelines as well as clinician judgement, and as such this result needs to be considered cautiously in interpretation across different neonatal units.

Interestingly, the relationship between apnoea duration and the likelihood of bradycardia/desaturation appeared to plateau or diverge between subgroups after approximately 20 seconds. One possible explanation is the influence of clinical interventions, as apnoea alarms are commonly triggered after 20 seconds, potentially prompting interventions that prevent further cardiorespiratory deterioration. The likelihood and nature of these interventions may also vary depending on an infant's perceived clinical stability. Infants considered more vulnerable may receive earlier or more intensive intervention, whereas those considered relatively stable may receive less immediate intervention, potentially contributing to some of the counterintuitive patterns observed. For example, although male infants had a higher overall likelihood of bradycardia/desaturation, female infants appeared to have a higher proportion of events associated with instability during longer apnoeas. Similarly, infants receiving low flow respiratory support appeared to have a higher likelihood of instability during prolonged apnoeas than those receiving high flow support, despite the latter potentially representing a more clinically vulnerable population. This may partly reflect differences in the respiratory support provided and clinical responses to apnoea. However, as the timing and nature of clinical interventions were not available, these explanations remain speculative. These findings further highlight the potential limitations of uniform apnoea alarm thresholds and support the investigation of personalised monitoring strategies that account for individual physiological vulnerability.

This study has several other limitations. The dataset included only central and mixed apnoeas

and did not capture obstructive apnoeas as nasal air flow is not routinely monitored on our neonatal unit. Consequently, analyses stratified by apnoea type were not possible. Chapter 3 identified literature that showed apnoea type significantly modulated how infants respond to an apnoea. This result was not able to be reproduced in this chapter.

Moreover, the majority of infants were older than 28 weeks PMA at the time of recording; evaluating cardiorespiratory instability risk in the youngest infants should be considered in a separate study. Conversely, as our hospital is a regional centre, preterm infants who are clinically stable are more likely to be discharged or transferred to other hospitals at an earlier PMA. As a result, infants with recordings available at late preterm ages in this dataset are more likely to represent a clinically unstable or complex subgroup, introducing potential bias into analyses involving PMA.

The algorithms used to obtain inter-breath intervals from the IP signal have a false-positive rate of around 10%, meaning that a similar proportion of apnoea events detected in this study may not have been true events. Manual screening and machine learning algorithms have been combined to avoid some false positives, but this limitation is likely to prevent the predictive models from achieving higher accuracy. Nevertheless, the models still achieved promising performance.

Another limitation is that apnoea events were required to occur from a stable cardiorespiratory baseline to allow reliable assessment of the associated changes in heart rate and oxygen saturation. Consequently, a subsequent apnoea occurring within 90 seconds of a preceding event associated with bradycardia and/or desaturation would be excluded if physiological re-

covery had not occurred. This may have resulted in an underrepresentation of closely spaced apnoeas, particularly those occurring before recovery from a preceding episode of cardiorespiratory instability, potentially influencing the observed relationship between apnoea clustering and subsequent physiological responses.

I was also unable to extract precise information from the clinical database on whether infants were receiving caffeine treatment at the time of each apnoea event, or the exact dosage administered, as medication records did not reliably capture the timing of administration. Similarly, infection status at the time of apnoea events could not be determined accurately due to limitations in the available data. In addition, infants who were infected and clinically unstable may be less likely to be enrolled in research studies, potentially biasing the dataset towards non-infected infants. Therefore, infection status was not included as a feature in the model, and infants receiving antibiotics at the time of apnoea episodes (who were likely to be infected) were excluded to minimise confounding effects.

Finally, the set of potential modulating factors within this data set is not exhaustive. For example, the impact of anaemia, steroids, surfactant, other medication, concurrent respiratory pathologies (e.g. BPD, RDS) and cardiac pathologies (e.g. PDA) would also be important to consider and should be addressed in future predictive models to obtain a tool suitable for all infants. Despite these limitations, this study provides strong evidence that infant demographic characteristics, recent apnoea history, and baseline physiological state significantly influence the cardiorespiratory response to apnoea. The severity of apnoea-related instability varies considerably between individuals, and for some infants, respiratory pauses shorter than 10 seconds may already be related to a bradycardia or desaturation, suggesting that the definition of apnoea

may warrant further consideration.

In summary, I demonstrate that multiple factors modulate the risk of bradycardia and/or desaturation from apnoea. Infants born at younger gestational ages, current younger PMA, and receiving high flow therapy were at higher risk than their older counterparts who were self-ventilating or receiving low flow therapy. Moreover, apnoea with bradycardia and/or desaturation could be accurately predicted from clinical, demographic and dynamic features suggesting the potential for future development of clinical tools for personalised apnoea alarm limits. Given the long-term neurodevelopmental impact of frequent hypoxaemia, understanding infants at greater risk of cardiorespiratory instability is key to improve outcomes for preterm infants.

Chapter 5

Assessing preterm infants' functional respiratory age

5.1 Introduction

Biological age models have previously been used to assess human maturation and ageing trajectories [77–82]. In infants, cardiac and brain signals have been shown to capture developmental maturity [87, 177]. Preterm infants are born with an immature respiratory system and undergo rapid respiratory development during the postnatal period [26, 27]. The findings from the previous chapters suggest that respiratory maturation is not fully captured by chronological age alone: both apnoea frequency and the physiological responses to apnoeas vary substantially within and between infants, indicating that respiratory development follows heterogeneous trajectories.

This chapter therefore aims to develop models for estimating functional respiratory age (FRA) in preterm infants. Chapter 1 introduced the developmental immaturity of the preterm respiratory system and the concept of physiology-based biological age models. Respiratory control and pulmonary function were presented as dynamic systems that mature substantially after preterm birth, suggesting that routinely collected respiratory signals may contain information about an infant's underlying respiratory maturity. A reliable FRA model would therefore pro-

vide a non-invasive and low-burden means of assessing respiratory maturity using signals already available in clinical care. Such a model could help identify infants whose respiratory development appears delayed relative to their PMA and support closer monitoring or timely intervention when appropriate.

The aim of this chapter is to develop a reference trajectory that reflects the expected pattern of respiratory development among clinically stable preterm infants. The functional respiratory age should be interpreted relative to this preterm reference population rather than as an absolute measure of respiratory maturity. The purpose of the model is to identify infants whose respiratory physiology appears more or less mature than expected for clinically stable preterm infants of the same PMA. As there is currently no established objective reference measure of FRA, this study uses PMA as a surrogate for expected respiratory maturity. The central assumption is that, among clinically stable preterm infants without significant respiratory pathology, respiratory maturation follows a sufficiently consistent developmental trajectory for PMA to provide a reasonable reference. Consequently, the FRA estimated by the model represents the age at which an infant's respiratory characteristics would typically be observed within the reference population, rather than a direct measurement of their underlying physiological respiratory maturity. This follows the same general rationale as other biological age models, such as brain age and heart age, in which physiological measurements are used to estimate biological rather than chronological age [74–76, 87]. These approaches are based on the premise that organ systems undergo characteristic maturational or ageing trajectories that are reflected in measurable physiological signals [178, 179]. By learning the relationship between physiological features and chronological age in a reference population, the resulting model provides an estimate of

the age most consistent with an individual's observed physiology. Deviations between estimated biological age and chronological age may therefore indicate that the organ system has developed or aged differently from that expected for a typical individual of the same age [85, 180]. The proposed functional respiratory age model adopts the same conceptual framework, with the important distinction that it characterises maturation of the respiratory system during early postnatal development rather than ageing later in life.

In addition to developmental maturation, respiratory function in preterm infants may also be influenced by clinical exposures. Morphine is commonly used in neonatal care for analgesia and sedation, but it may alter breathing patterns [108]. This makes morphine a relevant clinical factor to consider when interpreting respiratory features and model-derived estimates of respiratory maturity.

The specific objectives of this chapter are to:

1. Investigate the correlation between respiratory features and PMA in clinically stable preterm infants to establish the expected trajectory of respiratory maturation.
2. Develop and validate a model for estimating functional respiratory age (with PMA as the ground truth) from respiratory features relative to this reference trajectory.
3. Investigate how morphine affects individual respiratory features and/or FRA.

5.2 Methods

5.2.1 Study design and modelling framework

This chapter used Dataset 2 (NeoVital dataset), described in Chapter 2.2, to address the first two objectives. The aim was to derive a FRA from respiratory features and evaluate its relationship with PMA. The modelling framework comprised both regression and classification approaches, implemented using strict subject-level cross-validation to prevent data leakage. Model performance was assessed at the recording level and, where appropriate, at the window level with subsequent aggregation. To address the third objective, Dataset 3 (POPPI dataset), described in Chapter 2.2, was used. This dataset enabled comparisons between infants who received morphine and those who received placebo.

The medication-based exclusion criteria for Dataset 2 were intended to minimise the influence of pharmacological treatments on respiratory physiology when establishing the reference maturational trajectory. Certain medications may alter respiratory drive, breathing patterns and respiratory variability independently of developmental maturity. Including recordings obtained during such treatment could therefore introduce physiological changes unrelated to PMA, potentially confounding the relationship between respiratory features and respiratory maturation. Excluding infants receiving these medications was intended to provide a more consistent reference population in which age-related respiratory changes could be investigated with fewer pharmacological influences.

The overall analytical strategy involved four stages. First, respiratory features were exam-

ined individually to determine whether they had systematic relationships with PMA and access whether respiratory features contains measurable maturational information. Second, multivariable machine learning models were developed to estimate PMA from combinations of respiratory features, allowing assessment of whether multiple physiological characteristics together provide a more accurate representation of respiratory maturity than any single feature. Third, feature selection and dimensionality reduction approaches were evaluated to determine whether predictive performance could be improved by reducing redundancy within the respiratory feature set. Finally, the best-performing model was applied to an independent morphine dataset to investigate whether morphine admission resulted in measurable changes in functional respiratory age.

5.2.2 Data processing

IBIs were derived from the IP signal using the breath detection and processing methods described in the general signal processing methods section (Chapter 2). All subsequent feature extraction was performed using the IP signal and the corresponding IBI series.

For each eligible recording, a set of 66 respiratory features was computed. Feature extraction was performed independently for each recording. A subset of 50 features was derived directly from the IBI time series. These include basic descriptive statistics, such as the mean, median, standard deviation, interquartile range and kurtosis of IBI, to capture the central tendency and dispersion. In addition, measures of signal complexity and temporal structure were computed from the IBI series. These included entropy-based metrics and frequency-domain features originally developed for cardiac interval analysis. These entropy and spectral features have

previously been applied to compute functional autonomic age by Iyer et al. (2021) [177]. The same definitions and parameter settings were applied here to respiratory IBIs.

Further features were extracted directly from the IP signal to characterise breath morphology and respiratory dynamics. These features included metrics related to breath amplitude, timing, and variability, as defined in prior work on impedance pneumography–derived respiratory measures. Feature definitions and computational procedures followed those described by Noto et al. (2018) [181].

Table B.1 in the Appendix shows the list of all the included features and a brief explanation of each feature.

5.2.3 Correlation analysis

Row-level Pearson correlation coefficients were computed between each respiratory feature and PMA. The association between each respiratory feature and PMA was assessed using linear mixed-effects models to account for repeated measurements within infants. For each feature, the following model was fitted:

$$\mathrm{PMA}_{ij} = \beta_0 + \beta_1 \cdot \mathrm{Z}(\mathrm{feature}_{ij}) + \mathrm{u}_i + \varepsilon_{ij} \tag{5.1}$$

where:

- i indexes infants and j indexes recordings,
- Z(feature) denotes the standardized (z-scored) respiratory feature,

- u_i is a random intercept for infant i,
- ε_{ij} is the residual error.

Feature standardisation allowed direct comparison of effect sizes across metrics. Model coefficients represent the expected change in PMA in weeks per one standard deviation increase in the feature. Statistical significance of fixed effects was assessed using Wald tests [182]. To control for multiple testing across the full set of respiratory features, p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure [183]. Features with FDR-adjusted p-values < 0.05 were considered statistically significant.

5.2.4 Cross-validation strategy

All machine learning models were developed and evaluated using a nested cross-validation framework with strict infant-level grouping to prevent data leakage and obtain an unbiased estimate of model performance [94, 184, 185]. Nested cross-validation separates model selection from model evaluation, which reduced the risk of overly optimistic performance estimates that may arise when the same data are used for both purposes.

The outer cross-validation loop was used to evaluate model performance. A leave-one-infant-out (LOIO) strategy was adopted [186, 187]. All recordings from one infant were held out as an independent test set, while recordings from all remaining infants were used for model development. This process was repeated until every infant had served once as the test subject. It was also ensured that all recordings from an individual infant were assigned exclusively to either the training or test set. In this way, the model was always evaluated on completely unseen

infants.

Within each outer training set, an inner cross-validation procedure was used to optimise model hyperparameters. Hyperparameters are settings that control how a machine learning algorithm learns from the data (for example, the number of trees in a tree-based model or the maximum tree depth) and must be specified before model training. To select these settings, grouped five-fold cross-validation was performed, ensuring that all recordings from the same infant remained within the same fold. Candidate hyperparameter combinations were sampled using random search [188], and their performance was evaluated across the five folds. The combination yielding the best average performance was selected and used to train the final model for that outer iteration. Importantly, the held-out test infant was not used during hyperparameter optimisation, ensuring that the final performance estimates reflected evaluation on completely unseen data.

5.2.5 Prediction of PMA

Regression models

To evaluate whether respiratory features reflect maturational status in clinically stable preterm infants, supervised regression models were trained to predict PMA from 66 extracted respiratory features. If respiratory physiology follows a characteristic developmental trajectory, the respiratory features should collectively contain sufficient information to estimate the PMA at which they would typically be observed. Under this framework, the predicted PMA is interpreted as an estimate of FRA relative to the reference population. As PMA was used as a

surrogate for expected respiratory maturity, agreement between predicted and observed PMA was used to evaluate model performance, but does not independently validate the estimated FRA as a measure of physiological respiratory maturity.

Six regression algorithms were considered: Random Forest [104, 189], Gradient Boosting [190], Extreme Gradient Boosting (XGBoost) [166], Support Vector Regression (SVR) [191], Elastic Net [192], and K-Nearest Neighbours (KNN) [193]. These algorithms differ in their assumptions regarding the relationship between predictors and the outcome, allowing assessment of whether respiratory maturation is better characterised by linear or nonlinear models. Ensemble tree-based methods (Random Forest, Gradient Boosting and XGBoost) can capture complex nonlinear relationships and interactions among predictors while being robust to correlated features. SVR uses kernel methods to model nonlinear relationships in high-dimensional feature spaces. Elastic Net provides a regularised linear model that performs variable selection while reducing overfitting, whereas KNN estimates PMA from the most similar recordings in the feature space without assuming a predefined functional relationship. Together, these methods provided a broad range of modelling strategies and a comprehensive assessment of the predictive information contained within the respiratory features.

A nested cross-validation approach was used as described in Section 5.2.4. For the inner cross-validation, K was set to five, and a randomised search with 200 iterations was performed for each model. The model achieving the lowest mean absolute error (MAE) within the inner cross-validation was selected and subsequently evaluated on the held-out infant in the outer loop.

Model performance was quantified using MAE, root mean squared error (RMSE), coefficient of determination (R^2), Pearson correlation coefficient (r) to the actual PMA, and the slope of the fitted linear relationship between predicted and actual PMA. Pearson's r quantifies the strength of the association between predicted and observed PMA, whereas the slope assesses calibration, with a slope of one indicating proportional agreement across the full PMA range.

Machine learning regression models may exhibit systematic calibration bias, with predictions being consistently compressed towards the mean age. Such bias can arise from regularisation or model averaging, and may reduce agreement between predicted and observed PMA even when the underlying maturational information is well captured. A post hoc linear calibration step was therefore evaluated to determine whether systematic bias could be corrected without altering the fitted machine learning model [194–196]. For each outer training fold, predictions generated by the selected model on the training data were regressed against the corresponding true PMA values using a simple linear model:

$$\mathrm{PMA}_{\mathrm{true}} = \mathrm{a} + \mathrm{b} \cdot \mathrm{PMA}_{\mathrm{predicted}}.$$

The estimated calibration parameters (a, b) were then applied to the predictions of the held-out test infant to obtain calibrated predictions. Performance metrics were computed both before and after calibration to assess whether linear recalibration improved agreement between predicted and actual PMA.

This calibration assumes that any systematic prediction bias can be approximated by a lin-

ear relationship between predicted and observed PMA, such that errors are primarily due to a constant offset and/or proportional scaling. The calibration parameters were estimated exclusively from the training data within each outer cross-validation fold and subsequently applied unchanged to the corresponding held-out test infant. Consequently, the test data remained completely independent throughout model fitting and calibration, preventing information leakage.

Comparing model performance before and after calibration allowed assessment of whether prediction errors were arised predominantly from systematic calibration bias or from limitations in the respiratory features themselves. Substantial improvement after calibration would suggest that the model had successfully captured maturational information but produced biased age estimates, whereas minimal improvement would indicate that prediction errors were primarily due to limited predictive information within the respiratory features.

Each recording was also divided into windows of one-hour duration, and features were generated using the one-hour windows. The same regression modelling approaches were applied to this dataset, with predictions first generated at the window level. For each recording, the final predicted PMA was defined as the median prediction across all windows belonging to that recording. For each recording, the variance of window-level predictions was also computed to quantify intra-recording prediction variability.

Finally, the model and procedure that performed best in initial experiments were examined using a more comprehensive hyperparameter search. Out-of-fold (OOF) performance metrics were obtained using LOIO cross-validation. The final hyperparameter set was selected as the most frequently occurring parameter configuration across outer folds. A new pipeline was then

developed with these chosen parameters and retrained on the entire dataset. This final model was saved for morphine effect analysis and for future evaluation on independent test sets. To avoid optimistic bias, no performance metrics were computed on the training set based on this model.

