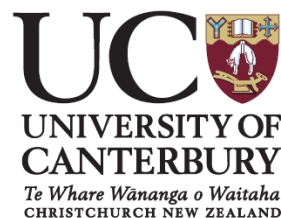


Analysis of Oxygenation and Other Risk Factors of Retinopathy of Prematurity in Preterm Babies

A thesis submitted in partial fulfilment of the
requirements for the Degree of
Doctor of Philosophy in Statistics
By Marina Zahari
School of Mathematics and Statistics

University of Canterbury

2015



Abstract

Maintaining adequate and stable blood oxygen level is important for preterm babies to avoid the risk of brain, lung and retinal injury such as retinopathy of prematurity (ROP). However, wide disparities in policies and practices of oxygenation in preterm babies exist among neonatal care providers as it is still unclear which best method of monitoring and what features of oxygen measurements are important to clinician's interpretations for assessing preterm babies at risk of developing severe ROP or unstable health condition. This thesis consists of two projects: NZ-ROP that examines multiple factors of severe ROP including summary statistics (mean, standard deviation (SD), coefficient of variation (CV) and desaturation) for oxygen saturation (OS) features in very extreme preterm babies, and NZ-LP that investigates the efficacy of some of these statistics for health monitoring of late preterm babies.

The OS data in NZ-ROP were recorded using modified oximeters that have offsets and inherent software artefact, both of which mask the actual saturation for certain OS ranges and may complicate the choice of methods in the analyses. Therefore, novel algorithms involving linear and quadratic interpolations are developed, implemented on the New Zealand data, and validated using the data of a UK preterm baby, as recorded from offsets and non-offsets oximeters. For all data sets, the algorithms produced saturation distributions that were very close to those obtained from the non-offset oximeter. The algorithms perform within the recommended standards of commercial oximeters currently used in the clinical practice.

ROP is a multifactorial disease, with oxygenation fluctuations as one of the key contributors. The all-subsets logistic regression, robust and generalised additive statistical modelling, along with a model averaging approach, are applied in NZ-ROP to determine the relationship of variability and level of OS with severe ROP, and the extent of contribution of various clinical predictors to the severity of this eye disease. Desaturation, as a measure of OS variability, has the strongest association with severe ROP among all OS statistics, in particular, the risk of severe ROP is almost three times higher in babies that exhibit greater occurrences of desaturation episodes. Additionally, this study identifies longer periods of ventilation support, frequent

desaturation events, extreme prematurity and low birth weight as the most important factors that substantially exacerbate the severity of ROP, and therefore signify babies' underlying condition of being severely ill.

Persistent cardiorespiratory instabilities prior to hospital discharge may expose preterm babies to a greater risk of neuro-developmental impairments. In NZ-LP, the statistical summaries of mean, SD and CV are computed from the OS measurements of healthy stable and unstable babies, and the performance of these statistics in detecting the unstable babies is evaluated using an extremeness index for outlying data and a hierarchical clustering technique. With SD and CV, the clinically unstable babies were very well separated from the group of stable babies, wherein, the separation was even more apparent with the use of CV. These suggest that measures of variability could be better than saturation level for highlighting babies' underlying instability due to immature physiological systems, but the combination of variability and level through the CV are believed to be even better.

Identification and summarisation of useful OS features quantitatively hold great promise for improved monitoring of oxygenation instability and diagnosis of severe ROP for preterm babies.

Acknowledgements

I would like to express my very special gratitude to Dr Dominic Savio Lee for his encouragement, tremendous patience and continuous belief in me throughout my research years. He selflessly taught me to pay attention to every detail of my research, and to appreciate the life in New Zealand. You are indeed a great teacher during my postgraduate life.

My heartfelt thanks to Professor Dr Brian Darlow and Dr Glynn Russell for kindly organising the NZ-ROP and NZ-LP data, and for their valuable suggestions and discussion on this neonatal research. I also want to thank Associate Professor Dr Marco Reale for taking on the role of my Administrative supervisor during the last months of my study. I wish to thank Nicolette McNeill and Patricia Graham at the Christchurch Women's Hospital for their kind assistance during my data collection. Thank you to Paul and Steve, and the faculty members of the School of Mathematics and Statistics at UC for their support and help.

Special appreciation also goes to Professor Dr Joseph Hilbe, Professor Dr Frank Harrell, Professor Dr Andreas Ruckstuhl and Dr. Xavier Robin for our fruitful discussion on statistical analysis and R functions. My gratitude to Professor Dr. Kamarulzaman Ibrahim, Professor Dr. Carey Priebe and Professor Ronald Broadbent for critically examining this thesis and providing constructive suggestions.

I am also grateful to Dr Xin Zhao, Dr. Miriam Hodge, Dr Mashitoh, Dr Norshidrah and En Jaafar, Sabariah and Haneef, Puan Hajar and En Kam, Puan Jah and En Hassan, Ann and Joe Goh, Angeline and many friends whom my family and I have shared good times with during the years at New Zealand.

I would like to thank my sponsors: the Malaysian Ministry of Education and the National University of Malaysia for the scholarship, the New Zealand government for the student allowance, and the School of Mathematics and Statistics at UC for sponsoring me to attend relevant conferences.

Importantly, special thanks to my family for bearing with me throughout the critical times. To my husband, Tarmizi for his endless love, encouragement and. To my delightful children: Fatin, Haziq, Hariz and Hafiz for being my greatest supporters. Thank you God for everything. The journey at the University of Canterbury has been a wonderful and fulfilling one.

Contents

List of Figures	x
List of Tables	xiii
Abbreviations	xv
1 Introduction	1
1.1 Neonatal Oxygenation.....	1
1.1.1 Complications of prematurity.....	2
1.1.2 Maturation of physiological systems.....	3
1.2 Previous Research of Oxygenation Studies.....	3
1.3 Thesis Objectives.....	7
1.4 Study Method.....	8
1.5 Contributions and Significance of Research.....	10
1.6 Overview of the Thesis.....	13
1.7 Thesis related publications and papers.....	15
2 NZ-ROP Study - Background	17
2.1 Introduction.....	17
2.1.1 ROP in preterm babies.....	17
2.1.2 Severe ROP worldwide.....	18
2.1.3 Significance of ROP problem.....	19
2.1.4 Assessment, classification and treatment of ROP.....	20
2.2 Pathophysiology of ROP.....	23
2.3 Previous and Current Research in ROP.....	24
2.3.1 Guidelines for blood oxygen levels to minimise the risk of ROP.	24
2.3.2 Oxygen fluctuations and ROP	25
2.3.3 Postnatal and clinical determinants of ROP.....	26
2.4 NZ-ROP Study	29
2.4.1 BOOST-NZ trial.....	29
2.4.2 NZ-ROP data.....	30
2.6 Summary.....	32

3	Pre-processing of Oxygen Saturation Data for NZ-ROP	35
3.1	Introduction.....	35
3.2	Data Clean-up.....	35
3.3	Operation of Masimo Oximeters for BOOST-NZ.....	36
3.4	Adjustments for Offsets and Dip Artefact.....	37
3.4.1	Procedure for identifying high or low saturation oximeters.....	37
3.4.2	Algorithm for adjusting offsets and dip artefact.....	39
3.4.3	Validation of the adjustment method.....	45
3.5	Summary Statistics for the OS Data.....	46
3.5.1	Checking for incomplete data.....	47
3.5.2	Data summary.....	47
3.6	Summary.....	48
4	Methodology for NZ -ROP	49
4.1	Introduction.....	49
4.2	Time Interval to Summarise the OS Statistics.....	49
4.3	Logistic Regression Modelling.....	50
4.3.1	The model.....	50
4.3.2	The response and predictors of the logistic regression.....	51
4.3.3	Collinearity.....	53
4.3.4	Predictor selection.....	53
4.3.5	Single- and multi-predictor analyses.....	54
4.3.6	Model averaging analysis.....	55
4.4	Diagnostics.....	56
4.4.1	Linearity assumption for continuous predictors.....	57
4.4.2	Logistic regression diagnostics.....	59
4.4.3	Robust logistic regression diagnostics.....	61
4.4.4	Outliers detection.....	62
4.5	Assessment of Models.....	63
4.5.1	Fitting of model.....	63
4.5.2	Measuring the performance of model.....	64
4.6	Validation of Model.....	66
4.6.1	Basic idea of cross-validation.....	66

4.6.2	Assessment of several validation schemes.....	67
4.6.3	Validation for NZ-ROP data.....	68
4.7	Computational Aspects.....	70
4.8	Summary.....	70
5	Results and Discussion for NZ-ROP	72
5.1	Introduction.....	72
5.2	Time Interval for Computing OS Summary Statistics.....	73
5.2.1	Determining the time interval.....	73
5.2.2	Discussion.....	78
5.3	The Response and Predictors for Logistic Regression.....	80
5.3.1	Characteristics of severe ROP and no severe ROP babies.....	81
5.3.2	Collinearity among the predictors.....	85
5.4	Logistic Regression Model Fitting and Diagnostics.....	88
5.4.1	Examination of structure of continuous predictors.....	88
5.4.2	Detection of outlying observations.....	95
5.5	Comparative Performance of Models.....	102
5.5.1	Single-predictor models.....	102
5.5.2	Multi-predictor models.....	105
5.5.3	Model averaging approach.....	108
5.6	Discussion.....	111
5.7	Summary.....	117
6	NZ –LP Study: Investigation of Some Statistical Summaries of OS Data	119
6.1	Instability in Neonatal Oxygenation	119
6.2	Studies on Oxygenation in Healthy Preterm Babies.....	120
6.2.1	Full term babies studies.....	121
6.2.2	Preterm babies studies.....	122
6.3	NZ-LP Study.....	123
6.3.1	Patients.....	123
6.3.2	Pre-processing of OS data.....	125
6.4	Methodology.....	127