Classification models

To evaluate whether respiratory features capture clinically meaningful differences in maturity, classification models were trained to distinguish the degree of prematurity. Preterm infants are commonly categorised as extremely preterm (<28 weeks PMA), very preterm (28 to <32 weeks PMA), moderate preterm (32 to <34 weeks PMA), and late preterm (34 to <37 weeks PMA). However, the dataset used in this study contained only 7 recordings obtained when infants were extremely preterm. Therefore, recordings with PMA <28 weeks were combined with the very preterm group, and classification models were trained to distinguish three classes: 0 = <32 weeks PMA, 1 = 32 to <34 weeks PMA, and 2 = 34 to <37 weeks PMA.

Six classifiers (Random Forest, Gradient Boosting, XGBoost, Support Vector Classifier, Logistic Regression, and KNN) were evaluated. Class imbalance was addressed using class weighting [197, 198]. Model performance was assessed using balanced accuracy and the area under the receiver operating characteristic curve (AUROC).

Permutation test

The performance of all regression and classification models was evaluated using permutation tests. The labels in the dataset were randomly permuted 10,000 times, and models trained on the permuted datasets were compared with the original models. The p-value was calculated as the proportion of permuted models that performed better than the original model. For regression models, performance was assessed using MAE, while for classification models, performance was assessed using balanced accuracy.

5.2.6 Feature selection

To examine whether reducing the dimensionality of the respiratory feature set could improve model stability and generalisability, two feature selection strategies were evaluated in addition to a model using all available features. These approaches comprised a supervised univariate selection method and an unsupervised dimensionality reduction method. Both were implemented within the cross-validation framework so that feature selection was performed exclusively on training data in each fold.

Supervised feature selection - Select K-Best

Supervised feature selection was performed using a univariate filtering approach, through the Select K-Best function (`SelectKBest`) in Python sklearn package. In this method, each respiratory feature was evaluated individually with respect to its statistical association with PMA

within the training data. Features were ranked according to the strength of this association, and only the highest-ranking features were retained for model training. Different numbers of top-ranked features were considered in the inner cross-validation procedure to identify the optimal subset size.

Unsupervised dimensionality reduction

In contrast, unsupervised dimensionality reduction was performed using principal component analysis (PCA). This method does not consider the outcome variable when constructing new variables. Instead, it identifies orthogonal linear combinations of the original respiratory features that explain the greatest variance within the training data. A reduced set of principal components was then used as input to the regression models.

By transforming correlated respiratory features into a smaller number of independent components, PCA reduces redundancy and multicollinearity within the feature set. Different numbers of retained components were evaluated within cross-validation to determine the degree of dimensionality reduction that best balanced information retention and model simplicity.

5.2.7 Assessing the effect of morphine on respiratory function

To investigate the effect of morphine on respiratory function, a separate dataset (Dataset 3, see Section 2.2) was analysed comprising infants who received either morphine or placebo. Each infant had approximately 24 hours of continuous respiratory recordings before and after administration of morphine or placebo.

Pre–post comparison in the morphine group

To assess the immediate effect of morphine on respiratory dynamics, analyses were first restricted to infants who received morphine. For each infant, respiratory features were extracted from recordings obtained prior to morphine administration and from recordings obtained following administration. Feature values before and after morphine were compared using paired t-tests to determine whether morphine exposure was associated with significant changes in respiratory function. Statistical significance was assessed at a threshold of FDR-adjusted $\mathrm{p} < 0.05$.

Time-resolved analysis using moving windows

To characterise temporal changes in respiratory features surrounding drug administration, recordings from both morphine and placebo groups were segmented into one-hour moving windows spanning from 24 hours before to 24 hours after administration. For each window, respiratory features were computed independently. If no valid signal was available within a given window, the corresponding feature values were recorded as missing.

For each time window, the mean and standard error (SE) of each respiratory feature were calculated separately for the morphine and placebo groups. Each data point was plotted at the midpoint of the corresponding one-hour window. This approach enabled visualisation of the temporal trajectory of respiratory dynamics across the peri-administration period and facilitated comparison between morphine and placebo groups.

Time-resolved functional respiratory age change estimation

To further evaluate the physiological impact of morphine, FRA was computed for each one-hour window using the best-performing regression model identified in earlier analyses. The trained model was applied to window-level feature sets to generate FRA estimates over time. Change in FRA was then defined as FRA minus the average FRA before morphine administration.

For each window, mean change in FRA and SE were calculated separately for morphine and placebo groups. Trajectories of change in FRA were plotted as a function of time relative to drug administration, with separate curves for morphine and placebo infants and shaded regions representing SE. This analysis allowed assessment of whether morphine administration was associated with deviations in respiratory maturity.

Cluster-based permutation testing was used to assess group differences in trajectories of FRA change over time while controlling for multiple comparisons across the sliding windows. For each window, an independent-samples t-statistic comparing groups (e.g., infants who had morphine versus infants who had placebo, or infants who had severe reactions to morphine versus infants who had placebo) was computed. The resulting time series of t-values was thresholded at a critical value corresponding to $\alpha = 0.05$, and the sliding windows were grouped into clusters. For each cluster, a cluster-level statistic was calculated as the sum of t-values within the cluster (cluster mass). To determine statistical significance, a null distribution of the maximum cluster mass was generated using 10,000 permutations, in which group labels were randomly reassigned across subjects while preserving the temporal structure of the data.

For each permutation, the same clustering procedure was applied and the largest cluster mass recorded. Observed cluster statistics were then compared against this null distribution, and cluster-level p-values were computed as the proportion of permutations with a maximum cluster mass greater than or equal to the observed value.

5.3 Results

The final dataset included 205 recordings from 105 preterm infants as described in Section 2.2. The median of infants' PMA at the time of the study was 32.9 weeks (ranged = 26.9 to 36.9 weeks). Each infant contributed 1 to 7 records. The total records of all infants amounted to 4,596.5 hours, with a median record time of 23.4 hours (ranging from 6.1 to 24.0 hours).

5.3.1 Relationship between respiratory features and PMA

Across all respiratory features analysed, 15 features were significantly associated with PMA after FDR correction (Table 5.1). The mean standardised effect sizes (β) for these 15 features was -0.26 (ranged: -0.65 $-$ 0.54) weeks of PMA per standard deviation increase in the feature. Of the significant features, 4 showed positive associations with PMA, indicating increasing values with advancing maturation, while 11 showed negative associations. Median peak inspiratory flow (PIF) was the feature with the largest positive association with PMA (β = 0.54, SE = 0.16, FDR-adjusted p = 0.0083), and apnoea rate was the feature with the strongest negative association (β = -0.65, SE = 0.15, FDR-adjusted p = 0.0013).

Feature	β	**SE**	**FDR-adjusted p**	**Pearson r**	**Pearson p**
Apnoea ($\geq$15s) rate per hour	-0.65	0.15	0.0013	-0.28	0.0001
mean I:E ratio	-0.58	0.15	0.0047	-0.21	0.0028
sd2	-0.56	0.16	0.0083	-0.22	0.0017
median PIF	0.54	0.16	0.0083	0.20	0.0056
skewness	-0.53	0.16	0.0083	-0.20	0.0061
max PSE	-0.50	0.15	0.0085	-0.23	0.0017
max IBI	-0.49	0.15	0.0085	-0.22	0.0024
IBI range	-0.49	0.15	0.0085	-0.22	0.0024
IBI s.d.	-0.53	0.16	0.0088	-0.21	0.0039
sd1	-0.49	0.16	0.0177	-0.19	0.0089
median I:E ratio	-0.47	0.16	0.0177	-0.11	0.1288
BWHm	-0.47	0.16	0.0177	-0.18	0.0142
HF normalised	0.47	0.17	0.0268	0.15	0.0339
ibi 5 centile	0.46	0.17	0.0365	0.23	0.0014
mean PIF	0.42	0.16	0.0422	0.14	0.0509

Table 5.1: Correlation between respiratory features and PMA for the 15 features that were found to be statistically significant. β represents the standardised effect size obtained from mixed-effects models, indicating the expected change in PMA (in weeks) per one standard deviation increase in the feature. SE denotes the standard error of the estimated effect size. I:E = inspiratory duration : expiratory duration; PIF = peak inspiratory flow; PSE = power spectral entropy; s.d. = standard deviation; BWHm = burst width-height gradient; HF = high-frequency power.

Pearson correlation coefficients between individual recordings and PMA were consistent in direction with the mixed-effects model estimates, supporting the observed associations (r = -0.28 to 0.23). Scatter plots of respiratory features against PMA showed visually apparent trends across recordings (Figure 5.1), particularly when the data were grouped into 2-week PMA intervals (Figure 5.2).

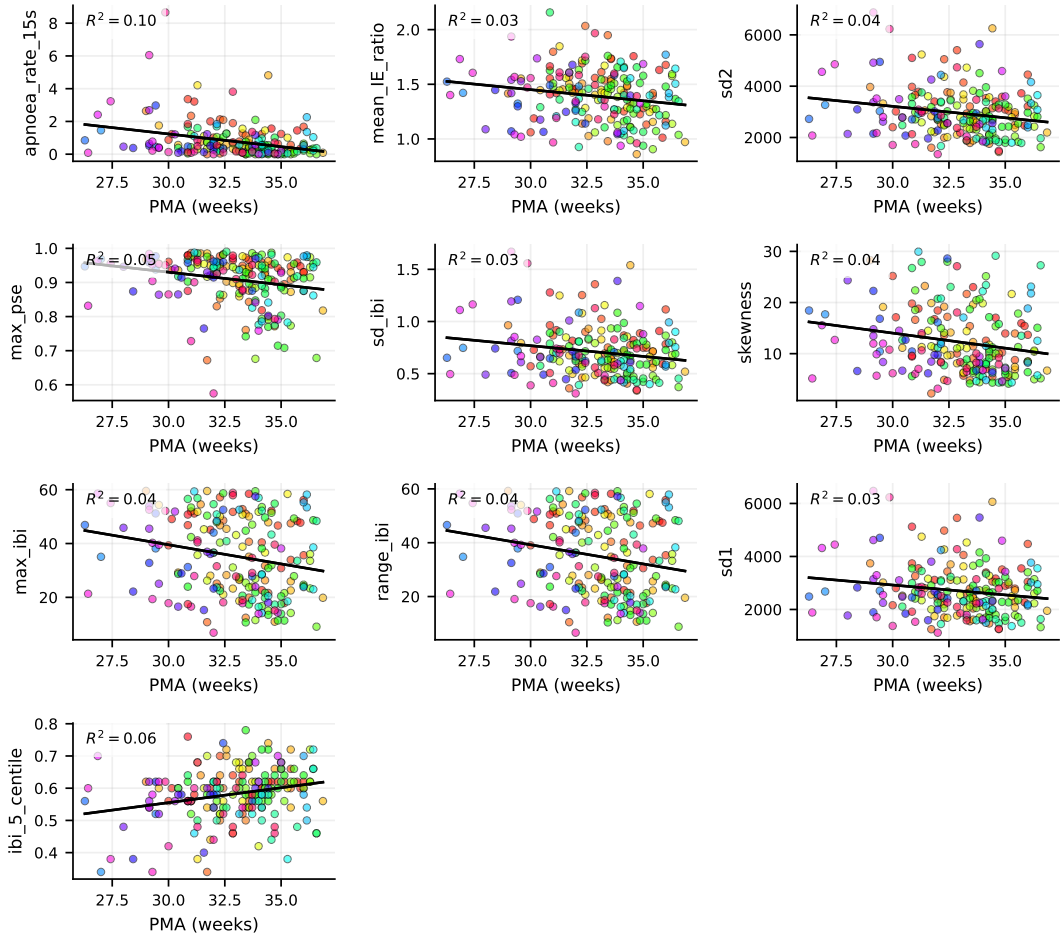

Figure 5.1: Scatter plots showing the relationship between PMA and respiratory features that were significantly associated with PMA in the mixed-effects models after FDR correction. Each point represents one recording, coloured by infant, and the black line indicates the line of best fit.

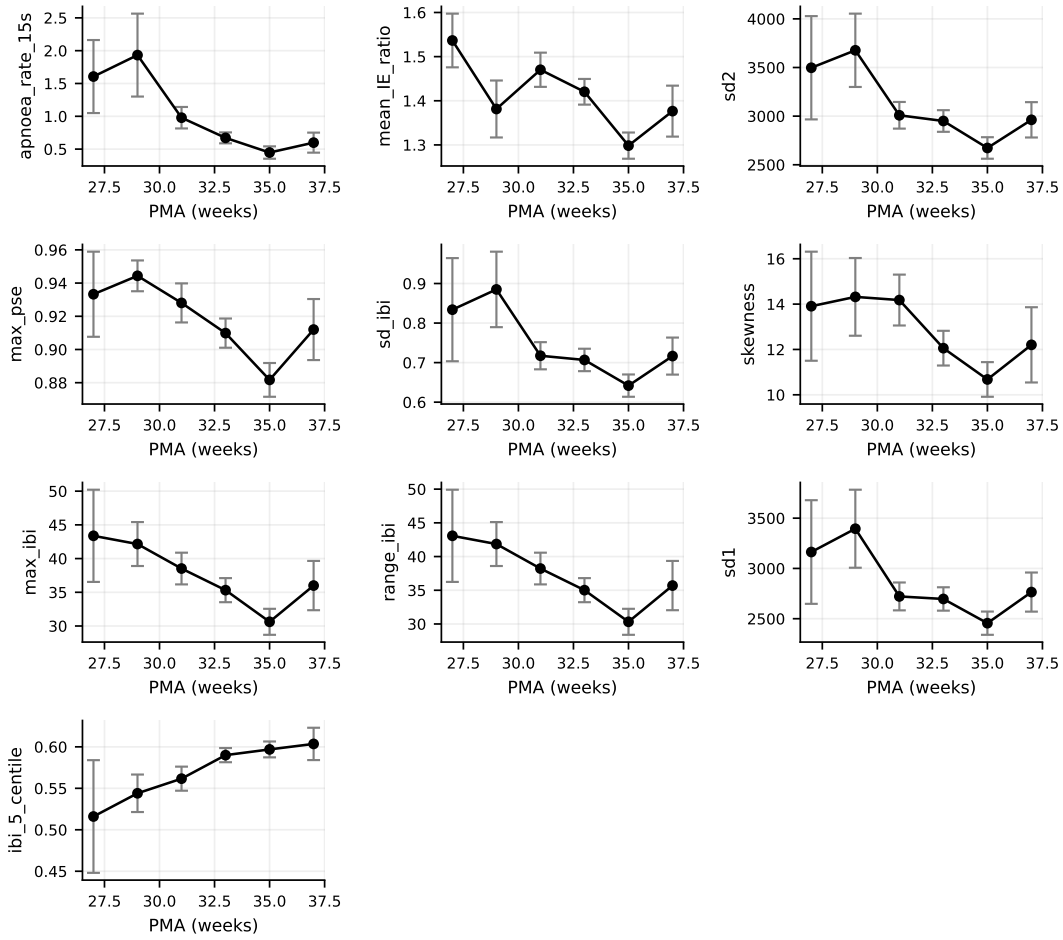

Figure 5.2: Mean respiratory feature values plotted against PMA after grouping recordings into 2-week PMA intervals for features that were significantly associated with PMA in the mixed-effects models after FDR correction. Points represent bin means and error bars indicate the standard error of the mean.

5.3.2 Can we predict PMA from respiratory features?

Regressors based on complete recordings

Multiple regression algorithms were evaluated to predict PMA from the 66 respiratory features. Across all models, the MAE ranged from 1.61 to 1.70 weeks, indicating modest predictive performance, but all models performed significantly better than the null model (Table 5.2). The scatter plots of predicted versus actual PMA (Figure 5.3) demonstrate a similar qualitative pattern across all algorithms, whereby the predictions are broadly centred around the population mean rather than tracking individual PMA values across the full range.

Statistical correlation analyses yielded small but statistically significant Pearson correlation coefficients (r ranging from 0.27 to 0.39, Table 5.2). However, the fitted linear slopes were close to zero (range = 0.07 to 0.21), indicating that the magnitude of the association was negligible.

Model	MAE	Perm. p	RMSE	R^2	r	Pearson p	Slope
RandomForest	1.63	0.0001	2.09	0.15	0.39	< 0.0001	0.16
SVR	1.66	0.0001	2.11	0.13	0.39	< 0.0001	0.21
Gradient Boosting	1.67	0.0001	2.12	0.12	0.37	< 0.0001	0.19
XGBoost	1.70	0.0002	2.15	0.11	0.35	< 0.0001	0.17
KNN	1.74	0.0002	2.20	0.06	0.25	0.0002	0.07
ElasticNet	1.80	0.0009	2.26	0.01	0.27	< 0.0001	0.14

Table 5.2: Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, without calibration. MAE = mean absolute error; Perm. p = p-value from permutation test between model and null model's MAE; RMSE = root mean squared error; r = Pearson correlation coefficient, Pearson p = two-sided p-value from the Pearson correlation test between predicted and actual PMA; slope = the slope of the best-fit linear regression line relating predicted to actual PMA.

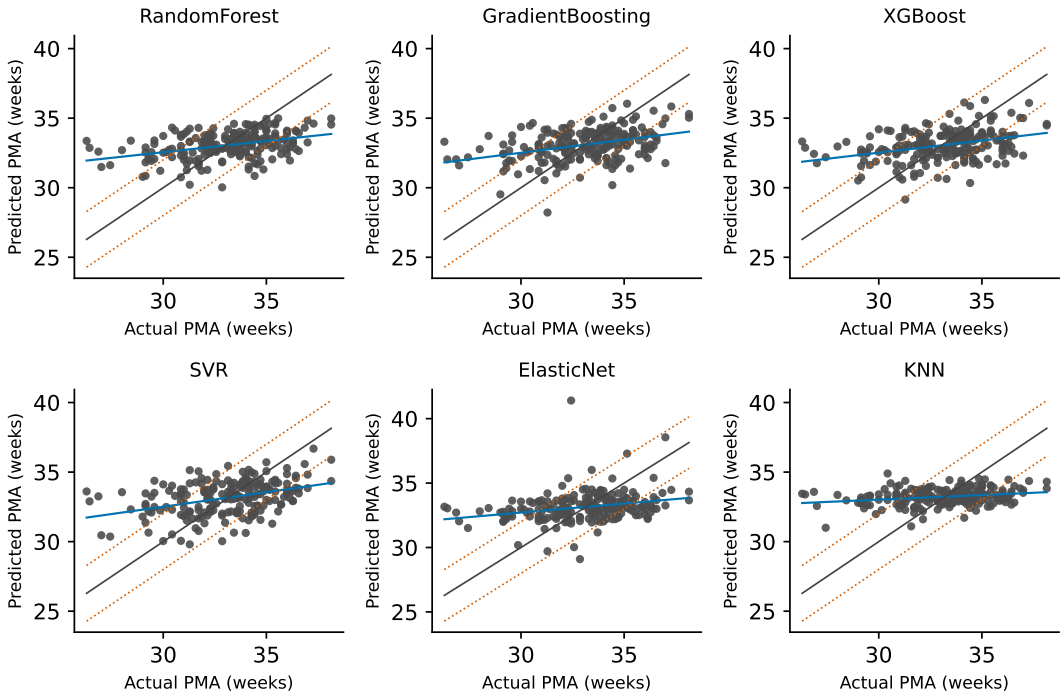

Figure 5.3: Predicted versus actual PMA obtained using six machine learning regression models. Grey dots represent individual predictions. The solid blue line shows the fitted linear regression line, the solid black line represents the identity line (prediction = actual), and the red dotted lines indicate the $\pm$2-week prediction interval.

After applying calibration to the regression outputs, the six machine learning models achieved MAE values ranging from 1.66 to 1.83 weeks (Table 5.3). The highest correlation coefficient and slope of the fitted relationship between predicted and actual PMA was 0.42 and 0.28 respectively, observed for the XGBoost model (Table 5.3, Figure 5.4). There were slightly higher than the slopes observed in the models without calibration.

Model	**MAE**	**Perm. p**	**RMSE**	**R^2**	**r**	**Pearson p**	**Slope**
RandomForest	1.66	0.0001	2.13	0.11	0.40	< 0.0001	0.25
XGBoost	1.68	0.0001	2.14	0.11	0.42	< 0.0001	0.28
SVR	1.69	0.0001	2.13	0.12	0.41	< 0.0001	0.26
Gradient Boosting	1.73	0.0001	2.19	0.07	0.39	< 0.0001	0.26
KNN	1.73	0.0001	2.19	0.07	0.26	0.00013	0.08
ElasticNet	1.83	0.0001	2.30	-0.03	0.31	< 0.0001	0.21

Table 5.3: Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, with calibration. MAE = mean absolute error; Perm. p = p-value from permutation test between model and null model's MAE; RMSE = root mean squared error; r = Pearson correlation coefficient, Pearson p = two-sided p-value from the Pearson correlation test between predicted and actual PMA; slope = the slope of the best-fit linear regression line relating predicted to actual PMA.

Therefore, a more comprehensive grid search was conducted using the XGBoost model with calibration applied. This optimised model yielded an MAE of 1.53, a statistically significant Pearson correlation between predicted and actual PMA ($r = 0.37$, $p < 0.0001$), and a slope of the best-fit line of 0.24.

Regressors based on selected features

When the Select K-Best feature selection method was applied, model performance did not improve. The MAE ranged from 1.75 to 1.81 weeks, and the slope ranged from 0.10 to 0.31

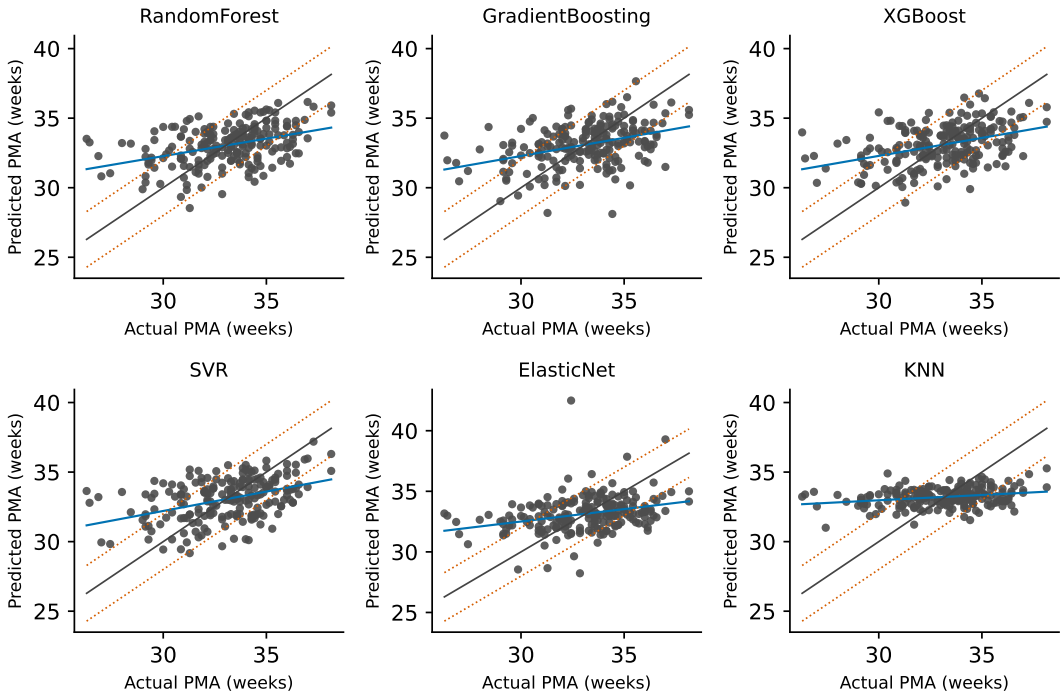

Figure 5.4: Predicted versus actual PMA obtained using six machine learning regression models with calibration. Grey dots represent individual predictions. The solid blue line shows the fitted linear regression line, the solid black line represents the identity line (prediction = actual), and the red dotted lines indicate the ±2-week prediction interval.

(Table 5.4, Figure 5.5). Similar results were observed when PCA was applied, with MAE ranging from 1.65 to 1.84 weeks and slope ranging from 0.08 to 0.26 (Table 5.5, Figure 5.6). Feature selection was not used in the subsequent analyses since no obvious improvements in the performances were observed.

Model	**MAE**	**Perm. p**	**RMSE**	**R^2**	**r**	**Pearson p**	**Slope**
XGBoost	1.75	0.001	2.25	0.02	0.40	< 0.0001	0.31
ElasticNet	1.76	0.001	2.22	0.04	0.29	< 0.0001	0.14
KNN	1.78	0.001	2.24	0.02	0.23	0.0008	0.10
SVR	1.78	0.001	2.21	0.05	0.32	0.0002	0.18
Gradient Boosting	1.78	0.001	2.32	-0.04	0.34	< 0.0001	0.26
RandomForest	1.81	0.001	2.27	0.00	0.36	< 0.0001	0.25

Table 5.4: Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, with Select K-Best feature selection and calibration. MAE = mean absolute error; Perm. p = p-value from permutation test between model and null model's MAE; RMSE = root mean squared error; r = Pearson correlation coefficient, Pearson p = two-sided p-value from the Pearson correlation test between predicted and actual PMA; slope = the slope of the best-fit linear regression line relating predicted to actual PMA.

Model	**MAE**	**Perm. p**	**RMSE**	**R^2**	**r**	**Pearson p**	**Slope**
SVR	1.65	0.001	2.11	0.13	0.42	< 0.0001	0.26
RandomForest	1.71	0.001	2.14	0.11	0.38	< 0.0001	0.22
KNN	1.72	0.001	2.18	0.07	0.28	< 0.0001	0.08
ElasticNet	1.73	0.001	2.16	0.09	0.35	< 0.0001	0.18
Gradient Boosting	1.78	0.001	2.29	-0.02	0.29	< 0.0001	0.18
XGBoost	1.84	0.001	2.31	-0.03	0.30	< 0.0001	0.19

Table 5.5: Result metrics for the six regressors used to predict PMA from respiratory features for each complete recording, with PCA feature selection and calibration. MAE = mean absolute error; Perm. p = p-value from permutation test between model and null model's MAE; RMSE = root mean squared error; r = Pearson correlation coefficient, Pearson p = two-sided p-value from the Pearson correlation test between predicted and actual PMA; slope = the slope of the best-fit linear regression line relating predicted to actual PMA.

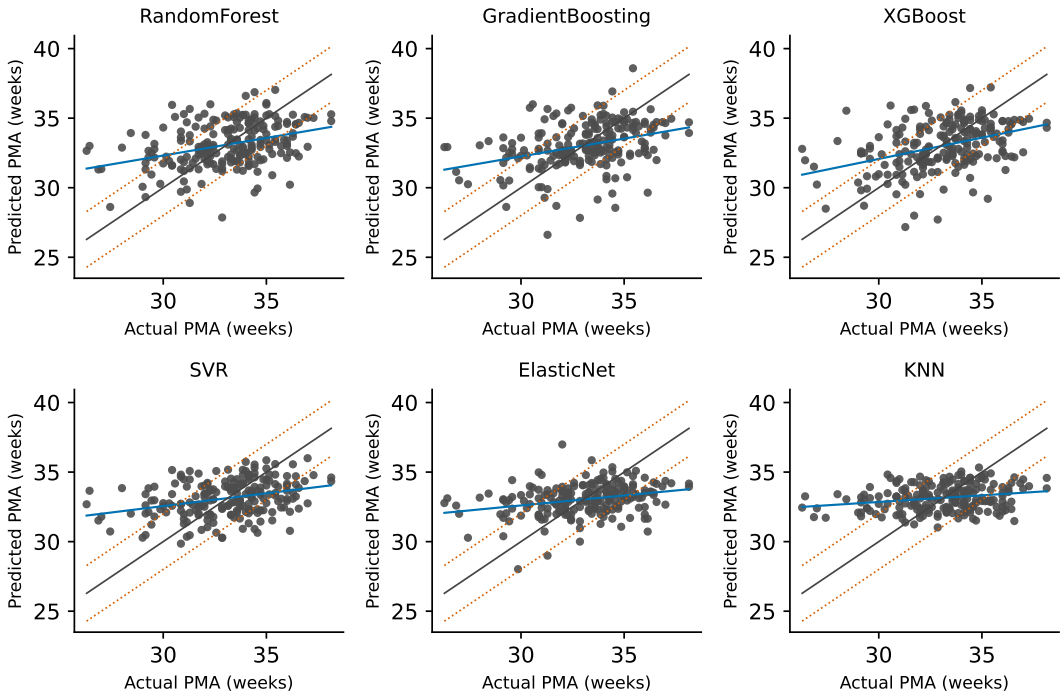

Figure 5.5: Predicted versus actual PMA obtained using six machine learning regression models with select k-best features selection and calibration. Grey dots represent individual predictions. The solid blue line shows the fitted linear regression line, the solid black line represents the identity line (prediction = actual), and the red dotted lines indicate the $\pm$2-week prediction interval.

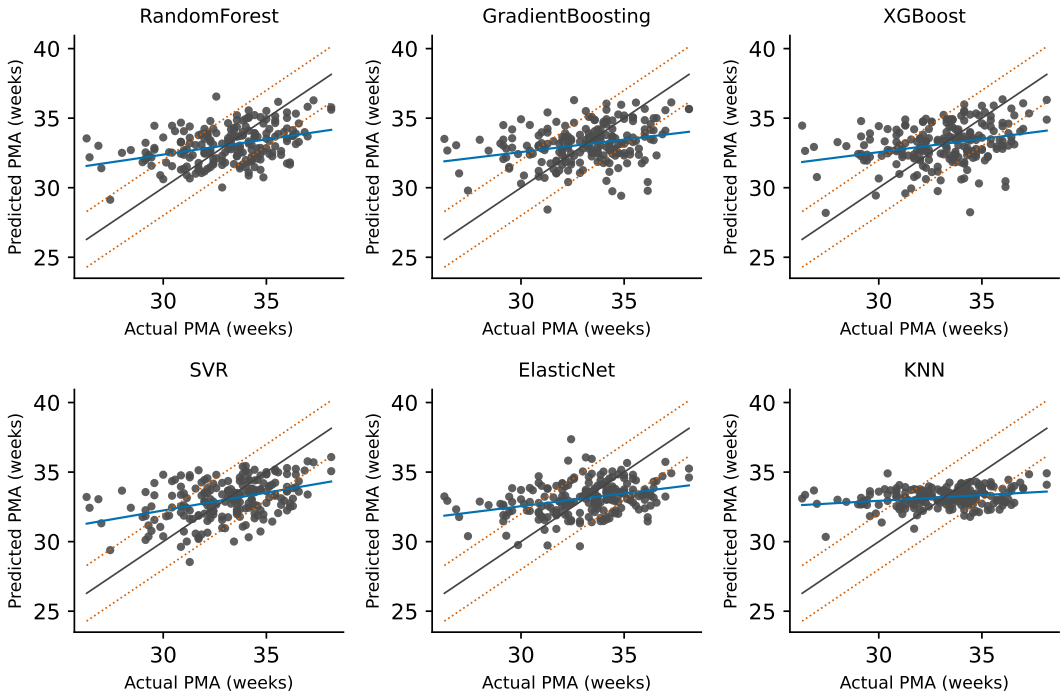

Figure 5.6: Predicted versus actual PMA obtained using six machine learning regression models with PCA features selection and calibration. Grey dots represent individual predictions. The solid blue line shows the fitted linear regression line, the solid black line represents the identity line (prediction = actual), and the red dotted lines indicate the $\pm$2-week prediction interval.

Regressors based on one-hour epoched data

By extracting features based on 1 hours epochs and taking averages across the predictions for all epochs from the same recording, the models achieved MAE ranged from 1.55 to 1.61 without calibration (Table 5.6), and 1.51 to 1.59 with calibration (Table 5.7). All models demonstrated statistically significant Pearson correlations between predicted and actual PMA. In the non-calibrated models, correlation coefficients ranged from 0.41 to 0.51, while in the calibrated models they ranged from 0.43 to 0.51.

Model	**MAE**	**Perm. p**	**RMSE**	**R^2**	**r**	**Pearson p**	**Slope**
SVR	1.55	< 0.0001	1.98	0.22	0.51	< 0.0001	0.16
RandomForest	1.59	< 0.0001	2.01	0.20	0.46	< 0.0001	0.17
ElasticNet	1.60	< 0.0001	2.02	0.19	0.47	< 0.0001	0.14
KNN	1.61	< 0.0001	2.04	0.18	0.46	< 0.0001	0.13
GradientBoosting	1.62	< 0.0001	2.05	0.16	0.42	< 0.0001	0.13
XGBoost	1.62	< 0.0001	2.06	0.16	0.41	< 0.0001	0.14

Table 5.6: Result metrics for the six regressors used to predict PMA from respiratory features for for one-hour epoched data, without calibration. MAE = mean absolute error; RMSE = root mean squared error; r is the Pearson correlation coefficient and p its two-sided p-value from the Pearson correlation test between predicted and actual PMA; "slope" denotes the slope of the best-fit linear regression line relating predicted to actual PMA.

The slope of the best-fit line between predicted and actual PMA ranged from 0.13 to 0.17 for the models trained on epoched data without calibration (Table 5.6, Figure 5.7), and from 0.15 to 0.23 for the models with calibration (Table 5.7, Figure 5.8). Overall, model performance on epoched data was not improved compared with models based on complete recordings.

Model	MAE	Perm. p	RMSE	R^2	r	Pearson p	Slope
SVR	1.51	< 0.0001	1.95	0.25	0.51	< 0.0001	0.20
RandomForest	1.58	< 0.0001	2.00	0.21	0.46	< 0.0001	0.23
ElasticNet	1.58	< 0.0001	2.00	0.21	0.48	< 0.0001	0.16
XGBoost	1.59	< 0.0001	2.03	0.18	0.43	< 0.0001	0.21
GradientBoosting	1.59	< 0.0001	2.04	0.18	0.43	< 0.0001	0.21
KNN	1.59	< 0.0001	2.03	0.18	0.45	< 0.0001	0.15

Table 5.7: Result metrics for the six regressors used to predict PMA from respiratory features for one-hour epoched data, with calibration. MAE = mean absolute error; RMSE = root mean squared error; r is the Pearson correlation coefficient and p its two-sided p-value from the Pearson correlation test between predicted and actual PMA; "slope" denotes the slope of the best-fit linear regression line relating predicted to actual PMA.

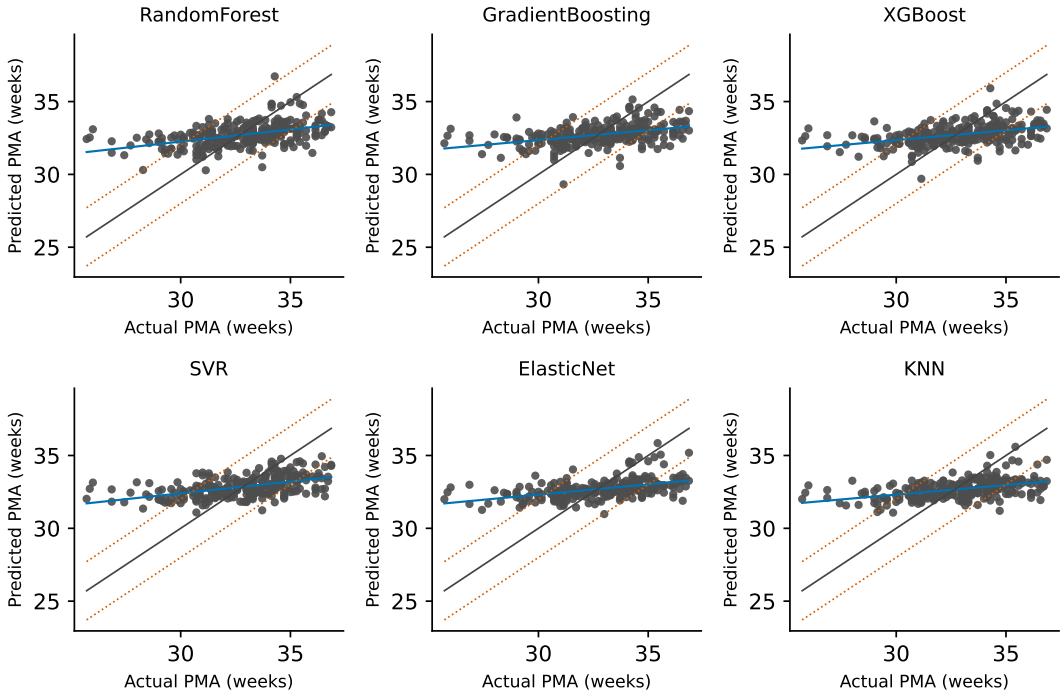

Figure 5.7: Predicted versus actual PMA obtained using six machine learning regression models for one-hour epoched data, without calibration. Grey dots represent individual predictions. The solid blue line shows the fitted linear regression line, the solid black line represents the identity line (prediction = actual), and the red dotted lines indicate the $\pm$2-week prediction interval.