6.4.1	Graphical assessment.....	127
6.4.2	Clustering analysis.....	128
6.5	Results.....	130
6.5.1	Box plot.....	134
6.5.2	Clustering.....	135
6.6	Discussion.....	138
6.7	Summary.....	140
7	Concluding Remarks	142
7.1	Overview and Summary.....	142
7.2	Data Management and Adjustments.....	144
7.3	Further Work.....	145
7.3.1	Curve-fitting methods.....	145
7.3.2	Measures of OS dynamics.....	146
7.3.3	Investigation on the behaviour of OS.....	146
7.3.4	Range of measure values.....	147
7.3.5	Time-dependent predictors.....	147
7.4	Concluding Remarks.....	147
A	MATLAB programs	151
A.1	Algorithm for cleaning and assembling NZ-ROP data	151
A.2	Algorithms for adjusting the offsets and artefact of Masimo oximeters	160
	References	163

List of Figures

2.1	Stages of retinopathy based on a diagrammatic view of retina. The white area in the centre is the optic nerve, the grey lines are the arteries, and the black lines are the veins. (Reprinted from “UK Retinopathy of Prematurity Guideline 2008” (Wilkinson et al., 2008) with permission of the Royal College of Paediatrics and Child Health)..	22
2.2	Flowchart of study population in NZ-ROP and ROP outcome.....	31
3.1	Flowchart of data clean-up procedure.....	36
3.2	A typical OS pattern for low saturation oximeters: SD(85-87%) is smaller than SD(93-95%) and a peak occurs at 96%.....	38
3.3	A typical OS pattern for high saturation oximeters: SD(93-95%) is smaller than SD(85-87%) and a peak occurs at 84%.....	38
3.4	Effects of offset and ‘dip’ artefact, as indicated by the circles, for low saturation oximeters.....	39
3.5	Corrections for offsets and dip artefact for low saturation oximeters.....	42
3.6	Effects of offsets and ‘dip’ artefact, as indicated by the circles, for high saturation oximeters.....	42
3.7	Results of corrections for offsets and dip artefact for high saturation oximeters.....	43
3.8	Comparison of unadjusted and adjusted readings from a Masimo low saturation oximeter and a Siemens oximeter.....	46
3.9	Comparison of unadjusted and adjusted readings from a Masimo high saturation oximeter and a Siemens oximeter.....	46
3.10	The relative frequency of infants with percentage of missing points in NZ-ROP data.....	47
4.1	Procedure of K -fold cross-validation for $K = 3$, which displays (a) partition of data and rotation of folds for each of the three successive iterations, and (b) training and testing of data within each iteration.....	67
5.1	Group means (with 95% CIs) of four OS statistics for 28 to 35 weeks PMA: first panel, cvOS; second panel, sdOS; third panel, meanOS; and fourth panel, dstOS.....	74

5.2	Kernel density plots (a), and box plots (b) for continuous variables of NZ-ROP according to ROP status.....	84
5.3	Scatterplots for each pair of predictors in NZ-ROP with binary ROP status encoded in red and blue (red dots indicate severe ROP cases and blue dots indicate no severe ROP cases).....	85
5.4	Partial residual plots of eight continuous predictors. The circles denote the residuals and the solid line depicts the smoothed residuals using a lowess smoother.....	89
5.5	Plots of GAM fit for eight continuous predictors, with pointwise standard errors and partial residuals included. The black solid line represents the fitted cubic spline $s(\cdot)$, the dotted lines represent the upper and lower pointwise twice-standard-error curves, and the dots represent the residuals. The rug plots at the bottom of each graph show the occurrences of the data values.....	91
5.6	Diagnostic plots for model $\text{dstOS} + \text{weight} + \text{daysVentilate}$. Jackknife residuals (top left); Cook's distance against leverage (top right); Marlow's weight against leverage (bottom left); and Marlow's weight against Cook's distance (bottom right). The dotted lines denote the thresholds, $8/(n - 2p - 2)$ for Cook's distance and $2(p + 2)/n$ for leverage. The numbers beside the circles correspond to the observation numbers.....	96
5.7	Plots of GAM fit (excluded observations 157, 130, 204, 8, and 207) for four statistics of OS (cvOS, sdOS, meanOS, and dstOS), with pointwise standard errors and partial residuals included. The black solid line represents the fitted cubic spline $s(\cdot)$, the dotted lines represent the upper and lower pointwise twice-standard-error curves, and the dots represent the residuals. The rug plots at the bottom of each graph show the occurrences of the data values.....	101
5.8	Unadjusted ORs (with 95% CIs) of severe ROP based on 252 babies in NZ-ROP for all predictors. For continuous variables, the value depicted reflects the odds of severe ROP for babies at the 75 th percentile of the distribution of the variable versus patients at the 25 th percentile.....	104

5.9	Box plots of fitted probabilities of severe ROP and no severe ROP for a single-predictor model, dstOS. The 25 th , 50 th , and 75 th quartiles are 0.1361, 0.246, and 0.370 for severe ROP, and 0.047, 0.079, and 0.171 for no severe ROP, respectively.....	105
5.10	Box plots of fitted probabilities of severe ROP and no severe ROP for a multi-predictor model, dstOS+weight+daysVentilate. The 25 th , 50 th , and 75 th quartiles are 0.174, 0.384, and 0.622 for severe ROP, and 0.016, 0.039, and 0.138 for no severe ROP, respectively.....	107
5.11	Model-averaged ORs (with 95% CIs) of severe ROP based on 252 babies in NZ-ROP for all predictors. For continuous variables, the value depicted reflects the odds of severe ROP for babies at the 75 th percentile of the distribution of the variable versus patients at the 25 th percentile.....	110
5.12	Box plots of fitted model-averaged probabilities of severe ROP and no severe ROP. The 25 th , 50 th , and 75 th quartiles are 0.147, 0.275, and 0.426 for severe ROP, and 0.035, 0.069, and 0.176 for no severe ROP, respectively.....	111
6.1	Kernel densities of OS for two babies, P02 (up) and P17 (down).....	132
6.2	Box plots for mean (left), SD (middle) and CV (right) at 34, 35 and 36 weeks PMA.....	134
6.3	Box plots for the mean (left), SD (middle) and CV (right) at 36 weeks PMA.....	135
6.4	Dendograms for hierarchical clustering with the mean (top), SD (middle) and CV (bottom).....	137

List of Tables

2.1	Severe ROP incidence in some high-, middle-, and low-income countries (adapted with permission from Zin and Gole, 2013).....	19
4.1	Evaluation of different validation methods.....	68
5.1	Group means of (a) cvOS, (b) sdOS, (c) meanOS, and (d) dstOS with respect to ROP status and PMA.....	75
5.2	Group means of cvOS, sdOS, meanOS, and dstOS, with respect to ROP status, based on a combined data set of 28 and 29 weeks PMA....	77
5.3	<i>P</i> -values from <i>t</i> -test for comparing differences in means of OS statistics between pairwise time periods in severe ROP and no severe ROP babies.....	78
5.4	Summary of severe ROP and no severe ROP babies.....	82
5.5	Estimates of Pearson correlation (with 95% CI) for each pair of predictors.....	86
5.6	VIF for eight continuous variables.....	87
5.7	Comparison between models with linear term, and GAM with nonlinear term for continuous predictors.....	92
5.8	Comparison between models with a linear term, and models with a nonlinear parametric term for meanOS and dstOS.....	94
5.9	Identification of candidate outliers from diagnostic measures for model dstOS+weight+daysVentilate.....	96
5.10	Candidate outliers and their occurrences in 21 models, as identified by four diagnostic measures.....	97
5.11	Comparison of parameter estimates between logistic and robust logistic estimation for a model, meanOS+weight+daysVentilate+O2W36.....	98
5.12	Comparison of parameter estimates between logistic and robust logistic estimation for a model, sdOS+GA+daysO2+wGA.....	99
5.13	Comparison of parameter estimates between MLE and robust estimation for a model, cvOS+weight+gender.....	100
5.14	Single-predictor logistic regression models of NZ-ROP.....	103

5.15	Predictor weight-for-GA and severe ROP.....	104
5.16	Selected multi-predictor regression models of NZ-ROP.....	106
5.17	Model-averaged <i>p</i> -values and ORs of severe ROP based on 252 babies in NZ-ROP.....	109
5.18	Findings on duration of oxygen therapy and ROP status.....	116
6.1	Summary of GA, birth weight and gender of the babies included in study.....	131
6.2	Summary of OS (%) data in healthy preterm babies at 34 to 36 weeks PMA.....	131
6.3	CV, SD and mean of OS for individual babies at 34, 35 and 36 weeks PMA.....	133

Abbreviations

AIC: Akaike Information Criterion
AUC: area under the ROC curve
CI: confidence interval
CH: Calinski-Harabasz
CV: coefficient of variation
GA: gestational age
GAM: generalised additive model
GLM: generalised linear model
IQR: interquartile range
IRWLS: iterative reweighted least squares
LL: log-likelihood
LR: likelihood ratio
MLE: maximum likelihood estimation
NICU: neonatal intensive-care unit
NZ-LP: New Zealand Late Preterm Babies
NZ-ROP: New Zealand Retinopathy of Prematurity
OR: odds ratio
OS: oxygen saturation
PMA: post-menstrual age
ROC: receiver operating characteristic
ROP: retinopathy of prematurity
RS10CV: repeat-stratified 10-fold cross-validation
SD: standard deviation
SE: standard error
VEGF: vascular endothelial growth factor
VIF: variance inflation factor
IGF-1: insulin-like growth factor
PaO₂: arterial oxygen tension
PtcO₂: transcutaneous oxygen

CHAPTER 1

INTRODUCTION

1.1 Neonatal oxygenation

Preterm babies are at an increased risk of morbidity, mortality, and hospital re-admissions than term babies (Petrini et al., 2009; Darlow et al., 1997). Each year, around 4,800 babies who are born within the period of less than 37 completed weeks gestation are admitted to neonatal intensive care units (NICUs) in New Zealand. Of these, about 4% or 200 are extremely preterm babies born before 28 weeks gestation and out of this figure, nearly 80% are discharged home alive (Cust et al., 2003). However, these babies do have considerably higher risk of poor growth, respiratory illness, visual deficits, cerebral palsy, sensori-neural and cognitive disabilities, educational and behavioural impairment than term babies (Darlow et al., 1997).