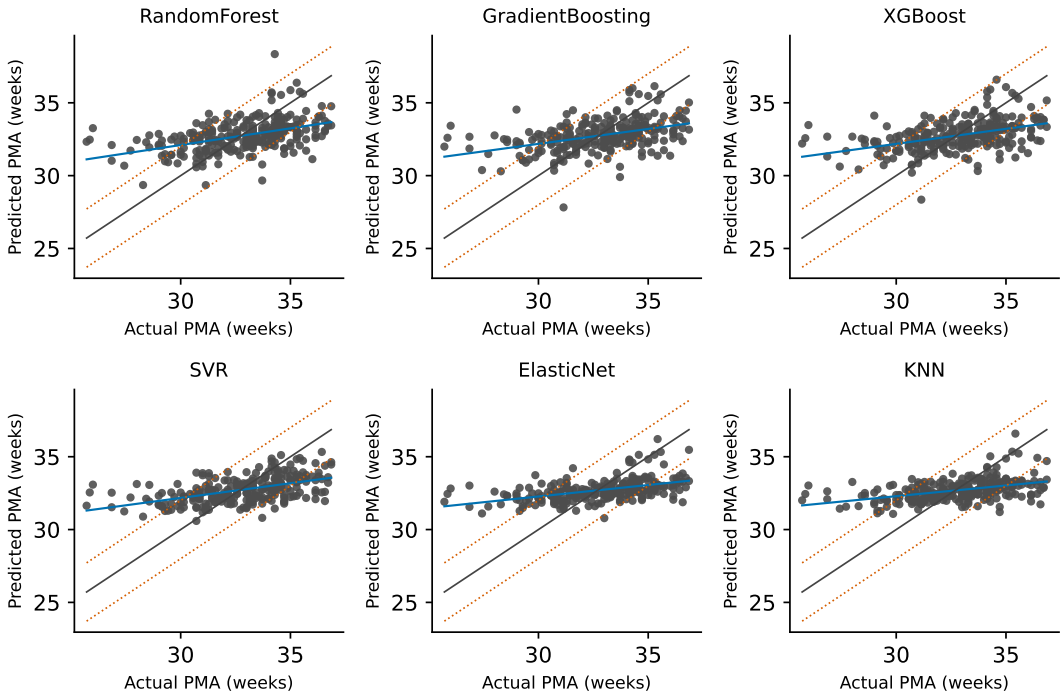

Figure 5.8: Predicted versus actual PMAobtained using six machine learning regression models for one-hour epoched data, with calibration. Grey dots represent individual predictions. The solid blue line shows the fitted linear regression line, the solid black line represents the identity line (prediction = actual), and the red dotted lines indicate the $\pm$2-week prediction interval.

Classifying the degree of maturity

To complement the regression analyses, classification models were also developed to assess whether respiratory features could distinguish broader stages of respiratory maturation. Rather than predicting PMA as a continuous outcome, this approach grouped recordings into three clinically relevant maturity categories: PMA < 32 weeks (extremely or very preterm), PMA 32–34 weeks (moderate preterm), and PMA ≥ 34 weeks (late preterm). These categories were chosen to test whether respiratory features could identify infants at different stages of maturation, even if precise prediction of continuous age remained challenging.

The three-class classification models demonstrated the ability to predict infants' degree of maturity based on respiratory features. Among the evaluated models, Support Vector Classifier (SVC) achieved the best overall performance, with a balanced accuracy of 53.9% and an AUROC of 0.67 (Table 5.8). Although overall predictive performance was modest, the model correctly identified the true maturity category as the most frequently predicted class, as illustrated in its confusion matrix (Figure 5.9). These findings suggest that respiratory features contain information about maturational status, although classification performance remains limited.

5.3.3 Morphine affects respiratory function in preterm infants

Based on the findings from the preceding analyses, all regressors had a similar range of MAE, and all performed significantly better than the null model. However, because the range of included PMA values was relatively narrow, models that predicted most values close to the median could still achieve a good MAE. Therefore, slope was used to identify the best-performing

Model	Balanced accuracy	AUROC	Perm. p
SVC	53.9	0.67	0.0010
XGBoost	52.7	0.68	0.0010
Logistic Regression	51.3	0.67	0.0010
Gradient Boosting	49.7	0.66	0.0010
Random Forest	47.5	0.64	0.0010
KNN	45.3	0.63	0.0020

Table 5.8: Performance metrics of the three-class classification models for predicting maturational category from respiratory features. Balanced accuracy accounts for class imbalance, AUROC represents the macro-averaged area under the receiver operating characteristic curve across the three maturity classes, and Perm. p represents the p-value from the permutation test between accruacy of the models and null model. SVC = Support Vector Classifier, KNN = K-Nearest-Neighbour.

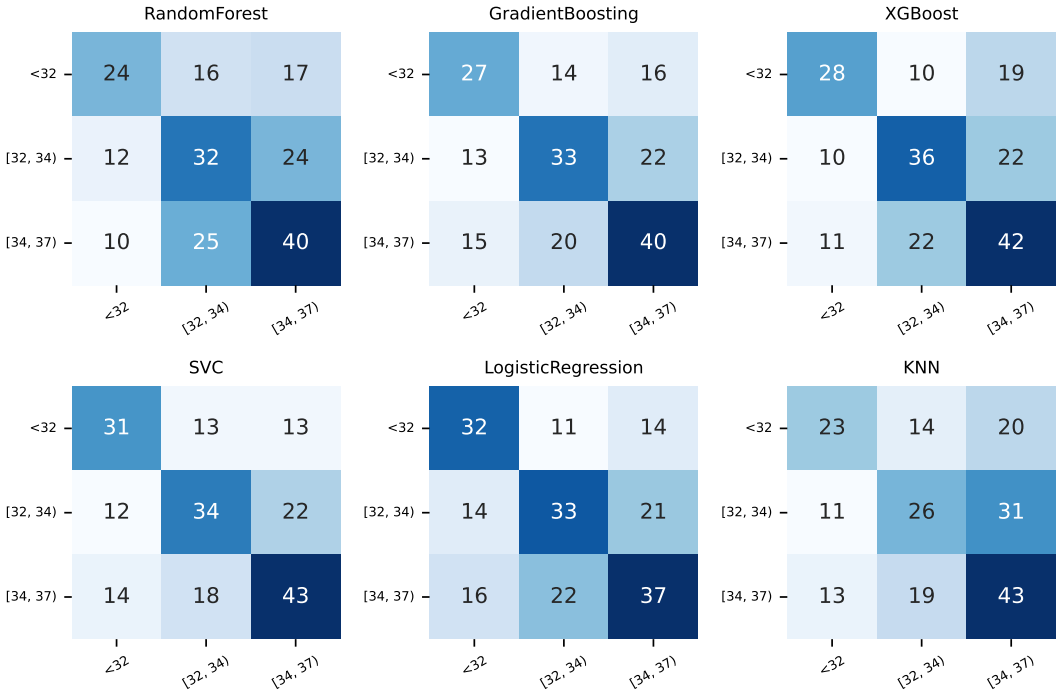

Figure 5.9: Confusion matrices for the six classification models predicting maturational category from respiratory features. Rows represent the true maturity class and columns represent the predicted class. Values indicate the number of recordings classified into each category. The diagonal elements correspond to correct classifications. SVC = Support Vector Classifier, KNN = K-Nearest-Neighbour.

model. The selected model was XGBoost with linear calibration, using all features and applied to complete recordings. This model was then used to assess morphine-related effects on respiratory maturity. This analysis was conducted using Dataset 3, described in Section 2.2, which comprised 30 infants (morphine group, $\mathrm{N} = 15$; placebo group, $\mathrm{N} = 15$).

Respiratory features were first extracted from recordings of the 15 infants who received morphine. Pre- and post-administration feature values were compared to assess the acute effect of morphine on respiratory dynamics. Thirteen features demonstrated statistically significant differences following morphine administration (Table 5.9), indicating measurable alterations in respiratory variability and signal characteristics associated with the intervention.

To further characterise temporal dynamics, respiratory features were subsequently extracted using one-hour moving windows with 30 minutes overlap across the recording period. Significant post-administration shifts in multiple respiratory features were observed in the morphine group, whereas no comparable systematic changes were evident in the placebo group (Figure 5.10). These findings suggest that morphine exerts measurable effects on respiratory physiology at the feature level.

FRA was then estimated for each time window using the selected XGBoost model. The change in FRA after drug administration (predicted FRA minus the average FRA before drug administration) showed a decrease following morphine administration (Figure 5.11A), indicating that the breathing pattern induced by morphine is more similar to that of physiologically younger infants. This finding is consistent with the known respiratory depressant effects of morphine, which reduces respiratory drive and alters breathing patterns.

Feature	Mean before morphine	Mean after morphine	Mean differences	FDR-adjusted p
IBI s.d.	0.50	0.57	0.07	0.028
Long-term IBI variability	2098.25	2391.56	293.31	0.028
Short-term IBI variability	1878.71	2182.85	304.14	0.028
LF normalised	0.67	0.57	-0.10	0.028
BWHm	0.00	0.00	0.00	0.028
HF normalised	20.57	17.88	-2.70	0.028
BA average	125.15	142.04	16.90	0.033
median SNLEO	0.27	0.31	0.04	0.033
IBI 95 centile	1.65	1.76	0.11	0.033
Apnoea ($\geq$5s) rate per hour	8.64	12.62	3.98	0.033
Minimum Burst Value	0.81	0.98	0.17	0.033
max PSE	0.79	0.82	0.02	0.037
Band power 2	0.10	0.09	0.00	0.046

Table 5.9: Results of paired t-tests comparing respiratory feature values before and after morphine administration. The p-values were obtained from paired t-tests using FDR adjustment. Only features that demonstrated statistically significant differences (FDR-adjusted $p < 0.05$) are presented in this table. s.d. = standard deviation; LF = low-frequency power; HF = high-frequency power; BWHm = burst width-height gradient, BA average = mean area of each burst; SNLEO = smoothed non-linear energy operator; PSE = power spectral entropy.

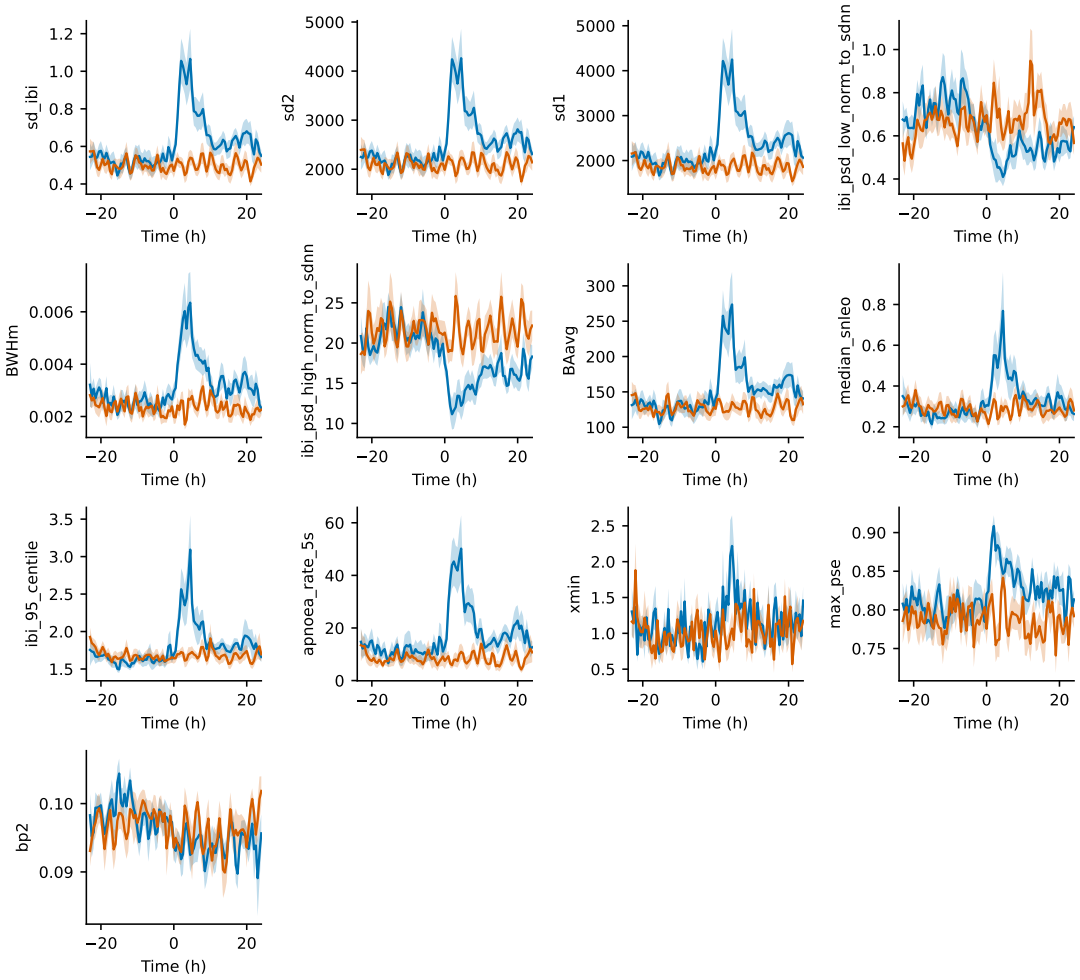

Figure 5.10: Selected respiratory features plotted against time relative to drug administration. Time 0 denotes the moment of drug administration. Blue curves represent the morphine group, and orange curves represent the placebo group. Solid lines indicate the mean value within each one-hour window, and shaded areas denote the standard error of the mean.

Six infants experienced severe respiratory depression following morphine administration. These infants also exhibited substantially larger decreases in FRA (Figure 5.11B), indicating that larger changes in respiratory physiology were associated with greater deviations from the expected respiratory maturity trajectory. This observation supports the hypothesis that FRA reflects underlying respiratory physiology and is sensitive to clinically significant changes in respiratory function.

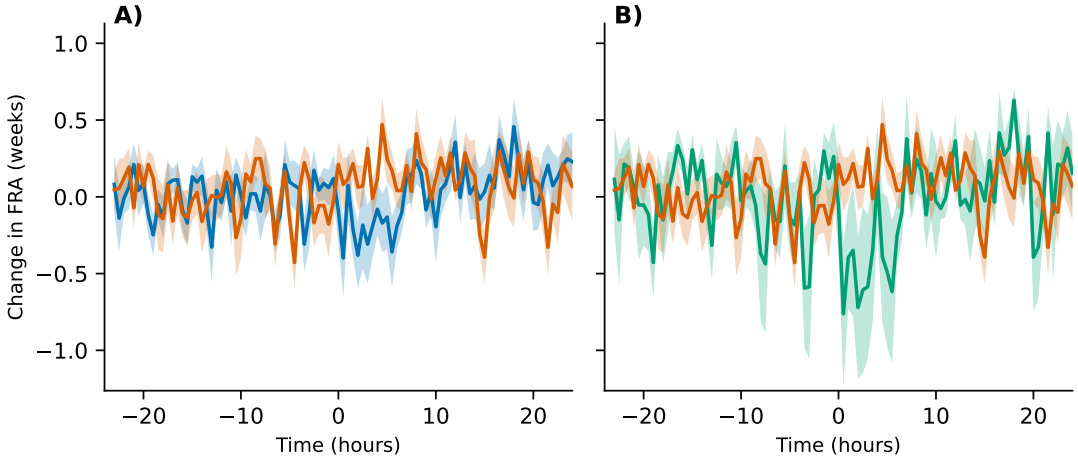

Figure 5.11: Change in FRA against time relative to drug administration. Time 0 denotes the moment of drug administration. Change was calculated by FRA after drug administration minus the average FRA before administration. Orange curves represent the placebo group, blue curves represent the morphine group, and green curves represent the infants who had severe reaction to morphine. Solid lines indicate the mean value within each one-hour window, and shaded areas denote the standard error of the mean.

5.4 Discussion

This chapter evaluated whether respiratory features derived from impedance pneumography and inter-breath interval signals follow a sufficiently consistent maturational trajectory in clinically stable preterm infants to support estimation of functional respiratory age. Specifically,

it investigated whether respiratory physiology changes systematically with PMA such that the underlying maturational stage of an infant could be inferred from respiratory features alone. Both regression and classification approaches were implemented to assess whether respiratory dynamics encode information about expected respiratory maturity within this preterm population.

Across all regression models, prediction of PMA from respiratory features yielded modest performance. Although calibration improved numerical metrics, the overall magnitude of improvement was small. Moreover, scatter plots of predicted versus actual PMA consistently demonstrated a relatively flat trend, with fitted slopes substantially below unity (maximum slope = 0.31). This indicates that, despite statistically significant correlations, the models explained only a limited proportion of the variability in PMA across the study population.

The discrepancy between statistically significant Pearson correlations and small slopes highlights an important distinction between statistical and practical significance. From a predictive standpoint, the models largely regressed toward the cohort mean rather than tracking individual developmental trajectories. In practical terms, respiratory features alone were insufficient to produce clinically meaningful estimates of PMA at the individual level.

An alternative interpretation is that respiratory maturation in preterm infants may not follow a single, tightly constrained developmental trajectory. Rather, respiratory development is influenced by multiple physiological systems, including pulmonary development, autonomic regulation, central nervous system maturation, previous respiratory support, and other clinical factors, resulting in substantial heterogeneity between infants of the same PMA [26, 27, 73].

Consequently, the modest regression performance observed in this study may reflect genuine biological variability rather than solely limitations of the modelling approach. Furthermore, although the respiratory feature set was extensive, it may capture variability in breathing patterns without isolating the specific maturational components that change consistently with age. Within-infant variability and measurement noise may further weaken the relationship between respiratory dynamics and PMA.

An important assumption underlying the FRA approach is that PMA provides a suitable reference for respiratory maturity. However, infants at the same PMA may not have reached the same stage of respiratory development, particularly if they were born at different gestational ages. As discussed in previous chapters, infants born at lower GA tend to experience apnoea until a later PMA [41], suggesting that respiratory maturation is influenced by GA at birth as well as current PMA. Furthermore, infants born at lower GA have experienced a longer period of extrauterine life at a given PMA, potentially exposing them to different respiratory support, treatments and clinical complications that may influence respiratory development. Therefore, using PMA as the reference for FRA may not fully capture individual differences in respiratory maturity. This could contribute to the variability in respiratory features observed between infants of the same PMA and partly explain the limited performance of the regression models. Future studies could investigate whether incorporating GA at birth and postnatal clinical history would improve the estimation of functional respiratory maturity.

Applying linear calibration to regression outputs consistently improved performance metrics. Calibration reduced systematic bias by correcting slope shrinkage and intercept offset observed in the raw model outputs [194–196]. This suggests that a component of prediction error was

attributable to model miscalibration rather than complete absence of signal. Nevertheless, even after calibration, the fitted slopes remained small, indicating that the underlying predictive signal was weak.

Although regression models failed to produce accurate continuous age predictions, classification models demonstrated more promising results. The confusion matrices showed that, for each maturity class, the correct class was the most frequently predicted label. This indicates that respiratory signals contain valuable information to distinguish a wide range of mature categories, even if they cannot address fine-grained age differences. Differences between $<$32 weeks, 32-34 weeks, and 34-37 weeks PMA may reflect structural and functional milestones in respiratory control systems. Classification performance therefore supports the hypothesis that respiratory dynamics encode clinically meaningful maturational information.

This study demonstrates that morphine administration is associated with measurable alterations in individual respiratory features as well as functional respiratory age in preterm infants. Thirteen respiratory features changed significantly following morphine administration, reflecting more irregular breathing patterns consistent with the known respiratory depressant effects of morphine [108]. These changes were not observed to the same extent in infants who received placebo, indicating that they were attributable to morphine rather than spontaneous temporal variability. Importantly, the observed changes were consistent with respiratory physiology becoming less mature rather than more mature, supporting the interpretation that morphine transiently suppresses respiratory function.

Time-resolved analysis using one-hour windows further revealed that respiratory features in

the morphine group showed significant deviations during the six hours following drug administration when compared with the placebo group. This temporal pattern is consistent with the known pharmacokinetics and physiological effects of morphine, supporting the validity of the feature-based analysis.

In addition to feature-level changes, a modest but detectable reduction in FRA was observed following morphine administration. Because PMA remained unchanged throughout the recording period, this reduction indicates that morphine altered respiratory physiology and changed the respiratory characteristics to a stage that is physiologically closer to younger infants. Both infants with severe respiratory depression and those without severe reactions showed reductions in FRA, although the reduction was substantially greater in the severe-reaction subgroup. This finding is biologically plausible because morphine depresses respiratory drive in all infants, with severe respiratory depression representing a more pronounced effect. The larger reduction in FRA observed in infants with severe respiratory depression therefore suggests that FRA is sensitive to the magnitude of respiratory physiological change rather than merely reflecting background variability or chronological age.

Although the overall drop in FRA change was smaller than the feature-level changes, this finding is not unexpected. The FRA model represents a composite measure derived from multiple respiratory features and is trained to capture maturational trajectories rather than acute pharmacological effects. Therefore, transient physiological disturbances may manifest as relatively subtle shifts in FRA compared with individual feature changes.

This study has several limitations. First, the dataset was not uniformly distributed across PMA

groups. A greater proportion of infants were in the moderate preterm range (32–34 weeks PMA), whereas fewer recordings were available from infants in the younger and older PMA groups. This imbalance reflects the clinical realities of neonatal care and may have introduced selection bias. At younger PMAs, only relatively stable infants who required no more than high flow respiratory support and were not receiving the excluded medications could be included. These infants may therefore represent a particularly stable subgroup for their age, rather than the wider population of extremely preterm infants. Conversely, healthy late preterm infants are more likely to be discharged or transferred to other hospitals, meaning that older infants remaining in the neonatal unit may represent a more clinically complex population. As a result, the younger infants included in this study may have been relatively more physiologically mature for their age, while the older infants may have had greater respiratory vulnerability than expected for their PMA. This selection bias could have reduced the apparent differences in respiratory features between PMA groups, potentially contributing to the modest regression performance and limiting the generalisability of the FRA models. This may also explain why several respiratory features showed a reversal or flattening of their maturational trends beyond approximately 35 weeks PMA (Figure 5.2). Future studies could investigate whether excluding infants older than 35 weeks PMA improves the consistency of the observed maturational trajectories and the performance of FRA models. However, such an approach would reduce the age range covered by the model, and therefore its potential applicability to older preterm infants.

Second, although exclusion criteria were applied, they may not have been sufficiently comprehensive to eliminate all physiological confounders. Some included infants may still have

had underlying clinical conditions that influenced respiratory function. More fundamentally, an important limitation of this study is the absence of an independent, objective reference measure of functional respiratory age. PMA was therefore used as a surrogate for expected respiratory maturity, based on the assumption that respiratory development follows a sufficiently consistent trajectory in clinically stable preterm infants. However, PMA reflects chronological developmental age rather than directly measured respiratory maturity, and infants at the same PMA may differ substantially in their respiratory function. Consequently, the models were evaluated on their ability to predict PMA rather than their ability to estimate an independently validated measure of respiratory maturity. This distinction is important when interpreting the modest regression performance, as prediction errors may reflect both limitations of the respiratory features and differences between chronological and physiological maturity. Therefore, the FRA estimates derived in this study should be considered exploratory indicators of respiratory maturity rather than validated measurements of functional respiratory age.

Finally, due to limited data availability, the model was not evaluated on an independent external dataset. External validation is essential to determine the robustness and generalisability of predictive models. Future work should therefore include validation in independent cohorts, ideally collected at different clinical sites. Without such validation, it remains possible that the modest regression performance observed in this study reflects cohort-specific variability rather than intrinsic limitations of respiratory biomarkers.

In summary, this chapter demonstrates that respiratory signals derived from impedance pneumography contain clinically meaningful information about developmental maturity and physiological state in preterm infants. Although regression-based estimation of PMA yielded only

modest performance and limited sensitivity to fine-grained maturational differences, classification analyses confirmed that respiratory dynamics can discriminate between broad maturity categories. Furthermore, morphine administration was associated with significant feature-level changes and detectable shifts in the functional respiratory age gap, particularly in infants with severe clinical reactions. These findings suggest that while respiratory features alone may not support precise respiratory age prediction, they do encode biologically relevant information related to both maturation and acute physiological perturbation. Collectively, the results support the conceptual framework of functional respiratory age as a physiologically grounded marker of respiratory maturity, while highlighting the need for improved modelling strategies, broader datasets, and external validation to fully realise its potential as a clinically useful biomarker.

Chapter 6

Conclusion and future work

6.1 Summary of the main results

This thesis examined how physiological signals can be used to better understand respiratory instability and maturation in preterm infants, with the overarching aim of supporting more individualised neonatal monitoring. Across three main studies in the thesis, the findings show that the physiological consequences of apnoea in preterm infants vary with event duration, infant age, ventilation mode, recent apnoea history, and baseline physiological state. These results provide evidence that current one-size-fits-all approaches to apnoea detection and respiratory monitoring do not adequately reflect the heterogeneity of risk in this population.

The first study, a systematic review and meta-analysis, established the overall evidence base linking apnoea duration with physiological changes in preterm neonates. Across published studies, longer apnoeas were consistently associated with greater reductions in heart rate, oxygen saturation, and cerebral oxygenation. These findings support the biological expectation that prolonged apnoea produce greater cardiorespiratory and cerebrovascular instability. At the same time, the review highlighted substantial heterogeneity both across studies and between individual infants, indicating that apnoea duration alone is not sufficient to explain physio-

logical response. Importantly, PMA was shown as a significant modulator of the relationship between apnoea duration and oxygen desaturation, suggesting that maturational state influences tolerance to respiratory pauses. The review also showed that the literature is limited by small sample sizes, inconsistent definitions, older monitoring technologies, and incomplete reporting of potentially important confounders. This provided the rationale for analysing a large contemporary cohort in the subsequent chapters.

The second study extended these findings using a large modern dataset of preterm infants and confirmed that both bradycardia and desaturation are strongly related to apnoea duration, but with marked between-infant variability. Although longer apnoeas increase the probability of cardiorespiratory instability overall, some infants exhibit substantial cardiorespiratory instability even after relatively short apnoea episodes, whereas others remain more stable during longer events. This finding challenges the common definition of apnoeas, which typically only focuses on apnea lasting more than 10 or 20 seconds, and suggests that fixed duration-based definitions of apnoea may fail to identify clinically important events in some infants.

A major contribution of this thesis is the identification of multiple factors that modulate the risk of apnoea-associated bradycardia and desaturation. Younger PMA and GA were both associated with greater cardiorespiratory instability, consistent with developmental immaturity of respiratory control, pulmonary function, and oxygen transport. In addition, infants experiencing recurrent or clustered apnoeas were at greater risk of instability during subsequent events. Baseline physiological status and respiratory support also provide information indicating that the immediate cardiorespiratory condition of the infant changes the impact of a given apnoea.

These modulating factors were then incorporated into predictive models, which demonstrated that apnoeas associated with bradycardia and/or desaturation can be predicted with promising accuracy. Models including dynamic physiological features performed best, but models based only on demographic and clinical variables also retained useful predictive ability. This is an important translational result, as it suggests that individual risk stratification may be feasible even when continuous advanced signal features are unavailable. The prototype framework developed in this chapter therefore shows a potential of having personalised apnoea alarm thresholds or infant-specific definitions of clinically significant apnoea.

The third study addressed a related but distinct question on whether respiratory features derived from impedance pneumography and inter-breath interval signals encode maturational information in preterm infants. The results suggested that respiratory dynamics do contain information related to developmental status, but that this information is better suited to broad maturity classification than to precise continuous age prediction. Regression models predicting PMA from respiratory features yielded only modest performance and showed clear shrinkage toward the cohort mean, indicating that respiratory features alone are insufficient for accurate individual-level estimation of PMA. However, classification models were more successful, with broad maturity groups being distinguishable from respiratory signals. This study also demonstrated significant changes in multiple respiratory features and a detectable reduction in functional respiratory age following morphine administration. The larger reductions in infants with more severe respiratory depression suggest that functional respiratory age is sensitive to clinically meaningful perturbations in respiratory physiology.

Overall, the findings of this thesis support three main conclusions. First, apnoea-related physi-

ological instability in preterm infants is highly heterogeneous and cannot be adequately characterised by duration alone. Second, this heterogeneity is partly explainable and predictable using infant-specific clinical, demographic, and dynamic physiological factors. Third, although respiratory features contain some information related to respiratory maturation, their limited ability to predict PMA suggests that further development and validation are required before they can be used to estimate functional respiratory age reliably. Taken together, the findings can facilitate more personalised neonatal monitoring, although further work is required before these approaches can be implemented clinically.

6.2 Limitations and future directions

Several limitations should be considered when interpreting the findings of this thesis. Some are specific to individual studies, while others apply more broadly to the overall programme of work.

First, although this thesis combined evidence synthesis with large-scale observational analysis, much of the underlying data remains observational in nature. As a result, identified associations should not be interpreted as evidence of causality. For example, factors such as ventilation mode, baseline physiological instability, or recent apnoea clustering were associated with increased risk of bradycardia and desaturation, but these factors may also be markers of overall illness severity or underlying pathology rather than direct causal determinants of poorer apnoea tolerance. Future prospective studies should therefore move beyond identifying statistical associations and instead investigate the physiological mechanisms underlying. One potential

approach would be to combine continuous respiratory, cardiovascular, and cerebral monitoring with time-resolved clinical annotation to characterise the sequence of physiological events preceding apnoea-related bradycardia and desaturation. Such studies could help distinguish factors that directly contribute to cardiorespiratory instability from those that only reflect underlying illness severity.

Second, the observational analyses in this thesis were limited by the variables available in the pre-collected dataset. Some clinically relevant factors could not be reliably measured at the level of individual apnoea events. In particular, infection status, caffeine exposure and timing, and the exact timing or dose of some medications could not be determined with sufficient precision. The lack of these features in the model may have reduced model performance and may also mean that part of the observed variability in physiological response remains unexplained. Future predictive studies should therefore integrate richer, time-resolved clinical and physiological information. Models should incorporate variables such as trends in cardiorespiratory signals, and responses to preceding apnoeas. Such information would enable prediction models to distinguish developmental maturation from transient physiological influences. This would allow researchers to move towards estimating an infant's current risk to clinically significant cardiorespiratory instability, rather than simply characterising individual apnoea events.

Third, apnoea detection itself is imperfect. The impedance pneumography-derived algorithms used in this work have a false-positive rate of around 10%, meaning that some detected apnoeas may not have been true respiratory pauses. Although steps were taken to reduce artefactual detections, including screening approaches and machine learning-based processing, measurement error is likely to have limited predictive accuracy. This is a common challenge in neonatal

physiological monitoring. Future work should focus on improving apnoea detection.

Fourth, the analyses of apnoea-related instability were limited to central and mixed apnoeas only, as obstructive apnoeas could not be identified without nasal airflow monitoring. This is an important limitation because the systematic review suggested that apnoea type may influence physiological response. Future studies incorporating airflow measurements or other approaches to apnoea classification will be needed to determine whether personalised risk models should also account for apnoea types. In addition, the effects of clustered apnoeas were not explicitly investigated. Although the number of apnoeas within the preceding five minutes and the time since the previous apnoea were incorporated into the predictive models, this study did not examine whether clusters of apnoeas have cumulative physiological effects. Future studies should therefore characterise the temporal dynamics of apnoea clustering and determine whether modelling the sequence and recovery between successive apnoeas improves prediction of clinically significant events.

Fifth, cohort composition may have introduced selection biases. The Oxford dataset contained relatively fewer recordings from the youngest and oldest PMA ranges. Very immature infants were less likely to meet inclusion criteria because they were often clinically unstable, mechanically ventilated, or exposed to exclusionary treatments, while healthier older infants were more likely to have been discharged or transferred to other hospitals. This imbalance and bias in the datasets likely affected both the modelling of apnoea-associated instability and the functional respiratory age analyses. Broader multicentre datasets spanning a wider range of maturities and clinical trajectories will be important for improving generalisability.

Sixth, although the prototype personalised apnoea threshold framework is promising, it remains an early-stage proof of concept. Its clinical utility will depend not only on predictive accuracy but also on how it performs within real monitoring environments. For example, infant-specific alarm thresholds must improve sensitivity to clinically important events without generating excessive false alarms or increasing alarm fatigue. They must also be interpretable and trustworthy to clinicians, and robust to missing data, artefact, and workflow constraints. Future work should therefore move beyond retrospective model performance and evaluate implementation questions such as human factors, clinician usability, prospective calibration, and effects on clinical decision making.

These limitations also point directly to several future research directions. One priority is the development of large, multicentre, prospective neonatal physiology datasets with standardised data collection and signal processing pipelines, as well as comprehensive longitudinal clinical metadata. Such datasets would not only enable robust external validation of predictive models but also allow investigation of how respiratory maturation varies across different neonatal populations, clinical practices, and disease states. Importantly, they would provide sufficient statistical power to study relatively uncommon but clinically important subgroups, including extremely preterm infants, infants requiring prolonged respiratory support, and those with significant neurological or cardiorespiratory comorbidities.

More fundamentally, future studies should move beyond analysing individual apnoeas as isolated events. The findings of this thesis suggest that clinically significant cardiorespiratory instability depends not only on the characteristics of a single apnoea but also on the physiological state in which that apnoea occurs. Future modelling approaches should therefore explic-

itly characterise the evolving physiological context surrounding apnoeas, including preceding respiratory events, recovery dynamics, baseline cardiorespiratory stability, and interactions between respiratory, cardiovascular, and neurological systems.

The concept of functional respiratory age also offers potential for further development. The modest regression performance observed in this thesis suggests that respiratory maturation in preterm infants is more heterogeneous than can be captured by PMA alone. Future work should therefore investigate alternative reference frameworks that combine chronological maturity with physiological and clinical indicators of development. Rather than attempting to reproduce PMA or compare respiratory maturity between infants, FRA may have greater clinical value as a tool for tracking changes in respiratory function within an individual infant over time. This approach could allow each infant to serve as their own reference, potentially reducing the influence of between-infant variability. The reduction in FRA observed following morphine administration demonstrates that the model can detect acute changes in respiratory physiology. Similarly, other clinical conditions, such as sepsis, may produce measurable and reproducible changes in FRA, potentially allowing earlier identification of physiological deterioration. Longitudinal studies are needed to investigate whether FRA can reliably track individual respiratory development and detect clinically meaningful deviations from an infant's usual respiratory state.

Ultimately, the greatest clinical value of the works build within and beyond this thesis would lie in supporting personalised neonatal care. Future translational studies should investigate whether we could continuously and dynamically update the prediction models to identify infants at increased risk of clinically significant apnoea, before any deterioration may occur. Such

systems could enable monitoring intensity, clinical interventions, and treatment decisions to be tailored to an individual infant's evolving physiological state rather than relying primarily on chronological age or fixed clinical thresholds.

6.3 Conclusions

This thesis shows that cardiorespiratory responses to apnoea in preterm infants are highly variable, clinically meaningful, and only partly captured by conventional duration-based definitions. Longer apnoeas are generally associated with higher chances of bradycardia and desaturation, but level of instability also depends strongly on infant-specific factors including age, sex, ventilation mode, recent apnoea history and baseline physiological state. These findings provide strong evidence against a uniform interpretation of apnoea across all preterm infants.