The condition of premature babies is characterised by physiological adaptation to extra-uterine life. Sudden unpredictable physiological instabilities, such as variations of blood oxygen level, heart rate, or respiratory patterns are common in preterm babies (Baird, 2004). Frequently, these conditions resulted from a complex combination of neonatal illnesses, immaturity of organ development and function, iatrogenic factors arising from parental handling and hospital procedures (Murdoch and Darlow, 1984), and environmental factors (Hay, 1969) such as room temperature, light, and noise. Hence, physiological variability is partly intrinsic to the baby and equally induced by external factors. This poses a challenge to clinicians of how best to treat the preterm babies prudently without inflicting unnecessary harm.

Now, modern intensive care has given hope in enabling many preterm babies to survive. One of the most important aspects of this care is help with the breathing and monitoring of the blood oxygen level because they have very immature lungs. Preterm babies frequently have breathing difficulties and require treatment with supplementary oxygen to improve the lung perfusion and the oxygen content, which changes every few seconds and is difficult to be controlled exactly. Oxygen content, however, should be maintained sufficiently high as insufficient blood flow or low

oxygen content in the blood (hypoxaemia) and excess of oxygen content (hyperoxaemia) do have adverse effects on the tissue oxygenation (Silverman, 2004). This, in turn, can lead to complications of prematurity or could trigger a cascade of respiratory events. It is extremely important that the oxygen level is monitored so as to ensure that oxygen content is adequate when caring for the babies.

1.1.1 Complications of prematurity

The process of blood flow and maintaining of oxygen content at certain level are integrated to meet the metabolic needs of tissues for ensuring a healthy human body. During breathing in a natural environment, molecules of oxygen are transported by the blood to the tissues of the body. Most of the molecules consumed by the tissue are metabolised. A small fraction of the consumption, however, is constantly converted to oxygen free radicals and reactive oxygen species. However, when overly produced, these by-products overwhelm the body's natural defence system and may, at least partly, play a role in development of tissue injury in preterm babies (Saugstad, 1998).

An inappropriate amount of oxygen content in the blood or inspired air may be harmful to preterm babies due to their under-developed biological systems. Oxygenation of the tissues depends on lung perfusion, which needs to adequately match the ventilation areas of blood capillaries. A ventilation-perfusion mismatch may cause hyperoxaemia and hypoxaemia, both of which have been implicated in diverse organ injury such as lung (Jobe & Bancalari, 2001), brain (McDonald, 1964) and retinal injury in preterm babies (Askie et al., 2009). On occasions when the amount of oxygen used exceeds what is needed, free radicals and reactive oxygen species are increased in hyperoxaemia and in re-oxygenation after hypoxaemia, which invariably lead to oxidative stress. However, according to Saugstad (1998), preterm babies are vulnerable to oxidative stress due to inadequate antioxidant protection, and this may expose the babies to oxidant-related morbidities such as bronchopulmonary dysplasia (chronic neonatal lung disease), retinopathy of prematurity (ROP) and periventricular haemorrhage.

1.1.2 Maturation of physiological systems

Physiological systems in preterm babies continue to develop after birth. Nevertheless, a delay in the maturation progress could place a baby at a greater risk of severe cardiorespiratory events (Gaultier, 1995), which are reflected by patterns of periodic breathing and apnea (a pause in the regular breathing), bradycardia (a slow heart rate) and hypoxaemia of oxygenation (Henderson-Smart et al., 1986). Some of the observed patterns are thought to be mediated through a sensory nervous system in blood vessels, also known as peripheral chemoreceptor reflex, which monitors chemical concentrations in blood with advancing gestation and postnatal age (Rigatto et al., 1975). It is common for apnea, bradycardia and hypoxemia to occur concurrently, with early onset of bradycardia during apnea being pre-mediated by hypoxaemia (Henderson-Smart et al., 1986). Of particular concern, persistent events, even if subtle, may cause delayed mental, motor and poorer language development at later years in preterm babies (Mattia and deRegnier, 1998). These ramifications, however, would largely depend on the ability of babies to recover from any single events, and to stabilise after being given appropriate interventions (Balfour-Lynn et al., 2009).

1.2 Previous research of oxygenation studies

Numerous studies have been conducted to differentiate instability arising from physiological immaturity or complications of prematurity so as to decrease mortality and improve morbidity in this high-risk group during postnatal period. Research has found that illnesses such as respiratory distress, anaemia of prematurity and gastro-oesophageal reflux may contribute to variations in heart rate and breathing patterns (Baird, 2004). The investigation and treatment for some of these complications have been initiated using physiologic data from electronic monitors, for instance, by utilising heart rate variability analysis that quantifies the oscillations in the ECG signals (Malik et al., 1996). Existing use of heart rate and respiratory rate variability suggests potential for predicting outcome therefore permitting timely intervention. Changes in heart rate variability have been found to be able to predict morbidity and mortality in fetal (Yum & Kim, 2003) and neonatal patients (Griffin et al., 2007) based on heart rate characteristics. Measures of respiratory rate variability and severity indices have been developed to study prolonged ventilation in adult patients

(Seely et al., 2014) and to describe pulmonary status in ventilated preterm patients (Madan et al., 2005).

Establishing evidence-based information on the links between varying oxygen levels and the development of neonatal organs, however, remains a challenge to the clinicians because an appropriate range of oxygen levels that would not cause hyperoxic (abnormal increase of oxygen supply to the organs and tissues) or hypoxic (insufficient oxygen supply to the organs and tissues) injury is still unknown (Higgins et al., 2007; Askie, 2013). Results from several randomised control trials (BOOST II Group, 2013; SUPPORT Study Group, 2010) have found that high oxygen saturation (OS) levels predispose to ROP, which is the leading cause of blindness and visual disability in newborns, while high inspired oxygen concentrations exacerbate neonatal chronic lung disease, which is associated with poor growth, neuro-developmental impairment and long term respiratory morbidity (Jobe & Bancalari, 2001). However, insufficient oxygen also leads to poor growth and neuro-developmental impairment due to brain damage (McDonald, 1964), and mortality due to hypoxic respiratory failure (BOOST II Group, 2013). These studies showed that there is a crucial need to seek “safe” OS ranges for the most vulnerable premature babies. In a 2004 editorial in *Pediatrics* (Silverman, 2004), Dr William Silverman, formerly of Columbia University, stated, “...*there has never been a shred of convincing evidence to guide limits for the rational use of supplemental oxygen in the care of extremely premature infants. For decades, the optimum range of oxygenation (to balance four competing risks: mortality, ROP-blindness, chronic lung disease and brain damage) was, and remains to this day, unknown*”.

Oxygen mediated retinal injury is one of the critical areas in need of urgent study (Cole, 2010). A major area of concern is investigating various possible oxygen-related determinants that contribute to the development and progression of ROP. Early findings of the role of oxygen in inducing ROP were conducted using transcutaneous electrodes for measuring transcutaneous oxygen or umbilical catheters for measuring oxygen tension in arterial blood gases; these technologies were later replaced by OS monitoring of pulse oximeters. Nevertheless, oxygenation monitoring using these methods has been shown not to improve ROP outcome in preterm babies requiring intensive care (Flynn et al., 1987; Kinsey et al., 1977; Quine and Stenson,

2008). Times series data generated by the monitors, however, could be more useful as the source data for more complex analysis by mathematical or statistical modelling. Studies using regression models have found that standard deviation (SD) of transcutaneous oxygen in the first 2 weeks of life was a significant predictor of severe ROP (Cunningham, et al., 1995); increased fluctuations in arterial oxygen tension, as quantified by coefficient of variation (CV), imposed a higher risk of developing threshold ROP in very-low-birth-weight premature babies (York, et al., 2004; Saito et al., 1993); and desaturation episodes (intermittent falls in oxygen levels) of OS were higher in babies with severe ROP (Di Fiore et al., 2010). Until now, comparative studies have not been done to evaluate which measures of oxygen could possibly provide better insight in describing the link between oxygenation fluctuations and ROP.

Several early clinical trials (1951-1969) that investigated the effects of restricted versus unrestricted oxygen concentrations on ROP and mortality had reported that severity of ROP was reduced with the restriction of oxygen (Askie et al., 2009). Nonetheless, it was observed that some babies in the restricted oxygen group still developed ROP whilst many babies never developed ROP in the unrestricted oxygen group. This indicates that oxygen is not the only independent risk factor of ROP, and that there are other factors that also contribute to the cause of ROP. For instance, lower birth weight and younger gestational age (GA) at birth (Darlow et al., 2004; McColm & Fleck, 2001; Good et al., 2005) have been identified as having the greatest association with the risk of ROP. Male gender (Jacobson et al., 2009; Darlow et al., 2004); greater illness severity (Hagadorn et al., 2007) measured by the Score for Neonatal Acute Physiology (Richardson et al., 1993); clinical oxygen (Kinsey et al., 1956; Lala-Gitteau et al., 2007); and upregulation of vascular growth factor (Chen and Smith, 2007) are amongst other important contributors of this eye disorder. ROP may also relate to poor postnatal growth (Wikstrand et al., 2011; Hellstrom et al., 2009), postnatal steroid exposure (Halliday et al., 2003) and oxidative stress (Saugstad, 1998).

Monitoring the development of physiological systems in preterm babies is an area in need of additional study (Raju et al., 2006). Several studies were set up to investigate symptoms of delayed maturation in babies born less than 35 weeks GA, of

whom occurrences of apnea, bradycardia, hyperoxemia and hypoxemia events were recorded longitudinally for six months (Hunt et al., 1996; Hunt et al., 2011). The risk for these events was highest at 36 weeks post-menstrual age (PMA) but declined with age at 44 weeks PMA. Further, such events that occurred more than four times (Hunt et al., 2004) and prolonged beyond 40 weeks PMA (Poets et al., 1991) were found to be associated with poor performance of gross and fine motor skills, and cognitive ability. The extent of morbidity in relation to the cardiorespiratory events may explain the concern on whether some preterm babies, particularly the late-preterm ones (babies of 33 to 36 weeks gestation), are clinically ready for NICU discharge (Engle et al., 2007) when their physiological systems are still immature. It is, therefore, important to detect pre-discharged babies with suspected unstable health conditions to ensure necessary treatments could be provided with the aim of improving the babies' outcome.

Whereas pre-discharge criteria have been established for respiratory and heart rates to guide interventions during neonatal transition prior to hospital discharge (American Academy of Pediatrics, 2004), evidence-based guidelines have been difficult to be obtained for oxygenation of a preterm baby (Higgins et al., 2007; Askie, 2013). Nevertheless, there have been several observational studies describing the oxygenation patterns and typical values of OS during the first days and months after birth in preterm babies. Most of these studies have considerable interest in saturation level (Rhein et al., 2014; Harigopal et al., 2011; Hunt et al., 2011; Beresford et al., 2005; Ng et al., 1998; Poets et al., 1992), and some common summary statistics that have been used are mean, median, percentiles, and median baseline (median of the means). These studies focus on OS level and not on variability in OS.

Some clinicians believe that information about the health state of preterm babies is contained in the variability of physiological systems. Studies have found that preterm babies with very low birth weight and/ or complications are more likely to exhibit lower level or higher variability of blood oxygen. For example, babies with severe ROP that require retinal surgery have higher oxygen variability (York et al., 2004; Saito et al., 1993); babies with chronic lung disease have greater number of desaturation episodes (Durand et al., 1992); and babies with infection have increased hypoxaemia and apnoea incidences (Hofstetter et al., 2007). A statistic that is often

used to represent variability in OS data is SD (Cunningham et al., 1995; DiPietro et al., 1994). An alternative but infrequently used measure to quantify OS is CV (Grove et al., 2005; York et al., 2004; Saito et al., 1993), which quantifies the variability of OS relative to the mean level.

Appropriate measures of blood oxygen need to be identified for distinguishing preterm babies at high risk of ROP or babies with unstable oxygenation. This would be a daunting task since the definition for optimal high and low limits for OS to guide oxygenation monitoring in either healthy or oxygen supplemented preterm babies is still not clearly defined, which in part, could be due to the absence of clarity on how often and/or what length and duration of an apneic episode and hypoxemia is safe (Cole et al., 2003; Silvestri et al., 2002), and lack of knowledge on what suitable measures can sufficiently summarise the instability in blood oxygen. That this knowledge has not been forthcoming is due to the complexity and multitude of information involved, which precludes most medical researchers from examining the physiology as the interrelated system that it is. Much existing research in this area has examined individual measures of oxygenation, but not all of them together, which is specifically what is being sought in this thesis. Such measures need to be carefully chosen to ensure that they possess blood oxygen features that reflect the instability of oxygenation in a baby.

1.3 Thesis objectives

There exist wide disparities in current policies and practices of oxygenation in preterm babies among neonatal care providers due to limitations in current knowledge on how to appropriately manage the oxygen for newborn babies. There is a need to determine useful measures of OS for saturation monitoring, to better understand the links between OS and retinal injury (Cole et al., 2003), and to investigate whether some indication of health status of preterm babies could be determined from certain measures of routinely collected OS data. These are the motivations for this research, which is structured as two studies, New Zealand Retinopathy of Prematurity (NZ-ROP) and New Zealand Late Preterm Babies (NZ-LP), with separate groups of babies.

The goal of NZ-ROP and NZ-LP studies is to provide some information on several aspects of oxygenation that could support the clinician's assessment when evaluating the risk of severe ROP and health status of preterm babies in NICUs. Oximeter OS data and other clinical information are utilised in the two studies for detailed statistical modelling and various techniques. Some statistical measures based on the continuously collected saturation data are then identified to provide additional insights into physiologic conditions in unstable premature babies and pathologic conditions in sick babies.

Altogether, there are three specific aims in this thesis, two of which are the aims of NZ-ROP and the other is the aim of NZ-LP. Specifically, the aims of NZ-ROP are:

1. to determine any plausible relationships between saturation levels, three measures of variability of OS (specifically SD, CV and desaturation), and severe ROP and
2. to identify dominant predictors that have stronger relationship with severe ROP.

The aim of NZ-LP is to determine whether three statistical summaries (mean, SD and CV) of the OS recordings are potentially useful in the clinical care of late preterm babies.

1.4 Study method

Various methods have been identified to achieve the research objectives of NZ-ROP and NZ-LP. With respect to the first two objectives, the study of NZ-ROP is carried out in two phases. The first phase involves pre-processing of OS data to render them suitable for analysis. The data were recorded from the Masimo SET oximeters and were obtained from the BOOST-NZ trial. This was a double-blind randomised trial (Darlow et al., 2014) with its own objectives in investigating OS targeting in two groups of preterm babies, of whom each baby was randomly allocated with one of two types of oximeters. The oximeters had two features: trial-imposed offsets of either +3% or -3% in the OS range 85-95% that masked the trial intervention, and machine artefact inherent in the oximeters. Both of these features influenced the OS readings. Part of the pre-processing includes identifying the type of oximeter used for each

baby, removal of the offsets and correction of the artefact. Novel algorithms through interpolation techniques are developed, implemented, and validated to adjust for these features. Subsequently, an appropriate period for summarising the OS data has to be determined. For this, some clinical factors such as the onset of ROP (Palmer, 2006) and age of the babies are taken into consideration.

The second phase of NZ-ROP involves the analysis of predictors that could contribute to the development of severe ROP. This includes identifying whether some measures of saturation levels or variability of OS are related to severe ROP. As opposed to previous research that had studied measures of oxygen fluctuations separately based on older technology, this research considers four summary statistics of OS readings of non-invasive oximeters: mean for measuring saturation level; SD for measuring saturation variability; CV as a composite measure of variability relative to the mean level, and desaturation as another variability measure. By doing so, the degree of relationship between each measure of OS and severe ROP could be evaluated. These statistics are computed from the OS readings and are analysed as part of the predictors in the regression analysis for predicting risk of severe ROP in very young preterm babies less than 28 weeks GA. The all-subsets logistic regression, which produced models with all possible combinations of predictors, is employed. During the model verification stage, identification of outlying observations and appropriate functional forms for the continuous predictors is checked using conventional, robust (Cantoni & Ronchetti, 2001) and generalised additive statistical modelling (Hastie and Tibshirani, 1990). The performance of the models is measured by the area under the receiver operating characteristic (ROC) curve (AUC), and validated using a repeat-stratified 10-fold cross-validation technique (Kohavi, 1995; Refaeilzadeh et al., 2009). Single-, multi-predictor, and model averaging approaches are used to assess the relationships between the OS statistics, seven other clinical predictors, and severe ROP, as well as to identify dominant predictors that are strongly related to the diagnosis of severe ROP. The analysis by the model averaging approach is based on several measures computed by averaging over all models.

Data for NZ-LP were obtained from the Christchurch Women's Hospital. These data consisted of OS readings of 31 preterm babies that were recorded from the Masimo Radical oximeters. Except for one baby, most of the recordings were

conducted when the babies were 34 to 36 weeks PMA. The babies were clinically well at the time of enrolment for the study, however, two babies were soon identified as being clinically mildly unstable. Using these data, some measures of OS features are evaluated whether they could be informative to the clinicians when monitoring late preterm babies potentially with unstable health conditions. To address this issue, statistical summaries of mean, SD and CV of OS are evaluated in a two-phase analysis. In the first phase, all suspected erroneous data are identified and assigned with a lesser weight through the use of an appropriate weighting function. In the second phase, the distribution of saturations for several stable and unstable babies is assessed using nonparametric kernel density technique (Wand & Jones, 1995). Subsequently, the OS readings of each baby are summarised using the three statistical measures, and the ability of these measures to distinguish the unstable babies is investigated. This could be done in two ways. One is through the use of graphical approach, i.e. by assessing the capability of each measure in identifying the two babies as outlying data points as well as quantifying the level of extremeness of these outliers from the main data set. Clustering technique is another approach that can be considered, i.e. each measure is used to group the babies into clusters, such that babies within a cluster looked more “similar” with respect to the measure than babies from different clusters. This is performed using the hierarchical clustering method (Sneath, 1973) in conjunction with a dendrogram and a clustering performance index (Calinski & Harabasz, 1974) to assess the overall results.

1.5 Contributions and significance of research

Oxygen has been established as one of the key factors of ROP (Cole, 2010). However, more information is required pertaining to the relationships between oxygenation fluctuations and ROP, particularly so with respect to GA, timeline of the onset, or relationship with other risk factors (Cole et al., 2003). Studies have been done to improve the understanding of the origin and mechanism of ROP, but not without limitations. There is a subjective interpretation of outcome due to the different oxygen monitoring devices used, a lack of knowledge of whether and which measures of OS are related to the severity of ROP, and a lack of understanding on the relationships between oxygenation, other clinical and postnatal factors, and severe ROP. These issues underscore the importance of conducting a detailed and comprehensive analysis in understanding the development (or healing) of severe ROP,

and these are addressed in NZ-ROP. It is a study that investigates multiple determinants, including four OS measures, of severe ROP in extremely low GA (<28 weeks GA) babies.

Defining the “appropriate” blood oxygen levels or dose of supplemental oxygen in preterm babies is one of the primary interests in the perinatal and neonatal medical community. However, this goal has been hindered as to which best method of monitoring and what features of the oxygen readings to analyse are still debated. Previous studies have attempted to address these issues by suggesting several typical measures of the saturation levels when investigating babies’ oxygenation during the early life. In NZ-LP, three summary statistics based on OS variability and levels are considered and assessed for saturation monitoring of late preterm babies.