By combining evidence synthesis, large-scale observational analysis, and predictive modelling, this thesis demonstrates that cardiorespiratory instability related to apnoea can be quantified and predicted with promising accuracy. It also shows that respiratory signals derived from routine monitoring encode information about maturational state and acute physiological perturbation. This thesis advances a more individualised view of apnoea in preterm infants, and provides initial evidence that personalised risk models and infant-specific alarm thresholds may be feasible. Although further methodological development, external validation, and prospective clinical evaluation are still required, the results presented here lay the groundwork for a new generation of personalised monitoring tools in neonatal care.

Overall, this thesis supports the view that more personalised, physiology-informed approaches

to neonatal monitoring are both necessary and achievable. Better identification of infants at risk of apnoea-associated instability may enable earlier intervention, more appropriate alarm settings, and ultimately improved care for preterm infants.

Appendix A

Searching strategy for the systematic review

MEDLINE (Ovid)

Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® Coverage: 1946 to present

1	infant/ or infant, newborn/ or infant, low birth weight/ or infant, small for gestational age/ or infant, very low birth weight/ or infant, extremely low birth weight/ or infant, postmature/ or infant, premature/ or infant, extremely premature/ or (infant* or newborn* or new-born* or perinatal* or neonatal* or neo-natal* or post-natal* or post-natal* or baby or babies or neonate* or neo-nate*).ti,ab,kf.
2	apnea/ or breath holding/ or (apno* or apne* or breath hold* or periodic breath* or pause in breath* or cessation of breath* or AOP or respiratory pause* or intermittent breath* or intermittent respiration or inter breath interval* or stop breath*).ti,ab,kf.

3 heart rate/ or bradycardia/ or tachycardia/ or oxygen saturation/ or hypoxia/ or cerebral blood volume/ or brain ischemia/ or hypoxia-ischemia, brain/ or

(heart rate* or cardiac rate* or heartbeat* or pulse rate* or bradycardia or tachycardia or oxygen saturation or desat* or oxygen deprivation or lack of oxygen or oxygen starvation or hypoxia or hypoxemia or hypoxeamia or SpO2 or SaO2 or PaO2 or cerebral blood volume or CBV or cerebral blood flow or brain ischemia or cerebral ischeamia or brain blood flow or brain blood supply or brain hemodynamics or brain haemodynamics or brain perfusion or cerebral oxygenation).ti,ab,kf.

4 1 and 2 and 3

5 exp animals/ not humans/

6 4 not 5

—

Embase (Ovid)

Coverage: 1974 to present

1 infant/ or baby/ or high risk infant/ or hospitalized infant/ or newborn/ or prematurity/ or

(infant* or newborn* or new-born* or perinatal* or neonatal* or neo-natal* or post-natal* or post-natal* or baby or babies or neonate* or neo-nate*).ti,ab,kf.

2 *apnea/ or *apnea attack/ or *newborn apnea/ or *newborn apnea attack/ or *breath holding/ or

(apno* or apne* or ”breath hold*” or ”periodic breath*” or ”pause in breath*” or ”cessation of breath*” or AOP or ”respiratory pause*” or ”intermittent breath*” or ”intermittent respiration” or ”inter breath interval*” or ”stop breath*”).ti,ab,kf.

3 heart rate/ or bradycardia/ or tachycardia/ or oxygen saturation/ or hypoxia/ or hypoxemia/ or newborn hypoxia/ or brain blood volume/ or brain ischemia/ or

(”heart rate*” or ”cardiac rate*” or heartbeat* or ”pulse rate*” or bradycardia or tachycardia or ”oxygen saturation” or desat* or ”oxygen deprivation” or ”lack of oxygen” or ”oxygen starvation” or hypoxia or hypoxemia or hypoxeamia or SpO2 or SaO2 or PaO2 or ”cerebral blood volume” or CBV or ”cerebral blood flow” or ”brain ischemia” or ”cerebral ischeamia” or ”brain blood flow” or ”brain blood supply” or ”brain hemodynamics” or ”brain haemodynamics” or ”brain perfusion” or ”cerebral oxygenation”).ti,ab,kf.

4 1 and 2 and 3

5 exp animal/ not human/

6	4 not 5
7	Systematic review.ti,ab. not (trial or study).ti.
8	(review.ab. and review.pt.) not trial.ti.
9	we searched.ab. and (review.ti. or review.pt.)
10	update review.ab.
11	(databases adj4 searched).ab.
12	(conference abstract or conference paper or ”conference review” or editorial).pt.
13	7 or 8 or 9 or 10 or 11 or 12
14	6 not 13

—

PsycINFO (Ovid)

Coverage: 1806 to present

1	premature birth/ or

(infant* or newborn* or new-born* or perinatal* or neonatal* or neo-natal* or post-natal* or post-natal* or baby or babies or neonate* or neo-nate*).ti,ab,id.

2 apnea/ or

(apno* or apne* or ”breath hold*” or ”periodic breath*” or ”pause in breath*” or ”cessation of breath*” or AOP or ”respiratory pause*” or ”intermittent breath*” or ”intermittent respiration” or ”inter breath interval*” or ”stop breath*”).ti,ab,id.

3 heart rate/ or bradycardia/ or tachycardia/ or anoxia/ or cerebral blood flow/ or cerebral ischemia/ or

(”heart rate*” or ”cardiac rate*” or heartbeat* or ”pulse rate*” or bradycardia or tachycardia or ”oxygen saturation” or desat* or ”oxygen deprivation” or ”lack of oxygen” or ”oxygen starvation” or hypoxia or hypoxemia or hypoxeamia or SpO2 or SaO2 or PaO2 or ”cerebral blood volume” or CBV or ”cerebral blood flow” or ”brain ischemia” or ”cerebral ischeamia” or ”brain blood flow” or ”brain blood supply” or ”brain hemodynamics” or ”brain haemodynamics” or ”brain perfusion” or ”cerebral oxygenation”).ti,ab,id.

4 1 and 2 and 3

—

Cochrane Library

#1 [mh ^infant] OR [mh ^"infant, newborn"] OR [mh ^"infant, low birth weight"] OR [mh ^"infant, small for gestational age"] OR [mh ^"infant, very low birth weight"] OR [mh ^"infant, extremely low birth weight"] OR [mh ^"infant, postmature"] OR [mh ^"infant, premature"] OR [mh ^"infant, extremely premature"] OR

(infant*:ti,ab,kw OR newborn*:ti,ab,kw OR new-born*:ti,ab,kw OR perinatal*:ti,ab,kw OR neonatal*:ti,ab,kw OR neo-natal*:ti,ab,kw OR postnatal*:ti,ab,kw OR post-natal*:ti,ab,kw OR baby:ti,ab,kw OR babies:ti,ab,kw OR neonate*:ti,ab,kw OR neo-nate*:ti,ab,kw)

#2 [mh ^apnea] OR [mh ^"breath holding"] OR

(apno*:ti,ab,kw OR apne*:ti,ab,kw OR ("breath" NEXT hold*):ti,ab,kw OR ("periodic" NEXT breath*):ti,ab,kw OR ("pause in" NEXT breath*):ti,ab,kw OR ("cessation of" NEXT breath*):ti,ab,kw OR AOP:ti,ab,kw OR ("respiratory" NEXT pause*):ti,ab,kw OR ("intermittent" NEXT breath*):ti,ab,kw OR "intermittent respiration":ti,ab,kw OR ("inter breath" NEXT interval*):ti,ab,kw OR ("stop" NEXT breath*):ti,ab,kw)

#3 [mh ^"heart rate"] OR [mh ^bradycardia] OR [mh ^tachycardia] OR [mh ^"oxygen saturation"] OR [mh ^hypoxia] OR [mh ^"cerebral blood volume"] OR [mh ^"brain ischemia"] OR [mh ^"hypoxia-ischemia, brain"] OR

(("heart" NEXT rate*):ti,ab,kw OR ("cardiac" NEXT rate*):ti,ab,kw OR heartbeat*:ti,ab,kw OR ("pulse" NEXT rate*):ti,ab,kw OR bradycardia:ti,ab,kw OR tachycardia:ti,ab,kw OR "oxygen saturation":ti,ab,kw OR desat*:ti,ab,kw OR "oxygen deprivation":ti,ab,kw OR "lack of oxygen":ti,ab,kw OR "oxygen starvation":ti,ab,kw OR hypoxia:ti,ab,kw OR hypoxemia:ti,ab,kw OR hypoxeamia:ti,ab,kw OR SpO2:ti,ab,kw OR SaO2:ti,ab,kw OR PaO2:ti,ab,kw OR "cerebral blood volume":ti,ab,kw OR CBV:ti,ab,kw OR "cerebral blood flow":ti,ab,kw OR "brain ischemia":ti,ab,kw OR "cerebral ischeamia":ti,ab,kw OR "brain blood flow":ti,ab,kw OR "brain blood supply":ti,ab,kw OR "brain hemodynamics":ti,ab,kw OR "brain haemodynamics":ti,ab,kw OR "brain perfusion":ti,ab,kw OR "cerebral oxygenation":ti,ab,kw)

#4 #1 and #2 and #3

Appendix B

Respiratory features extracted from IP and IBI signals

Feature	Brief Description
Respiratory Event Measures	
Respiratory rate	Number of breaths per minute.
Apnoea rate (5s)	Frequency of IBI $\geq$ 5 seconds per hour.
Apnoea rate (15s)	Frequency of IBI $\geq$ 15 seconds per hour.
Basic Statistical Features of IBI	
Mean IBI	Average IBI duration.
Median IBI	Median IBI duration.
IBI 5th Percentile	5th percentile of IBI distribution.
IBI 95th Percentile	95th percentile of IBI distribution.
Minimum IBI	Shortest observed inter-breath interval.
Maximum IBI	Longest observed inter-breath interval.
Range	Difference between maximum and minimum IBI.
Area	Area under the IBI distribution curve.
Standard deviation	Dispersion of IBI values around the mean.
Skewness	Asymmetry of the IBI distribution.
Kurtosis	Tailedness (peakedness) of the IBI distribution.
Spectral Measures	
Band power 1	Power within predefined low-frequency respiratory band.
Band power 2	Power within predefined high-frequency respiratory band.
LF normalised	Low-frequency power of the IBI signal, normalised by standard deviation of the IBI.
HF normalised	High-frequency power of the IBI signal, normalised by standard deviation of the IBI.
HF:LF ratio	Ratio of high- to low-frequency power.
Frequency width at half maximum	Spectral bandwidth measured at half of peak power.
Frequency width	Total width of dominant spectral peak.
Instantaneous autospectrum	Time-resolved power spectral density estimate.
Entropy and Complexity Measures	

Feature	Brief Description
Sample entropy	Regularity measure quantifying unpredictability of the IBI time series.
Short-term IBI variability	Variability of IBI over short time scales.
Long-term IBI variability	Variability of IBI over longer time scales.
Fractal dimension	Measure of signal geometric complexity.
IBI MSE scale 1	Multiscale entropy at scale 1.
IBI MSE scale 2	Multiscale entropy at scale 2.
IBI MSE scale 3	Multiscale entropy at scale 3.
IBI MSE scale 4	Multiscale entropy at scale 4.
Median spectral entropy	Median entropy across spectral distribution.
Max PSE	Maximum power spectral entropy value.
Shannon entropy	Information-theoretic measure of signal randomness.
Renyi entropy	Generalised entropy measure sensitive to distribution tails.
Max permutation entropy	Maximum ordinal pattern entropy across embedding dimensions.
Approximate entropy	Measure of signal regularity and pattern predictability.
Fuzzy entropy	Entropy measure robust to noise via fuzzy membership functions.
Conditional entropy	Residual uncertainty given prior signal state.
Lempel–Ziv complexity	A measure of how many different patterns appear in the IBI sequence. Higher values indicate more variation.
Power Law / Fractal Scaling Measures	
BA average	The mean area of each burst, calculated by adding the IBIs for every burst in a recording or a selected window.
Burst number (BN)	The total number of detected bursts.
BN $\times$ BA average	The total number of detected bursts times the mean area of each burst.
BWHc	The burst width-height intercept, which is the intercept of a simple linear regression line for the height and width of each burst.
BWHm	The burst width-height gradient, which is the gradient of a simple linear regression line for the height and width of each burst.
Power law (alpha)	Scaling exponent describing long-range correlations.
PLBSKc	Power-law-based skewness coefficient.
PLBSKm	Power-law-based kurtosis coefficient.
PLXmin	Minimum threshold used for power-law fitting.
Non-linear Measures	
Median SNLEO	Median value of the smoothed nonlinear energy operator.
Hurst exponent	Estimate of long-range temporal correlations (persistence).
BreathMetrics-derived Breath and Volume Measures	
Median inhale duration (s)	Median duration of inspiratory phase across breaths (seconds).
Mean inhale duration (s)	Mean duration of inspiratory phase across breaths (seconds).
Median exhale duration (s)	Median duration of expiratory phase across breaths (seconds).

Feature	Brief Description
Mean exhale duration (s)	Mean duration of expiratory phase across breaths (seconds).
Median I:E ratio	Median breath-wise ratio of inspiratory to expiratory duration (I:E).
Mean I:E ratio	Mean breath-wise ratio of inspiratory to expiratory duration (I:E).
Median inhale volume	Median inspiratory volume estimate across breaths (volume surrogate).
Mean inhale volume	Mean inspiratory volume estimate across breaths (volume surrogate).
Median exhale volume	Median expiratory volume estimate across breaths (volume surrogate).
Mean exhale volume	Mean expiratory volume estimate across breaths (volume surrogate).
Median tidal volume (surrogate)	Median of tidal-volume surrogate computed from paired inhale/exhale volumes.
Mean tidal volume (surrogate)	Mean of tidal-volume surrogate computed from paired inhale/exhale volumes.
Median peak inspiratory flow (PIF)	Median of the maximum inspiratory flow values measured per breath.
Mean peak inspiratory flow (PIF)	Mean of the maximum inspiratory flow values measured per breath.
Median peak expiratory flow (PEF)	Median of the peak expiratory (trough) flow values measured per breath.
Mean peak expiratory flow (PEF)	Mean of the peak expiratory (trough) flow values measured per breath.

Bibliography

[1] Y. Chen et al., "Apnoea duration and changes in cardiorespiratory and cerebrovascular responses in preterm neonates: A systematic review and meta-analysis," *Pediatric Research*, pp. 1–12, 2025.

[2] Y. Chen et al., "Risk of apnoea-related cardiorespiratory instability in preterm infants is modulated by clinical, demographic and dynamic indicators," *medRxiv*, 2026. DOI: 10.64898/2026.05.13.26353101 eprint: https://www.medrxiv.org/content/early/2026/05/17/2026.05.13.26353101.full.pdf. [Online]. Available: https://www.medrxiv.org/content/early/2026/05/17/2026.05.13.26353101

[3] E. O. Ohuma et al., "National, regional, and global estimates of preterm birth in 2020, with trends from 2010: A systematic analysis," *The Lancet*, vol. 402, pp. 1261–1271, 10409 Oct. 2023, ISSN: 01406736. DOI: 10.1016/S0140-6736(23)00878-4

[4] W. M. Callaghan, M. F. MacDorman, S. A. Rasmussen, C. Qin, and E. M. Lackritz, "The contribution of preterm birth to infant mortality rates in the united states," *Pediatrics*, vol. 118, no. 4, pp. 1566–1573, 2006.

[5] D. Henderson-Smart, "The effect of gestational age on the incidence and duration of recurrent apnoea in newborn babies," *Journal of Paediatrics and Child Health*, vol. 17, pp. 273–276, 4 Dec. 1981, ISSN: 1034-4810. DOI: 10.1111/j.1440-1754.1981.tb01957.x

[6] W. J. Daily, M. Klaus, and H. B. Meyer, "Apnea in premature infants: Monitoring, incidence, heart rate changes, and an effect of environmental temperature.," *Pediatrics*, vol. 43, pp. 510–518, 4 1969.

[7] C. J. Upton, A. D. Milner, and G. M. Stokes, "Apnoea, bradycardia, and oxygen saturation in preterm infants.," *Archives of disease in childhood*, vol. 66, pp. 381–385, 4 Spec No 1991.

[8] G. Pichler, B. Urlesberger, and W. Muller, "Impact of bradycardia on cerebral oxygenation and cerebral blood volume during apnoea in preterm infants.," *Physiological measurement*, vol. 24, pp. 671–680, 3 2003.

[9] R. S. C. Horne, S. Sun, S. R. Yiallourou, K. L. Fyfe, A. Odoi, and F. Y. Wong, "Comparison of the longitudinal effects of persistent periodic breathing and apnoea on cerebral oxygenation in term- and preterm-born infants," *The Journal of Physiology*, vol. 596, pp. 6021–6031, 23 Dec. 2018, ISSN: 0022-3751. DOI: 10.1113/JP275686

[10] B. Araz-Ersan et al., "Epidemiological analysis of retinopathy of prematurity in a referral centre in turkey," *British Journal of Ophthalmology*, vol. 97, pp. 15–17, 1 Jan. 2013, ISSN: 0007-1161. DOI: 10.1136/bjophthalmol-2011-301411

[11] A. Janvier, M. Khairy, A. Kokkotis, C. Cormier, D. Messmer, and K. J. Barrington, "Apnea is associated with neurodevelopmental impairment in very low birth weight infants," *Journal of Perinatology*, vol. 24, pp. 763–768, 12 Dec. 2004, ISSN: 0743-8346. DOI: 10.1038/sj.jp.7211182

[12] C. S. Zandvoort et al., "Apnoea suppresses brain activity in infants," *Imaging Neuroscience*, vol. 2, pp. 1–14, Jul. 2024, ISSN: 2837-6056. DOI: 10.1162/imag_a_00236

[13] E. C. Eichenwald et al., "Apnea of prematurity," *Pediatrics*, vol. 137, no. 1, 2016.