The contributions of NZ-ROP and NZ-LP are the identification of OS measures of variability as the useful features of blood oxygen for examining and detecting babies who may be at future risk for severe ROP and babies that may have unstable health conditions prior to hospital discharge. Specifically in NZ-LP, measures of SD and CV perform better than mean for separating mildly unstable babies in one group and all stable babies in another group. In NZ-ROP, a measure for desaturation events is found to be strongly associated with severe ROP. Additionally, the study demonstrates that a combined impact of four factors: frequent episodes of desaturation, longer duration of ventilation support, extremely low birth weight and high degree of immaturity is substantial on the development of ROP sufficiently severe to require retinal surgery.

Influential evidence on the relationships between oxygen and ROP occurrences are now emerging, with past studies showing that fluctuations of oxygenation (Cole, 2010) and treatment of oxygen therapy (Kinsey et al., 1977) have a role in this neonatal complication. The data from the BOOST-NZ oxygen targeting trial could be used to investigate the risk factors of severe ROP. However, these are new data sets and have not been analysed in any other studies. In particular, the OS readings are influenced by the trial-imposed offsets and software artefact of the oximeter, and the oximeter’s type for each baby is masked from the parents and neonatal care providers. As such, these hinder the OS data to be directly used for NZ-

ROP. Substantial efforts are carried out to manage and process the NZ-ROP data set so as to ensure that the oximeter's type is identified and the affected OS values are adjusted using novel algorithms to reflect the actual readings.

Implementation of the algorithms is also essential for other studies on preterm babies. The OS data could be useful to study patients of neonatal morbidities associated with the use of oxygen such as bronchopulmonary dysplasia (Askie et al., 2003) and damage to the postnatal growth of length, weight, and head circumference (Tin et al., 2001). The presence of the offsets and artefact in the data, however, could mask certain borderline threshold of saturations for each morbidity; thereby, could result in failure to identify babies with potentially severe outcomes.

The research outcomes of NZ-ROP and NZ-LP are important to the health care providers and academia. The knowledge that the research seeks will contribute to the wider neonatal community (neonatologists, paediatricians, ophthalmologists, and neonatal nurses) and towards better understanding of oxygen instability and its monitoring, and management of high-risk babies within and across the population of New Zealand preterm babies. This information would increase awareness in enhancing the care of high-risk newborn babies and stimulate ideas to develop evidence-based preventive and corrective strategies for future neonatal treatment. The research can also be important for academia as it provides a thorough description of the projects, which can be analysed and critiqued from different angles, for instance, using the regression analysis for NZ-ROP. The investigation of ROP determinants, which includes measures of oxygenation fluctuations and other clinical predictors, may confer better understanding on the extent of their relationship with severity of ROP. Consequently, the outcomes of predictive models for ROP may be of use to clinicians when evaluating vulnerable young babies at risk of the sight-threatening disease.

Continuous oxygen monitoring plays an important role in assessing the health of preterm babies. In the clinical care of these vulnerable babies, appropriate tools are required for identifying medical complications (Higgins et al., 2007; Raju et al., 2006). Successful identification of OS measures is therefore important for monitoring the oxygenation status of premature babies during their early weeks of life. This

would allow the clinicians to control and reduce factors that aggravate oxygen lability, and evaluate current care practices for oxygen supplemental administration.

As well as contributing to health care provision, all New Zealand high-risk babies will benefit from up-to-date evidence-based research from the two studies, with the goal to reduce the likelihood of developing neuro-developmental disability or ROP sufficiently severe to require laser therapy. The reduction in such morbidities and financial burden would enhance their quality of life and benefit their family and the society (Saigal and Doyle, 2008).

Overall prematurity accounts for much of the cost and disability from NICU care. Preterm babies contribute disproportionately to the work loads of neonatal intensive care units (Cust et al., 2003). According to Darlow and colleagues (2007), these babies have increased morbidity across a range of health and developmental outcomes at two years of age compared with healthy controls. The outcomes from the studies could be essential to reduce the burden of disability on health service resources in the community, in terms of time commitments and workloads of the NICUs medical team and specialist visit. In turn, these could reduce the long-term costs to New Zealand's health system and social services. The exact size of this cost saving, however, can only be estimated accurately by a formal economic evaluation.

1.6 Overview of the thesis

The structure of the thesis is as follows. Consequences and issues related to instability in blood oxygen level as well as the current challenges in managing these concerns are addressed in Chapter 1. NZ-ROP is described in Chapter 2 until Chapter 5; Chapter 3 and Chapter 5 form the main parts of the thesis. NZ-LP is discussed in Chapter 6.

Chapter 2 reviews on the pathological details and risk factors of ROP, motivation and relevant background of NZ-ROP. Description is also given for the three main components (OS readings, diagnosis of ROP status, and clinical and postnatal information) of data that are used in the analysis.

Chapter 3 concentrates on the pre-processing of OS data that were recorded using Masimo SET oximeters. Some of the challenges faced in processing these data include unmasking the oximeter's type for each baby involved in the study, and removing the effects of artefact and trial-imposed offsets from the OS readings. An outline of procedure for identifying the type of oximeter, and description for development and implementation of two novel algorithms for adjusting the two features are provided in this chapter. These steps are important to render the data suitable for use in the regression modelling, which is the focus of Chapter 4 and 5. Additionally, validation of the algorithms is also demonstrated using the offset and non-offset data of a UK preterm baby.

Chapter 4 provides the background material of statistical techniques used for analysing plausible relationships between four measures of OS, seven other predictors, and severe ROP. These relationships are examined simultaneously using all-subsets regression. The approaches that are used here extend the work of previous studies, which have mainly focused on single regression model and single measure of oxygen, and utilised older technology of blood oxygen monitoring.

The results and discussion of NZ-ROP are presented in Chapter 5. Various ideas are forwarded in this chapter when analysing the OS features (saturation level and variability) and other predictors for severe ROP diagnosis. For instance, using subtle patterns in the oxygenation of babies with and without ROP to locate a suitable period for summarising the OS measures; fitting regression models with one predictor from each group of common predictors that are highly correlated for handling multicollinearity; and grouping of common predictors of similar risk factor groups for examining the relative importance of ROP's predictors with respect to their relationship with this eye disorder.

Chapter 6 describes the investigation of NZ-LP. A review is presented on the previous works of obtaining some typical measures of saturation levels as the common values for oxygenation monitoring. Treatments are outlined for missing or unreliable OS readings and description of techniques is given to evaluate the ability of three statistical summaries: mean, SD, and CV of OS. The results of the analysis are

illustrated using an extremeness index of the box plots, and a hierarchical clustering method.

Concluding remarks for the thesis and recommendations for future direction of research are given in Chapter 7.

1.7 Thesis related publications and papers

Paper in preparation

Zahari, M., Lee, D. S. & Darlow, B. A. Desaturation episodes and risk factors for severe retinopathy of prematurity among New Zealand very premature babies. In prep for the *Journal of Pediatrics*.

Refereed journal paper

Zahari, M., Lee, D. S. & Darlow, B. A. Algorithms that eliminate the effects of calibration artefact and trial-imposed offsets of Masimo oximeter in BOOST-NZ trial. *Journal of Clinical Monitoring and Computing*. DOI: 10.1007/s10877-015-9752-1.

Lee, D. S., Zahari, M., Russell, G., Darlow, B. A., Scarrott, C. J., & Reale, M. (2011). An exploratory investigation of some statistical summaries of oximeter oxygen saturation data from preterm babies. *ISRN Pediatrics*. DOI: 10.5402/2011/296418.

Conference presentations

Zahari, M., Lee, D.S., Russell, G., Scarrott, C., Reale, M., et al. (2008). Quantifying blood oxygen levels in preterm infants using coefficient of variation. *Australian Statistical Conference*, 30 June – 3 July 2008, Melbourne, Australia.

Lee, D.S., Darlow, B. & Zahari, M. (2010). Bayesian nonparametric modelling of risk factors for retinopathy of prematurity. *Ninth Valencia International Meeting on Bayesian Statistics/ 2010 World Meeting of the International Society for Bayesian Analysis*, 3 - 8 June 2010, Benidorm, Spain.

Zahari, M., Lee, D., Russell, G., Darlow, B., Scarrott, C. & Reale, M. (2009). Identifying unstable preterm infants using measures of oxygen saturation in different

behavioural states, 2009, *International Biometric Society Australasian Region Conference*, 2009, Lake Taupo, New Zealand.

CHAPTER 2

NZ-ROP STUDY-BACKGROUND

2.1 Introduction

This chapter reviews the background of ROP, which includes the historical aspect, pathophysiology or development in the developing retina, classification and treatment of ROP severity as well as the group of babies at risk and related risk factors. Previous and recent developments in ROP studies are also discussed. For instance, the guidelines for safe OS monitoring in clinical practice, investigations on the relationship between arterial oxygen tension, transcutaneous oxygen and OS, and on oxygenation fluctuations and ROP. A brief description of a study on ROP predictors is given in the final section.

2.1.1 ROP in preterm babies

ROP is an abnormality of the retina and disorder of immature retinal blood vessels in babies born prematurely. The incomplete formation of the retina is due to failure of retinal blood vessels to develop normally and to grow to the entire retina. In the absence of timely and appropriate treatment, it can progress aggressively to retinal detachment leading to visual loss. ROP is a visual impairment threat to babies in various parts of the world, particularly so in countries with access to neonatal care of resuscitation (Gilbert et al., 2005).

Oxygen therapy for neonatal treatment was introduced in the 1930s in the US and soon, it became widespread to other industrialised countries. As proper devices for measuring blood oxygen levels were unavailable, clinicians found it difficult to target supplemental oxygen that meets each baby's need. The first epidemic of ROP was first reported in 1942 (Kinsey et al., 1956) and by 1950s, the role of breathing unrestricted supplemental oxygen as the major cause was discovered by three randomised control trials. As a result of these trials, the use of oxygen in nurseries was drastically curtailed, where the babies were often administered with less than 40% of supplemental oxygen. Such a restriction resulted in a decrease in blinding ROP, but there were also increased mortality and morbidity rates (Cross, 1973). The

second epidemic occurred in the 1970s, in which advances in health systems for both antenatal and neonatal care have contributed to the survival of a greater proportion of high-risk babies. Extreme prematurity, with some cases as young as 23 weeks GA, was the driving force for this epidemic to persist (Slidsborg et al., 2008; Zin and Gole, 2013); thus, stimulating clinicians to develop methods for preventing and detecting potentially blinding ROP. In the last decade or so, the rate of children who develop blinding ROP has been increasing in newly industrialising and developing countries (Gilbert et al., 2005). This is termed as the third epidemic.

2.1.2 Severe ROP worldwide

ROP is one of the three leading causes of childhood blindness in the world for the past 10 years (Kong et al., 2012). In 2010, it was estimated that nearly 22% or 184,700 of surviving preterm babies had some degree of ROP. Of these, almost 6.7% or 12,300 babies developed mild or moderate visual impairment and a further 10.8% progressed to severe impairment or become blind (Bleincowe et al., 2013). Incidence for ROP, however, differs enormously per country as it depends on the standards of health care and socioeconomic development. ROP was once considered as a disease that mostly occurred in developed countries; nonetheless, incidence of ROP has been emerging in middle-income countries and urban areas of low income countries, in which these account for 65% of visual impairment in children. Lowest income regions, however, have very low percentage of ROP cases with mortality among premature babies as high as 50% due to lack of good quality health care facilities for babies and pregnant women (Gilbert et al., 2005).

More ROP incidences are now being reported for severe ROP. This generally refers to sight threatening ROP at advanced stages for which treatment will be offered. The incidence of severe ROP in several high-, middle-, and low-income countries over the past two decades (Zin and Gole, 2013) is summarised in Table 2.1. It can be seen that severe ROP is emerging as a major cause of blindness in middle-income nation such as China, and regions such as Southeast and South Asia, Latin America and parts of Eastern Europe. In these places, many of the cases have been reported to occur in more mature (32-36 weeks of GAs) and larger (>1500g of birth weights) babies. Such a phenomenon could be due to the rapid improvement in the maternal and neonatal facilities, which in part, has induced delivery of moderately to

late premature babies for non-medical reasons (Zin and Gole, 2013). In high-income countries, incidences of ROP are widely ranged from nearly 4% to 35% (Table 2.1). However, blindness and severe visual impairment is more prevalent among newborns with low GAs <28 weeks or extremely low birth weights <1000g (Horbar et al., 2012).

Table 2.1: Severe ROP incidence in some high-, middle-, and low-income countries (adapted with permission from Zin and Gole, 2013)

Country	Country group by income	Period	Number of babies	Incidence (%)
United States	High	2003-2007	6866	16.0
United Kingdom	High	1987-1998	484	5.2
Australia	High	1992-2009	373	15.0
Canada	High	2006-2007	1866	3.9
Germany	High	1993-2007	1473	14.8
Sweden	High	2007-2009	506	34.8
Argentina	Upper middle	2008-2010	1169	13.4
Brazil	Upper middle	2004-2006	3437	3.4
China	Upper middle	2009-2010	2185	13.1
Romania	Upper middle	2002-2007	1783	15.2
India	Lower middle	1999-2002	1083	11.0
Pakistan	Lower middle	2003-2006	68	20.6
Vietnam	Lower middle	2001	225	9.3
Bangladesh	Low	1998-2003	114	4.4

2.1.3 Significance of ROP problem

ROP is a public health issue with economic and social implications. Premature babies with ROP face visual impairment including blindness. The impact on these survivors of prematurity when entering the society with potential life-long blindness is particularly tragic. They may have more difficulty in securing employment; therefore, hindering them from producing economic benefits to the society and in turn, requires aid from their families or social services. Blindness due to severe ROP also negatively affects neuro-developmental function. Babies with unfavourable visual acuity fare

worse functional outcomes, with 77% have self-care disability, 50% continence disability, 43% motor disability, and 66% social cognitive disability (Msall et al., 2000). The adverse sequels resulting from these disabilities affect the quality of life of the children and inflict a long-term burden on their families, society, and health care resources (Saigal and Doyle, 2008).

Disabilities impose great strains on the family. Many parents with a disable child have reported additional pressures and stresses on their other children, health and marriage when coping with their child's disabilities. Also, these parents are less likely to be engaged in full-time employment; hence, they tend to have lower household incomes and savings (Saigal and Doyle, 2008).

Caring for the severely disabled individuals entails considerable utilisation and cost of health care resources. According to Tommiska and colleagues (2003), who studied the costs of caring extremely low birth weight babies in Finland for the period 1996-1997, the average health care cost of raising severely disabled babies in the first two years is 68-fold compared to normal healthy babies. The higher costs are related to the treatments received during hospitalisation, specialist consultation, other health care services, and parents' travel, accommodation and loss of earnings.

High incidence of severe ROP, associated with disability and rising cost of health services, represents a significant health impact in all countries. With recent developments showing improved survival rates of premature babies (Horbar et al., 2012) and large numbers of ROP babies entering the society (Blencowe et al., 2013), ROP remains an ongoing concern. Eradicating ROP would enhance the quality of life of these babies and benefit their family and the community.

2.1.4 Assessment, classification and treatment of ROP

With vast improvements in the medical care, it is now possible to diagnose ROP through retinal examination by an ophthalmologist. Currently, more hospitals in many countries are starting to impose a regulation that all babies at risk of sight-threatening ROP should be examined in screening programs. Guidelines for examining and classifying of ROP were first developed in 1984 by an international group of pediatric ophthalmologists and retinal specialists (Committee for the

Classification of Retinopathy of Prematurity, 1984). These guidelines were updated recently, which recommended that the first eye examination needs to be performed as early as 31 weeks PMA for babies born <28 weeks GA, and at four weeks after birth for more mature babies (Fierson et al., 2013). When no ROP is detected during the first visit, follow-up exams are needed every one to three weeks until the eyes are fully vascularised or worsening condition is found. As proper screening programs are not properly established in the less and moderate developing countries, some ophthalmologists follow the US guidelines, i.e. birth weight ≤ 1500 g or ≤ 30 weeks GA, or 1500 g - 2000 g or >30 weeks GA for high risk babies (American Association for Pediatric Ophthalmology, 2006), and some follow the UK guidelines, i.e. birth weight <1500 g and/ or <32 weeks GA (Wilkinson and Royal College of Ophthalmologists of the United Kingdom, 2008). However, according to Gilbert and colleagues (2005), these guidelines from two highly developed countries may not suit countries such as India, China, Latin America, and Eastern Europe, in which larger and more mature babies are developing threshold ROP. The researchers then suggested that wider screening criteria and appropriate timing of examination be adopted according to the local data.

Classification

By convention of the current guidelines (Fierson et al., 2013; American Association for Pediatric Ophthalmology, 2006; Wilkinson and Royal College of Ophthalmologists of the United Kingdom, 2008), the overall ROP status of an eye is indicated by the worst stage of retinopathy. ROP is graded into five stages of increasing severity (stages 1-5). In the early stages 1 and 2, the eyes could get better on their own or in some cases, they progress to stage 3. If untreated, the eyes are exposed to the risk of potentially blinding stages 4 and 5. The schematic view of retina in Figure 2.1 illustrates the progression of ROP from normal to mild ROP of stages 1 and 2, to severe ROP of stage 3. The quadrant labelled A shows the retinal pattern of premature blood vessels. Quadrant B shows the first stage of ROP. The blood vessels have stopped developing and the normal retina is now separated from the premature retina by a sharp line. Stage 2 ROP, as shown in quadrant C, occurs when the line of separation advances to an elevated ridge of tissue. In quadrant D, more serious stage 3 ROP is present when new blood vessels extend from the ridge toward the center of the eye. As stage 3 advances, the separation between the retina

that has blood vessels in the back of the eye, and the retina that does not have blood vessels in the front of the eye becomes even clearer. Treatment is then required to destroy the retina without the blood vessels. After the treatment, ROP could either subside on its own or continues to affect the retina to detach partially (stage 4) or totally (stage 5).