[14] J. V. Aranda, W. Gorman, H. Bergsteinsson, and T. Gunn, "Efficacy of caffeine in treatment of apnea in the low-birth-weight infant," *The Journal of Pediatrics*, vol. 90, pp. 467–472, 3 Mar. 1977, ISSN: 00223476. DOI: 10.1016/S0022-3476(77)80718-X

[15] E. A. Oliphant, C. J. McKinlay, D. McNamara, A. Cavadino, and J. M. Alsweiler, "Caffeine to prevent intermittent hypoxaemia in late preterm infants: Randomised controlled dosage trial," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 108, pp. 106–113, 2 Mar. 2023, ISSN: 1359-2998. DOI: 10.1136/archdischild-2022-324010

[16] J. E. Scanlon, K. C. Chin, M. E. Morgan, G. M. Durbin, K. A. Hale, and S. S. Brown, "Caffeine or theophylline for neonatal apnoea?" *Archives of Disease in Childhood*, vol. 67, pp. 425–428, 4 Spec No Apr. 1992, ISSN: 0003-9888. DOI: 10.1136/adc.67.4_Spec_No.425

[17] J. B. West, *Respiratory physiology: the essentials*. Lippincott Williams & Wilkins, 2012.

[18] E. Nattie, "Co2, brainstem chemoreceptors and breathing," *Progress in neurobiology*, vol. 59, no. 4, pp. 299–331, 1999.

[19] E. Nattie and A. Li, "Central chemoreceptors: Locations and functions," *Comprehensive physiology*, vol. 2, no. 1, pp. 221–254, 2012.

[20] P. G. Guyenet and D. A. Bayliss, "Neural control of breathing and co2 homeostasis," *Neuron*, vol. 87, no. 5, pp. 946–961, 2015.

[21] W. F. Boron and E. L. Boulpaep, *Medical Physiology E-Book: Medical Physiology E-Book*. Elsevier Health Sciences, 2016.

[22] N. R. Prabhakar and Y.-J. Peng, "Peripheral chemoreceptors in health and disease," *Journal of applied physiology*, vol. 96, no. 1, pp. 359–366, 2004.

[23] J. E. Hall, *Guyton and hall textbook of medical physiology, Jordanian Edition E-Book*. Elsevier Health Sciences, 2016.

[24] J. Nagels, F. Landser, L. Van der Linden, J. Clement, and K. Van de Woestijne, "Mechanical properties of lungs and chest wall during spontaneous breathing," *Journal of Applied Physiology*, vol. 49, no. 3, pp. 408–416, 1980.

[25] M. Williamson, R. Poorun, and C. Hartley, "Apnoea of prematurity and neurodevelopmental outcomes: Current understanding and future prospects for research," *Frontiers in Pediatrics*, vol. 9, p. 755 677, 2021.

[26] P. H. Burri, "Structural aspects of postnatal lung development–alveolar formation and growth," *Neonatology*, vol. 89, no. 4, pp. 313–322, 2006.

[27] I. Copland and M. Post, "Lung development and fetal lung growth," *Paediatric respiratory reviews*, vol. 5, S259–S264, 2004.

[28] A. A. Hislop, J. S. Wigglesworth, and R. Desai, "Alveolar development in the human fetus and infant," *Early human development*, vol. 13, no. 1, pp. 1–11, 1986.

[29] S. Rehman and D. Bacha, "Embryology, pulmonary," 2019.

[30] J. Stocks, A. Hislop, and S. Sonnappa, "Early lung development: Lifelong effect on respiratory health and disease," *The lancet Respiratory medicine*, vol. 1, no. 9, pp. 728–742, 2013.

[31] O. Hjalmarson and K. Sandberg, "Abnormal lung function in healthy preterm infants," *American journal of respiratory and critical care medicine*, vol. 165, no. 1, pp. 83–87, 2002.

[32] T. Gerhardt, L. Reifenberg, D. Hehre, R. Feller, and E. Bancalari, "Functional residual capacity in normal neonates and children up to 5 years of age determined by a n2 washout method," *Pediatric research*, vol. 20, no. 7, pp. 668–671, 1986.

[33] A. A. Hislop, "Airway and blood vessel interaction during lung development," *Journal of anatomy*, vol. 201, no. 4, pp. 325–334, 2002.

[34] D. Holditch-Davis, M. Scher, and T. Schwartz, "Respiratory development in preterm infants," *Journal of Perinatology*, vol. 24, no. 10, pp. 631–639, 2004.

[35] A. B. te Pas, C. Wong, C. O. F. Kamlin, J. A. Dawson, C. J. Morley, and P. G. Davis, "Breathing patterns in preterm and term infants immediately after birth," *Pediatric research*, vol. 65, no. 3, pp. 352–356, 2009.

[36] R. A. Darnall, R. L. Ariagno, and H. C. Kinney, "The late preterm infant and the control of breathing, sleep, and brainstem development: A review," *Clinics in perinatology*, vol. 33, no. 4, pp. 883–914, 2006.

[37] G. P. Heldt, "Development of stability of the respiratory system in preterm infants," *Journal of Applied Physiology*, vol. 65, no. 1, pp. 441–444, 1988.

[38] C. F. Poets et al., "Association between intermittent hypoxemia or bradycardia and late death or disability in extremely preterm infants," *JAMA*, vol. 314, p. 595, 6 Aug. 2015, ISSN: 0098-7484. DOI: 10.1001/jama.2015.8841

[39] M. Miller and R. Martin, "Apnea of prematurity.," *Clinics in perinatology*, vol. 19, no. 4, pp. 789–808, 1992.

[40] A. Milner, A. Boon, R. Saunders, and I. Hopkin, "Upper airways obstruction and apnoea in preterm babies.," *Archives of disease in childhood*, vol. 55, no. 1, pp. 22–25, 1980.

[41] D. Henderson-Smart, "The effect of gestational age on the incidence and duration of recurrent apnoea in newborn babies," *Journal of Paediatrics and Child Health*, vol. 17, no. 4, pp. 273–276, 1981.

[42] A. J. Jobe, "The new bpd: An arrest of lung development," *Pediatric research*, vol. 46, no. 6, pp. 641–641, 1999.

[43] J. R. Williamson, D. W. Bliss, and D. Paydarfar, "Forecasting respiratory collapse: Theory and practice for averting life-threatening infant apneas," *Respiratory physiology & neurobiology*, vol. 189, no. 2, pp. 223–231, 2013.

[44] C. F. Poets, V. A. Stebbens, M. P. Samuels, and D. P. Southall, "The relationship between bradycardia, apnea, and hypoxemia in preterm infants.," *Pediatric research*, vol. 34, pp. 144–147, 2 1993.

[45] C. F. Poets and D. P. Southall, "Patterns of oxygenation during periodic breathing in preterm infants.," *Early human development*, vol. 26, pp. 1–12, 1 1991.

[46] D. J. Henderson-Smart and A. G. De Paoli, "Methylxanthine treatment for apnoea in preterm infants," *Cochrane Database of Systematic Reviews*, no. 12, 2010.

[47] F. Pillekamp, C. Hermann, T. Keller, A. Von Gontard, A. Kribs, and B. Roth, "Factors influencing apnea and bradycardia of prematurity–implications for neurodevelopment," *Neonatology*, vol. 91, no. 3, pp. 155–161, 2007.

[48] J. M. Di Fiore et al., "A higher incidence of intermittent hypoxemic episodes is associated with severe retinopathy of prematurity," *The Journal of pediatrics*, vol. 157, no. 1, pp. 69–73, 2010.

[49] J. P. Mortola, "How to breathe? respiratory mechanics and breathing pattern," *Respiratory physiology & neurobiology*, vol. 261, pp. 48–54, 2019.

[50] A. L. Goldberger, L. A. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," *Proceedings of the national academy of sciences*, vol. 99, no. suppl_1, pp. 2466–2472, 2002.

[51] F. A. Selig, E. R. Tonolli, É. V. C. M. d. Silva, and M. F. d. Godoy, "Heart rate variability in preterm and term neonates," *Arquivos brasileiros de cardiologia*, vol. 96, pp. 443–449, 2011.

[52] D. E. Elder, A. J. Campbell, P. D. Larsen, and D. Galletly, "Respiratory variability in preterm and term infants: Effect of sleep state, position and age," *Respiratory physiology & neurobiology*, vol. 175, no. 2, pp. 234–238, 2011.

[53] K. Jost, A. N. Datta, U. P. Frey, B. Suki, and S. M. Schulzke, "Heart rate fluctuation after birth predicts subsequent cardiorespiratory stability in preterm infants," *Pediatric research*, vol. 86, no. 3, pp. 348–354, 2019.

[54] K. D. Fairchild, "Predictive monitoring for early detection of sepsis in neonatal icu patients," *Current opinion in pediatrics*, vol. 25, no. 2, pp. 172–179, 2013.

[55] M. W. Kuzniewicz et al., "A quantitative, risk-based approach to the management of neonatal early-onset sepsis," *JAMA pediatrics*, vol. 171, no. 4, pp. 365–371, 2017.

[56] E. A. Ashley, "Towards precision medicine," *Nature Reviews Genetics*, vol. 17, no. 9, pp. 507–522, 2016.

[57] E. D. Ozturk and C. O. Tan, "Human cerebrovascular function in health and disease: Insights from integrative approaches," *Journal of physiological anthropology*, vol. 37, no. 1, p. 4, 2018.

[58] M. Tsuji et al., "Cerebral intravascular oxygenation correlates with mean arterial pressure in critically ill premature infants," *Pediatrics*, vol. 106, no. 4, pp. 625–632, 2000.

[59] C. J. Rhee, C. S. da Costa, T. Austin, K. M. Brady, M. Czosnyka, and J. K. Lee, "Neonatal cerebrovascular autoregulation," *Pediatric research*, vol. 84, no. 5, pp. 602–610, 2018.

[60] O. Paulson, S. Strandgaard, and L. Edvinsson, "Cerebral autoregulation," *Cerebrovascular and brain metabolism reviews*, vol. 2, no. 2, pp. 161–192, 1990.

[61] J. S. Soul et al., "Fluctuating pressure-passivity is common in the cerebral circulation of sick premature infants," *Pediatric research*, vol. 61, no. 4, pp. 467–473, 2007.

[62] S. A. Back and S. P. Miller, "Brain injury in premature neonates: A primary cerebral dysmaturation disorder?" *Annals of neurology*, vol. 75, no. 4, pp. 469–486, 2014.

[63] G. Greisen, "Autoregulation of cerebral blood flow in newborn babies," *Early human development*, vol. 81, no. 5, pp. 423–428, 2005.

[64] O. Pryds, "Control of cerebral circulation in the high-risk neonate," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 30, no. 3, pp. 321–329, 1991.

[65] J. J. Volpe, "Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances," *The Lancet Neurology*, vol. 8, no. 1, pp. 110–124, 2009. DOI: 10.1016/S1474-4422(08)70294-1

[66] G. Greisen and K. Børch, "White matter injury in the preterm neonate: The role of perfusion," *Developmental neuroscience*, vol. 23, no. 3, pp. 209–212, 2001.

[67] T. Alderliesten, P. M. Lemmers, J. J. Smarius, R. E. van de Vosse, W. Baerts, and F. van Bel, "Cerebral oxygenation, extraction, and autoregulation in very preterm infants who develop peri-intraventricular hemorrhage," *The Journal of pediatrics*, vol. 162, no. 4, pp. 698–704, 2013.

[68] B. G. Sood, K. McLaughlin, and J. Cortez, "Near-infrared spectroscopy: Applications in neonates," in *Seminars in Fetal and Neonatal Medicine*, Elsevier, vol. 20, 2015, pp. 164–172.

[69] A. Pellicer and M. del Carmen Bravo, "Near-infrared spectroscopy: A methodology-focused review," in *Seminars in fetal and neonatal medicine*, Elsevier, vol. 16, 2011, pp. 42–49.

[70] C. F. Poets and D. P. Southall, "Noninvasive monitoring of oxygenation in infants and children: Practical considerations and areas of concern," *Pediatrics*, vol. 93, no. 5, pp. 737–746, 1994.

[71] H. Lee et al., "A new algorithm for detecting central apnea in neonates," *Physiological measurement*, vol. 33, no. 1, p. 1, 2011.

[72] M. Cvach, "Monitor alarm fatigue: An integrative review," *Biomedical instrumentation & technology*, vol. 46, no. 4, pp. 268–277, 2012.

[73] J. M. Di Fiore et al., "Cardiorespiratory events in preterm infants referred for apnea monitoring studies," *Pediatrics*, vol. 108, no. 6, pp. 1304–1308, 2001.

[74] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, "The hallmarks of aging," *Cell*, vol. 153, no. 6, pp. 1194–1217, 2013.

[75] J. Jylhävä, N. L. Pedersen, and S. Hägg, "Biological age predictors," *EBioMedicine*, vol. 21, pp. 29–36, 2017.

[76] D. W. Belsky et al., "Quantification of biological aging in young adults," *Proceedings of the National Academy of Sciences*, vol. 112, no. 30, E4104–E4110, 2015.

[77] S. Horvath, "Dna methylation age of human tissues and cell types," *Genome biology*, vol. 14, no. 10, p. 3156, 2013.

[78] R. B. D'Agostino Sr et al., "General cardiovascular risk profile for use in primary care: The framingham heart study," *Circulation*, vol. 117, no. 6, pp. 743–753, 2008.

[79] A. A. Lopez-Gonzalez et al., "Effectiveness of the heart age tool for improving modifiable cardiovascular risk factors in a southern european population: A randomized trial," *European journal of preventive cardiology*, vol. 22, no. 3, pp. 389–396, 2015.

[80] Z. I. Attia et al., "Age and sex estimation using artificial intelligence from standard 12-lead ecgs," *Circulation: Arrhythmia and Electrophysiology*, vol. 12, no. 9, e007284, 2019.

[81] K. Umetani, D. H. Singer, R. McCraty, and M. Atkinson, "Twenty-four hour time domain heart rate variability and heart rate: Relations to age and gender over nine decades," *Journal of the American College of Cardiology*, vol. 31, no. 3, pp. 593–601, 1998.

[82] R. Poplin et al., "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature biomedical engineering*, vol. 2, no. 3, pp. 158–164, 2018.

[83] J. Dubois, G. Dehaene-Lambertz, S. Kulikova, C. Poupon, P. S. Hüppi, and L. Hertz-Pannier, "The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants," *neuroscience*, vol. 276, pp. 48–71, 2014.

[84] M. H. Johnson, "Functional brain development in humans," *Nature Reviews Neuroscience*, vol. 2, no. 7, pp. 475–483, 2001.

[85] J. H. Cole and K. Franke, "Predicting age using neuroimaging: Innovative brain ageing biomarkers," *Trends in neurosciences*, vol. 40, no. 12, pp. 681–690, 2017.

[86] G. Ball et al., "Multimodal image analysis of clinical influences on preterm brain development," *Annals of neurology*, vol. 82, no. 2, pp. 233–246, 2017.

[87] F. Usman et al., "Brain activity as a candidate biomarker for personalised caffeine treatment in premature neonates," *medRxiv*, 2025.

[88] T. T. Brown et al., "Neuroanatomical assessment of biological maturity," *Current biology*, vol. 22, no. 18, pp. 1693–1698, 2012.

[89] S. Vanhatalo and K. Kaila, "Development of neonatal eeg activity: From phenomenology to physiology," in *Seminars in Fetal and Neonatal Medicine*, Elsevier, vol. 11, 2006, pp. 471–478.

[90] K. Lehnertz et al., "Seizure prediction by nonlinear eeg analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 1, pp. 57–63, 2003.

[91] A. C. Kevat, D. V. Bullen, P. G. Davis, and C. O. F. Kamlin, "A systematic review of novel technology for monitoring infant and newborn heart rate," *Acta Paediatrica*, vol. 106, no. 5, pp. 710–720, 2017.

[92] T. Kato, A. Ueno, S. Kataoka, H. Hoshino, and Y. Ishiyama, "An application of capacitive electrode for detecting electrocardiogram of neonates and infants," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2006, pp. 916–919.

[93] Y. Ren, D. Wu, Y. Tong, A. López-DeFede, and S. Gareau, "Issue of data imbalance on low birthweight baby outcomes prediction and associated risk factors identification: Establishment of benchmarking key machine learning models with data rebalancing strategies," *Journal of medical Internet research*, vol. 25, e44081, 2023.

[94] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC bioinformatics*, vol. 7, no. 1, p. 91, 2006.

[95] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[96] A. Esteva et al., "A guide to deep learning in healthcare," *Nature medicine*, vol. 25, no. 1, pp. 24–29, 2019.

[97] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," *Jama*, vol. 319, no. 13, pp. 1317–1318, 2018.

[98] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[99] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep ehr: A survey of recent advances in deep learning techniques for electronic health record (ehr) analysis," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1589–1604, 2017.

[100] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data," *PloS one*, vol. 8, no. 6, e66341, 2013.

[101] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England journal of medicine*, vol. 375, no. 13, p. 1216, 2016.

[102] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[103] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Scientific reports*, vol. 6, no. 1, p. 26 094, 2016.

[104] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[105] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," in *Healthcare*, MDPI, vol. 11, 2023, p. 1808.

[106] J. S. Calvert et al., "A computational approach to early sepsis detection," *Computers in biology and medicine*, vol. 74, pp. 69–73, 2016.

[107] M. Podda, D. Bacciu, A. Micheli, R. Bellù, G. Placidi, and L. Gagliardi, "A machine learning approach to estimating preterm infants survival: Development of the preterm infants survival assessment (pisa) predictor," *Scientific reports*, vol. 8, no. 1, p. 13 743, 2018.

[108] C. Hartley et al., "Analgesic efficacy and safety of morphine in the procedural pain in premature infants (poppi) study: Randomised placebo-controlled trial," *The Lancet*, vol. 392, no. 10164, pp. 2595–2605, 2018.

[109] T. Adjei et al., "New method to measure interbreath intervals in infants for the assessment of apnoea and respiration," *BMJ Open Respiratory Research*, vol. 8, no. 1, 2021.

[110] H. Vyas, A. D. Milner, and I. E. Hopkin, "Relationship between apnoea and bradycardia in preterm infants.," *Acta paediatrica Scandinavica*, vol. 70, pp. 785–790, 6 1981.

[111] D. J. Henderson-Smart, M. C. Butcher-Puech, and D. A. Edwards, "Incidence and mechanism of bradycardia during apnoea in preterm infants.," *Archives of disease in childhood*, vol. 61, pp. 227–232, 3 1986.