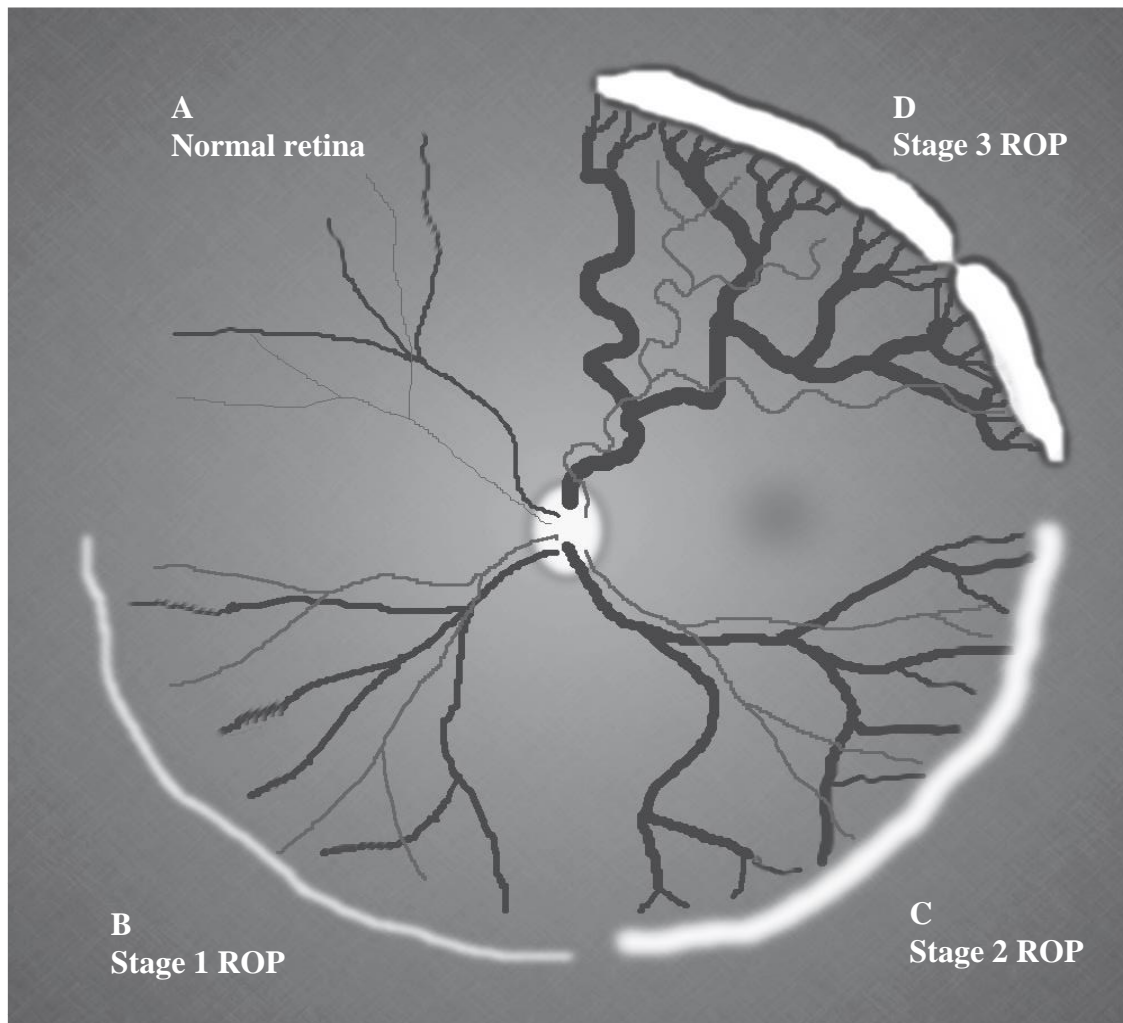


Figure 2.1: Stages of retinopathy based on a diagrammatic view of retina. The white area in the centre is the optic nerve, the grey lines are the arteries, and the black lines are the veins. (Reprinted from “UK Retinopathy of Prematurity Guideline 2008” (Wilkinson et al., 2008) with permission of the Royal College of Paediatrics and Child Health).

Treatment of established ROP

A majority of extremely preterm babies will show some degree of ROP, which sometimes resolves naturally without treatment. A small proportion, however, go on

to develop severe ROP in one or both eyes that progress to serious vision impairment when untreated. To prevent the adverse effects, treatments such as cryotherapy and laser therapy are used. The latter is often used as the current standard of care as laser-treated eyes have been reported to have better structural and visual acuity than cryotherapy (Ng et al., 2002). Recently, bevacizumab drug (Mintz-Hittner et al., 2011) was introduced in lieu of or in addition to laser treatment. The treatment of ROP with bevacizumab, however, is still at an experimental stage with patterns of recurrence and long term side effects remain to be further investigated. More evidence through randomised trials is needed to determine its safe and effective dose, and whether it is superior to laser (Darlow et al., 2013).

2.2 Pathophysiology of ROP

The development of ROP is surrounded by intricate connections with many factors including premature lung function (Askie et al., 2003), immature retinal tissue (Smith, 2005), variable tissue oxygenation (Cole, 2010) as well as demographic and postnatal factors (Darlow et al, 2005). Most recently, the role and impacts of vascular endothelial growth factor (VEGF) on the development of ROP have been explored (Chen and Smith, 2007). VEGF is excreted following from the deficiency of oxygen in the tissue and it is associated with the development of new blood vessels.

An adequate balance between oxygenation and regulation of VEGF is important for normal retinal formation (Chen and Smith, 2007; Pierce, 1996). A perturbation in oxygenation, however, will alter the delicate balance of the growth factor regulation and inevitably leads to the progression of clinical retinopathy. In early fetal life, the retina is devoid of blood vessels. After premature birth, treatment with supplemental oxygen may flood the immature retina with oxygen, resulting in higher exposure of the baby to oxygen concentration. This hyperoxia interrupts the natural hypoxia process in the retina. This leads to the downregulation of VEGF production and causes cessation of existing vessels growth. When the oxygen treatment stops, the retina becomes severely hypoxic, causing high secretion of VEGF. Such an overproduction stimulates new vessels to grow abnormally, extending from the ridge in the retina toward the center of the eye. These vessels are weak and prone to haemorrhage, scarring and contraction; thus, leading to detachments of retina and loss of vision.

2.3 Previous and current research in ROP

2.3.1 Guidelines for blood oxygen levels to minimise the risk of ROP

Monitoring blood oxygen levels is vital in neonatal care. Since 1970s and 1980s, three types of devices have been used to assess a baby's oxygenation status. These include transcutaneous oxygen electrode that requires regular calibration and can cause skin burns from prolonged use, and umbilical arterial catheter that has been considered as the optimum method for estimating blood oxygen levels. However, the latter is invasive and mostly limited to intermittent monitoring. A non-invasive pulse oximeter may be utilised to measure OS continuously. Its relative convenience of usage, which consists of a sensor that could be placed on the finger or toe of a patient, has made it as the predominant method of monitoring oxygenation in many NICUs.

Studies have found that arterial oxygen tension (PaO_2), transcutaneous oxygen (PtcO_2), and OS do not correspond linearly (Hintz et al., 2002). Hence, care must be taken when using the relationships between these measurements to interpret the readings. In particular, lower OS levels of less than 70% generally are less accurate (Giuliano and Higgins, 2005) whilst higher OS levels of 97-100% can correspond to a wide range of PaO_2 values starting from 60 mmHg to as high as 400 mmHg, or higher when a baby breathes supplemental oxygen. Likewise, a decreasing accuracy of linear relationship between PaO_2 and PtcO_2 has been observed at lower and higher readings of PaO_2 . For instance, a PtcO_2 of 45 mmHg may correspond to a PaO_2 of 50 mmHg or higher than 80 mmHg. As readings between these instruments differ, it remains a challenge to ascertain episodes of hypoxaemia, hyperoxaemia and desaturation in the oxygenation monitoring of preterm babies.

Reliable detection of the three events, as mentioned above, is crucial for avoidance of oxygen-related diseases such as ROP (Di Fiore et al., 2010) and chronic lung disease (American Academy of Pediatrics, 2002). Values of PaO_2 ranging from <40-50 mmHg have been considered to be the thresholds for defining episodes of hypoxaemia (Bohnhorst et al., 2000; Poets & Southhall, 1994) whilst an upper limit of 80 mmHg has been shown to have sufficient sensitivity to detect hyperoxaemia (Bohnhorst et al., 2002). Desaturation is the frequency of intermittent falls in OS. According to Hunt and colleagues (1999), occurrences of desaturation are common

either during short respiratory pauses or periodic breathing. OS levels of not more than 80% or 85% have been used as thresholds for defining desaturation episodes in full term babies (Hunt et al., 1999; Stebbens et al., 1991) and preterm babies (Di Fiore, 2004; Poets et al., 1992). A critical range of oxygen tension and OS in preterm babies receiving supplemental oxygen is still unknown. Flynn and colleagues (1992) later found that ROP occurred more often with oxygen tension above 80 mm Hg and thereafter, PaO₂ of 50-80 mmHg as a safe range to avoid ROP was accepted by professional consensus (American Academy of Pediatrics, 2002). Another study on the role of OS in fetal growth reported that normal fetuses develop with OS of 70-80% in utero (Nicolini, 1990); many clinicians, however, are relying on 85-95% in monitoring preterm babies (Cole et al., 2003).

With limited information on the link between oxygen tension and OS, efforts have been made to transpose oxygen tension of 50-80mm Hg into equivalent OS values. These included targeting a higher range of 90-95% (Vijaykumar, 1997) and a lower range of 80-95% (Tin et al., 2001; Anderson et al., 2004; Sun et al., 2002; Chow et al., 2003) shortly after birth. Outcomes from these observational studies suggested that retinal surgery may reduce by 61-100% if a lower range of OS policy is adopted. Such a recommendation, however, was met with a concern because lower saturation ranges may compromise neuro-developmental and survival outcomes (Cole et al., 2003). To guide future clinical practice in the administration of supplemental oxygen to newborn babies, a group of collaborated randomised control trials from nine countries are studying the effects of two different ranges of OS levels on the health of very preterm babies less than 28 weeks GA (Cole et al., 2003; Askie et al., 2011). Recent results from these trials suggested that incidence of severe ROP was reduced with lower saturation ranges below 90%; however, the risk of mortality was increased (SUPPORT Group, 2010; Vaucher et al., 2012; Schmidt et al., 2013; BOOST II Group, 2013).

2.3.2 Oxygen fluctuations and ROP

The role of oxygen fluctuations in the development of ROP was not fully understood until animal studies were carried out in the early 1950's (Ashton et al., 1954). Rats, cats, mice and rabbits were often used to study the effects of alternating or clustering of oxygen fluctuations on retinal development as their retinal structure at

birth is closely comparable to that in preterm babies. Mimicking oxygen-induced retinopathy, the newborn animals were exposed to varying concentrations of oxygen for continuous or intermittent periods. These studies later gave rise to other animal models of retinopathy and studies involving extremely premature babies as young as 23 weeks GA (Di Fiore et al., 2010).