[112] S. A. Sands, B. A. Edwards, V. J. Kelly, M. R. Davidson, M. H. Wilkinson, and P. J. Berger, "A model analysis of arterial oxygen desaturation during apnea in preterm infants," *PLoS computational biology*, vol. 5, no. 12, e1000588, 2009.

[113] M. J. Page et al., "The prisma 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, n71, Mar. 2021, ISSN: 1756-1833. DOI: 10.1136/bmj.n71

[114] N. R. Haddaway, M. J. Grainger, and C. T. Gray, "Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching," *Research Synthesis Methods*, vol. 13, pp. 533–545, 4 Jul. 2022, ISSN: 1759-2879. DOI: 10.1002/jrsm.1563

[115] A.
bibinitperiod D. for impact (DDI) Data, *World health statistics 2024: monitoring health for the SDGs, sustainable development goals*. 2024.

[116] T. H. Barker et al., "Revising the jbi quantitative critical appraisal tools to improve their applicability: An overview of methods and the development process," *JBI evidence synthesis*, vol. 21, pp. 478–493, 3 Nov. 2023, ISSN: 26898381. DOI: 10.11124/JBIES-22-00125

[117] Z. Munn et al., "Assessing the risk of bias of quantitative analytical studies: Introducing the vision for critical appraisal within jbi systematic reviews," *JBI evidence synthesis*, vol. 21, pp. 467–471, 3 Dec. 2023, ISSN: 26898381. DOI: 10.11124/JBIES-22-00224

[118] M. L. Spear, J. L. Stefano, and A. R. Spitzer, "Prolonged apnea and oxyhemoglobin desaturation in asymptomatic premature infants.," *Pediatric pulmonology*, vol. 13, pp. 151–154, 3 1992.

[119] A. Y. Trapeznikova, N. A. Petrova, G. I. Obraztsova, E. V. Boytsova, Y. V. Sviryaev, and D. O. Ivanov, "Episodes of apnea and periodic breathing in premature infants with bpd-associated pulmonary hypertension," *Rossiyskiy Vestnik Perinatologii i Pediatrii (Russian Bulletin of Perinatology and Pediatrics)*, vol. 67, pp. 94–99, 2 Jun. 2022, ISSN: 2500-2228. DOI: 10.21508/1027-4065-2022-67-2-94-99

[120] B. Urlesberger, A. Kaspirek, G. Pichler, and W. Muller, "Apnoea of prematurity and changes in cerebral oxygenation and cerebral blood volume.," *Neuropediatrics*, vol. 30, pp. 29–33, 1 1999.

[121] M. Gabriel and M. Albani, "Cardiac slowing and respiratory arrest in preterm infants.," *European journal of pediatrics*, vol. 122, pp. 257–261, 4 1976.

[122] G. M. Fenichel, B. J. Olson, and J. E. Fitzpatrick, "Heart rate changes in convulsive and nonconvulsive neonatal apnea.," *Annals of neurology*, vol. 7, pp. 577–582, 6 1980.

[123] N. N. Finer, K. J. Barrington, B. J. Hayes, and A. Hugh, "Obstructive, mixed, and central apnea in the neonate: Physiologic correlates.," *The Journal of pediatrics*, vol. 121, pp. 943–950, 6 1992.

[124] T. Carbone, L. C. Marrero, J. Weiss, M. Hiatt, and T. Hegyi, "Heart rate and oxygen saturation correlates of infant apnea.," *Journal of perinatology : official journal of the California Perinatal Association*, vol. 19, pp. 44–47, 1 1999.

[125] J. Beck et al., "Characterization of neural breathing pattern in spontaneously breathing preterm infants.," *Pediatric research*, vol. 70, pp. 607–613, 6 2011. DOI: https://dx.doi.org/10.1203/PDR.0b013e318232100e

[126] M. Gabriel, M. Albani, and F. J. Schulte, "[bradycardias and apneas in premature infants].," *Monatsschrift fur Kinderheilkunde*, vol. 124, pp. 437–8, 5 May 1976.

[127] O. P. Mathew, "Respiratory control during nipple feeding in preterm infants.," *Pediatric pulmonology*, vol. 5, pp. 220–224, 4 1988.

[128] A. P. Marshall, K. Lim, S. K. Ali, T. J. Gale, and P. A. Dargaville, "Physiological instability after respiratory pauses in preterm infants.," *Pediatric pulmonology*, vol. 54, pp. 1712–1721, 11 2019. DOI: https://dx.doi.org/10.1002/ppul.24451

[129] T. B. Waggener, I. D. 3. Frantz, B. A. Cohlan, and A. R. Stark, "Mixed and obstructive apneas are related to ventilatory oscillations in premature infants.," *Journal of applied physiology (Bethesda, Md. : 1985)*, vol. 66, pp. 2818–2826, 6 1989.

[130] D. P. Southall et al., "Undetected episodes of prolonged apnea and severe bradycardia in preterm infants.," *Pediatrics*, vol. 72, pp. 541–551, 4 1983.

[131] J. E. Hodgman, F. Gonzalez, T. Hoppenbrouwers, and L. A. Cabal, "Apnea, transient episodes of bradycardia, and periodic breathing in preterm infants.," *American journal of diseases of children (1960)*, vol. 144, pp. 54–57, 1 1990.

[132] O. P. Mathew, C. K. Thoppil, and M. Belan, "Motor activity and apnea in preterm infants. is there a causal relationship?.," *The American review of respiratory disease*, vol. 144, pp. 842–844, 4 1991.

[133] M. L. Nock, J. M. Difiore, M. K. Arko, and R. J. Martin, "Relationship of the ventilatory response to hypoxia with neonatal apnea in preterm infants.," *The Journal of pediatrics*, vol. 144, pp. 291–295, 3 2004.

[134] M. Seppa-Moilanen, S. Andersson, and T. Kirjavainen, "Spontaneous and apnea arousals from sleep in preterm infants.," *Pediatric research*, vol. 89, pp. 1261–1267, 5 2021. DOI: https://dx.doi.org/10.1038/s41390-020-1068-2

[135] L. Curzi-Dascalova, E. Christova, P. Peirano, B. B. Singh, C. Gaultier, and G. Vicente, "Relationship between respiratory pauses and heart rate during sleep in normal premature and full-term newborns.," *Journal of developmental physiology*, vol. 11, pp. 323–330, 6 1989.

[136] C. N. Storrs, "Cardiovascular effects of apnoea in preterm infants.," *Archives of disease in childhood*, vol. 52, pp. 534–540, 7 1977.

[137] J. Werthammer, J. Krasner, J. DiBenedetto, and A. R. Stark, "Apnea monitoring by acoustic detection of airflow.," *Pediatrics*, vol. 71, pp. 53–55, 1 1983.

[138] R. Falsaperla, G. Vitaliti, C. Cimino, S. Catanzaro, and G. Corsello, "Apnea events in neonatal age: A case report and literature review.," *Medical hypotheses*, vol. 131, p. 109 296, m0m, 7505668 2019. DOI: https://dx.doi.org/10.1016/j.mehy.2019.109296

[139] O. G. Jenni, M. Wolf, M. Hengartner, K. Siebenthal, M. Keel, and H. U. Bucher, "Impact of central, obstructive and mixed apnea on cerebral hemodynamics in preterm infants.," *Biology of the neonate*, vol. 70, pp. 91–100, 2 1996.

[140] D. E. Elder, J. Whale, D. Galletly, and A. J. Campbell, "Respiratory events in preterm infants prior to discharge: With and without clinically concerning apnoea.," *Sleep & breathing = Schlaf & Atmung*, vol. 15, pp. 867–873, 4 2011. DOI: https://dx.doi.org/10.1007/s11325-010-0457-x

[141] S. C. Muttitt, N. N. Finer, A. J. Tierney, and J. Rossmann, "Neonatal apnea: Diagnosis by nurse versus computer.," *Pediatrics*, vol. 82, pp. 713–720, 5 1988.

[142] C. J. Upton, A. D. Milner, and G. M. Stokes, "Upper airway patency during apnoea of prematurity.," *Archives of disease in childhood*, vol. 67, pp. 419–424, 4 Spec No 1992.

[143] C. F. Poets, V. A. Stebbens, D. Richard, and D. P. Southall, "Prolonged episodes of hypoxemia in preterm infants undetectable by cardiorespiratory monitors.," *Pediatrics*, vol. 95, pp. 860–863, 6 1995.

[144] J. A. Adams, I. A. Zabaleta, and M. A. Sackner, "Hypoxemic events in spontaneously breathing premature infants: Etiologic basis.," *Pediatric research*, vol. 42, pp. 463–471, 4 1997.

[145] M. A. Mohr et al., "Very long apnea events in preterm infants.," *Journal of applied physiology (Bethesda, Md. : 1985)*, vol. 118, pp. 558–568, 5 2015. DOI: https://dx.doi.org/10.1152/japplphysiol.00144.2014

[146] G. Varisco, I. Lensen, D. Kommers, P. Andriessen, P. Bovendeerd, and C. van Pul, "The effect of apnea length on vital parameters in apnea of prematurity - hybrid observations from clinical data and simulation in a mathematical model.," *Early human development*, vol. 165, p. 105 536, edh, 7708381 2022. DOI: https://dx.doi.org/10.1016/j.earlhumdev.2021.105536

[147] I. Abdulhamid, P. A. Vauthy, B. A. Barnett, D. R. Hufford, R. P. Reddy, and C. E. Hunt, "Comparison of 2-channel and 4-channel pneumograms.," *Pediatric pulmonology*, vol. 13, pp. 245–249, 4 1992.

[148] L. Curzi-Dascalova, J. Bloch, M. Vecchierini, A. Bedu, and P. Vignolo, "Physiological parameters evaluation following apnea in healthy premature infants.," *Biology of the neonate*, vol. 77, pp. 203–211, 4 2000.

[149] O. S. Fathabadi et al., "Hypoxic events and concomitant factors in preterm infants on non-invasive ventilation.," *Journal of clinical monitoring and computing*, vol. 31, pp. 427–433, 2 2017. DOI: https://dx.doi.org/10.1007/s10877-016-9847-3

[150] H. E. Suichies, J. G. Aarnoudse, A. Okken, H. W. Jentink, F. F. de Mul, and J. Greve, "Skin blood flow changes during apneic spells in preterm infants.," *Early human development*, vol. 20, pp. 155–163, 3-4 1989.

[151] P. Latzin et al., "Lung volume, breathing pattern and ventilation inhomogeneity in preterm and term infants," *PLoS ONE*, vol. 4, e4635, 2 Feb. 2009, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0004635

[152] S. K. Roy, A. K. Shrivastava, and P. R. Choudhary, "Study of fetal hemoglobin with different gestational age and birth weight of the newborn," *International Journal of Contemporary Pediatrics*, vol. 8, p. 434, 3 Feb. 2021, ISSN: 2349-3291. DOI: 10.18203/2349-3291.ijcp20210522

[153] C. D. Townsel, S. F. Emmer, W. A. Campbell, and N. Hussain, "Gender differences in respiratory morbidity and mortality of preterm neonates," *Frontiers in pediatrics*, vol. 5, p. 6, 2017.

[154] T. Dassios, C. Harris, E. E. Williams, and A. Greenough, "Sex differences in preterm respiratory morbidity: A recent whole-population study," *Acta Paediatrica*, vol. 113, no. 4, pp. 745–750, 2024.

[155] Z. Liew, A. C. Fenton, S. Harigopal, S. Gopalakaje, M. Brodlie, and C. J. O'Brien, "Physiological effects of high-flow nasal cannula therapy in preterm infants," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 105, no. 1, pp. 87–93, 2020.

[156] D. Campbell, P. Shah, V. Shah, and E. Kelly, "Nasal continuous positive airway pressure from high flow cannula versus infant flow for preterm infants," *Journal of Perinatology*, vol. 26, no. 9, pp. 546–549, 2006.

[157] B. Jiang, L. Yao, H. Zhao, J. Liang, and Y. Feng, "Low body weight predicted bradycardia and desaturation in retinopathy of prematurity surgeries: A retrospective cohort study," *Frontiers in Pediatrics*, vol. 8, p. 226, 2020.

[158] S. Ahmad, A. Tejuja, K. D. Newman, R. Zarychanski, and A. J. Seely, "Clinical review: A review and analysis of heart rate variability and the diagnosis and prognosis of infection," *Critical Care*, vol. 13, no. 6, p. 232, 2009.

[159] C. J. Upton, A. D. Milner, and G. M. Stokes, "Episodic bradycardia in preterm infants.," *Archives of disease in childhood*, vol. 67, pp. 831–834, 7 Spec No 1992.

[160] V. Cardot et al., "Ventilatory response to a hyperoxic test is related to the frequency of short apneic episodes in late preterm neonates.," *Pediatric research*, vol. 62, pp. 591–596, 5 2007.

[161] P. Tourneux et al., "Relationship between functional residual capacity and oxygen desaturation during short central apneic events during sleep in "late preterm" infants.," *Pediatric research*, vol. 64, pp. 171–176, 2 2008. DOI: `https://dx.doi.org/10.1203/PDR.0b013e318179951d`

[162] Royal College of Paediatrics and Child Health. "How the API works." RCPCH Growth Charts. [Online]. Available: `https://growth.rcpch.ac.uk/clinician/how-the-api-works/`

[163] WHO Multicentre Growth Reference Study Group, *WHO Child Growth Standards: Length/Height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age: Methods and Development*. Geneva, Switzerland: World Health Organization, 2006, ISBN: 924154693X.

[164] H. Madsen and P. Thyregod, *Introduction to general and generalized linear models*. CRC press, 2010.

[165] T. Stijnen, T. H. Hamza, and P. Özdemir, "Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data," *Statistics in medicine*, vol. 29, no. 29, pp. 3046–3067, 2010.

[166] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[167] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[168] H. He and Y. Ma, "Imbalanced learning: Foundations, algorithms, and applications," 2013.

[169] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance.," *Journal of machine learning research*, vol. 11, no. 6, 2010.

[170] B. Phipson and G. K. Smyth, "Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn," *arXiv preprint arXiv:1603.05766*, 2016.

[171] A. J. Mitchell, A. Green, D. A. Jeffs, and P. K. Roberson, "Physiologic effects of retinopathy of prematurity screening examinations," *Advances in Neonatal Care*, vol. 11, no. 4, pp. 291–297, 2011.

[172] E. van Mastrigt et al., "Structural and functional ventilatory impairment in infants with severe bronchopulmonary dysplasia," *Pediatric pulmonology*, vol. 52, no. 8, pp. 1029–1037, 2017.

[173] N. P. Kondamudi, L. Krata, and A. S. Wilt, "Infant apnea," 2017.

[174] A. K. Yee, L. S. Siriwardhana, G. M. Nixson, L. M. Walter, F. Y. Wong, and R. S. Horne, "Developmental consequences of short apneas and periodic breathing in preterm infants," *Journal of Perinatology*, vol. 43, no. 11, pp. 1420–1428, 2023.

[175] P. F. Decima, K. L. Fyfe, A. Odoi, F. Y. Wong, and R. S. Horne, "The longitudinal effects of persistent periodic breathing on cerebral oxygenation in preterm infants," *Sleep medicine*, vol. 16, no. 6, pp. 729–735, 2015.

[176] O. Jenni, H. Bucher, K. v. Siebenthal, M. Wolf, M. Keel, and G. Duc, "Cyclical variations in cerebral blood volume during periodic breathing," *Acta Paediatrica*, vol. 83, no. 10, pp. 1095–1096, 1994.

[177] K. K. Iyer et al., "Bedside tracking of functional autonomic age in preterm infants," *Pediatric Research*, vol. 94, no. 1, pp. 206–212, 2023.

[178] Y. E. Tian, V. Cropley, A. B. Maier, N. T. Lautenschlager, M. Breakspear, and A. Zalesky, "Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality," *Nature medicine*, vol. 29, no. 5, pp. 1221–1231, 2023.

[179] A. Zalesky, J. Wen, and Y. E. Tian, "From whole-body to organ-specific biological age clocks," *Nature aging*, vol. 6, no. 5, pp. 961–969, 2026.

[180] A. Salih, T. Nichols, L. Szabo, S. E. Petersen, and Z. Raisi-Estabragh, "Conceptual overview of biological age estimation," *Aging and disease*, vol. 14, no. 3, p. 583, 2023.

[181] T. Noto, G. Zhou, S. Schuele, J. Templer, and C. Zelano, "Automated analysis of breathing waveforms using breathmetrics: A respiratory signal processing toolbox," *Chemical senses*, vol. 43, no. 8, pp. 583–597, 2018.

[182] A. Kuznetsova, P. B. Brockhoff, and R. H. Christensen, "Lmertest package: Tests in linear mixed effects models," *Journal of statistical software*, vol. 82, pp. 1–26, 2017.

[183] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[184] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *The Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.

[185] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[186] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern recognition*, vol. 48, no. 9, pp. 2839–2846, 2015.

[187] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017.

[188] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of machine learning research*, vol. 13, no. 2, 2012.

[189] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.

[190] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[191] A. Smola, V. Vapnik, et al., "Support vector regression machines," *Advances in neural information processing systems*, vol. 9, no. 155-161, p. 3, 1997.

[192] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[193] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[194] E. W. Steyerberg et al., *Clinical prediction models*. Springer, 2019, vol. 201.

[195] B. Van Calster et al., "Calibration: The achilles heel of predictive analytics," *BMC medicine*, vol. 17, no. 1, p. 230, 2019.

[196] B. Van Calster, D. Nieboer, Y. Vergouwe, B. De Cock, M. J. Pencina, and E. W. Steyerberg, "A calibration hierarchy for risk models was defined: From utopia to empirical data," *Journal of clinical epidemiology*, vol. 74, pp. 167–176, 2016.

[197] K. R. M. Fernando and C. P. Tsokos, "Dynamically weighted balanced loss: Class imbalanced learning and confidence calibration of deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 2940–2951, 2021.

[198] M. Altalhan, A. Algarni, and M. T.-H. Alouane, "Imbalanced data problem in machine learning: A review," *IEEE access*, vol. 13, pp. 13 686–13 699, 2025.