Clinical human and animal studies have shown that oxygen fluctuations, hypoxia and hyperoxia in the early weeks of life during oxygen therapy increased the risk for severe or threshold ROP. PtcO₂ variability was reported to be a significant predictor of severe ROP in the first two weeks of life (Cunningham et al., 1995), and increased fluctuations of PaO₂ may expose premature babies to a higher risk of developing threshold ROP (Saito et al., 1993; York et al., 2004). In experiments done in newborn rats, alternating hypoxia and hyperoxia caused severe retinopathy (Werdich et al., 2004; Penn et al., 1994; McColm et al., 2004; Coleman et al., 2008). Severity is worsened when fluctuations in oxygen occurred in wider ranges (Penn et al., 1995), hypoxic clusters (Coleman et al., 2008), small increments (Werdich et al., 2004), or repeated hypoxic periods (McColm et al., 2004). These studies illustrate a plausible explanation that increased hypoxic episodes may be associated with the development of severe ROP, a finding which was later confirmed by Di Fiore and colleagues (2010). In their study, the babies in the group of ROP sufficiently severe to require laser therapy (LaserROP) had increased frequency of hypoxaemic events as well as longer period of supplemental oxygen compared with babies in the NoLaserROP group through the first four to five weeks of life. Because longer periods on oxygen therapy may subject a baby to greater episodes of hypoxia, hyperoxia and oxygenation fluctuations (Tin et al., 2001; Cole, 2010), clinical practices accompanied by frequent adjustments of supplemental oxygen need to be eradicated.

2.3.3 Postnatal and clinical determinants of ROP

ROP is a disease of prematurity that was discovered about seven decades ago after oxygen therapy was introduced into the newborn nursery. At that time, it was also noted that all of the initial patients weighed under 1600 g (Kinsey et al., 1956; Kinsey et al., 1977). The use of oxygen is now closely monitored and is no longer implicated as the sole or even the primary factor of ROP. Nevertheless, serious

retinopathy persists until today, with prematurity (Darlow et al., 2005) and birth weight (Good et al., 2005) remain as the dominant risk factors of this vision threatening disorder.

Many clinicians believe that events in the postnatal period of premature babies may be critical in determining the onset of ROP. A multitude of factors have been investigated as having associations with ROP. Whilst prematurity, low birth weight, and the use of oxygen therapy (McColm and Fleck, 2001) have been consistently found as the strongest risk factors of ROP, recent studies have also suggested that there are close relationships between postnatal weight gain (Wikstrand et al., 2011; Hellstrom et al., 2009), insulin-like growth factor (IGF-1) (Hellstrom et al., 2003) and severe ROP. Using a screening algorithm, these investigators demonstrated that poor development in weekly weight gain and IGF-1 serum levels that prolonged provide an early indication of babies at high risk of severe ROP. IGF-1 has been shown to be important for development of normal retina (Hellstrom et al., 2003), and growth of fetal and neonate through regulation of insulin glucose levels (Giudice et al., 1995). However, high overall glucose levels during the first month of life were found as a predictor of severe ROP (Chavez-Valdez et al., 2011). Infection is another area of concern in the neonatal development as it is a major cause of mortality and morbidity (Hofstetter et al., 2007; American Academy of Pediatrics, 2002). A meta-review by Bharwani and Dhanireddy (2008) showed that systemic infection such as candida sepsis is associated with a higher proportion of ROP severe enough to require surgical treatment.

There have been other clinical factors that have not been consistently confirmed with univocal results or clearly proven in large studies, but have been postulated to be associated with ROP. Among factors that have been studied are genetic disease, multiple gestation pregnancies, maternal bleeding, intakes of iron, vitamins E and A, breast milk feedings, light exposure. There are also several reports on the damaging effects of high carbon dioxide levels (Hauspurg et al., 2011) and frequent blood transfusions (Romagnoli, 2009) on exacerbating the severity of ROP. Although the findings are suggestive, additional investigations are warranted if these events persist in the neonatal care.

Gender has been identified as one of the risk factors since an early report of ROP in 1942, where male babies were noted to be more likely to acquire the disease than female babies. Whilst this outcome has been confirmed by several studies in the US (Di Fiore et al., 2010; Jacobson et al., 2009) and a large network in Australia and New Zealand (Darlow et al., 2005), others found no relationship between male gender and severity of ROP (Gunn et al., 2012; Austeng et al., 2010). The finding that male babies are prone to the risk of ROP is in accordance with some earlier evidence that being male is associated with increased chances of mortality and morbidity at extreme prematurity (Synnes et al., 1994). Likewise, babies with prenatal growth restriction have been suggested to be more susceptible to these adverse outcomes. Currently, many clinicians are starting to study whether restriction in intrauterine growth could have contributed towards the development of the eye disease. Consequently, there appears to be an increasing inclination to use weight-for-GA rather than birth weight as the basis of cohort selection (Qiu et al., 2012; Regev et al., 2003) since the former has been shown to provide a clearer relationship between morbidity, GA and birth weight (Synnes et al., 1994). The benefit of utilising weight-for-GA over birth weight, however, is still debated as not all studies consistently found similar associations, which in part, could be due to various criteria being used for defining weight-for-GA itself.

The efficacy of steroids on reducing mortality, morbidity and improving lung maturation in fetal and preterm babies has been reviewed by Halliday and colleagues (2003) based on 21 randomised controlled trials. Both prenatal and postnatal steroids appear to be associated with a reduction in the severity of ROP (Console et al., 1997; Jagielska et al., 2008). The protective effect of the latter, however, is demonstrated only when administered within 4 days of birth (Halliday et al., 2003), which also resulted in adverse side effects of body systems and neuro-developmental impairment. Timing of administration again plays an important role when a study reported that any late treatments given after three weeks of life or those that prolonged for more than two weeks may pose an increased risk of severe ROP in very low birth weight babies (Karna et al., 2005). A large multi-center randomised trial has been suggested to gain a consensus on recommended timing and dosage for both short and long-term use of steroids in clinical practice.

2.4 NZ-ROP study

The influence of clinical determinants of ROP in preterm babies is multifaceted (McColm and Fleck, 2001). Many of them are inter-related, and have been found to be associated not only with ROP but also with other various complications of prematurity. The simultaneous presence of these factors seems to indicate babies' underlying conditions of being more ill from severe disease status.

Reducing the severity of ROP is extremely important in protecting very premature babies from the risk of vision impairment or blindness. More studies are beginning to shed light on the many factors involved in the disease; in particular, fluctuations of blood oxygen levels of PaO₂ (York et al., 2004) and PtcO₂ (Cunningham et al., 1995). However, the role of OS, combined with other factors, on the development of ROP is still less clear today. A study, NZ-ROP, was then undertaken to examine whether some measures of OS could provide some insight on the link between fluctuations of OS and severity of ROP. NZ-ROP was analysed using a subset of data from the BOOST-NZ trial.

2.4.1 BOOST-NZ trial

The BOOST-NZ is the New Zealand arm of a group of collaborated international randomised controlled trials (described in Section 2.3.1) for investigating the optimal range of OS in extreme premature babies being treated with supplemented oxygen (Darlow et al., 2014). Specifically, the aim of the trials is to determine whether targeting the lower range 85-89% compare to upper range 91-95%, beginning within 24 hours of birth until 36 weeks PMA, leads to any difference in outcomes of death or survival with neuro-developmental disability diagnosed at two years of age. Additionally, postnatal growth and several neonatal illnesses such as ROP are investigated as the secondary outcomes.

The trial received approval from the Multi-Region Ethics Committee and obtained written consent from all parents. Enrolment of babies started in September 2006 and ended in December 2009. Recruitment of extremely preterm babies was undertaken at five NICUs in New Zealand. Some postnatal data were recorded upon enrolment such as gestation age and weight at birth, and gender whilst other variables were documented prospectively until the baby reached 36 weeks PMA or had an early

discharge, for instance, survival, OS, duration of oxygen therapy and neonatal illnesses. To achieve the trial's objectives, each enrolled baby was randomly assigned with a lower-target group with OS target of 85-89% or a higher-target group with OS target of 91-95%, and, therefore was allocated with one of two types of pulse oximeters to record the OS readings. These oximeters were specially programmed such that the readings between 85-95% were offset to either higher or lower than the actual saturation to blind the treatment allocation to family, clinical staff and outcome-assessors. Details of these offsets are described in Chapter 3, Section 3.3.

NZ-ROP has different objectives from the BOOST-NZ, although the data were taken from the latter. Whilst the BOOST-NZ attempted to assess the effect of two different OS ranges on the health of preterm babies, NZ-ROP is focused on identifying the potential relationships between four statistical summaries of OS, several postnatal factors and ROP alone. The differences in these objectives between the two studies have some implications on the OS data. Because the readings of the oximeters were influenced by the trial-imposed offsets, these effects must be removed before the OS data can be used in the NZ-ROP analysis. The procedures to adjust these offsets are detailed in Section 3.4.

2.4.2 NZ-ROP data

There were 340 premature babies that were enrolled in the BOOST-NZ. Babies were eligible for inclusion in NZ-ROP if they had been admitted to the NICUs between September 2006 and December 2009 with gestation less than 28 weeks, were less than 24 hours old, and were alive until 36 weeks PMA. Babies were excluded if they had known congenital anomalies affecting oxygenation. A total of 83 of the eligible babies did not meet the inclusion criteria for the following reasons: they were deceased before 36 weeks PMA, had not reached the end point of the eye examination, were still depending on supplemental oxygen, had withdrawn from the study before 36 weeks PMA, or had no OS data for at least two weeks after 27 weeks PMA. Overall, a total of 257 babies who met the inclusion criteria of NZ-ROP (Figure 2.2) and had the most complete information were included in the retinopathy study. Of these, 38 babies developed at least stage 3 ROP, and another 3 developed stage 2 ROP but were treated with laser retinal surgery. This resulted in 13.6% (or 41/301) of

severe ROP cases. The ROP status, clinical and postnatal information as well as the OS data for these 257 babies were utilised in the NZ-ROP analysis.

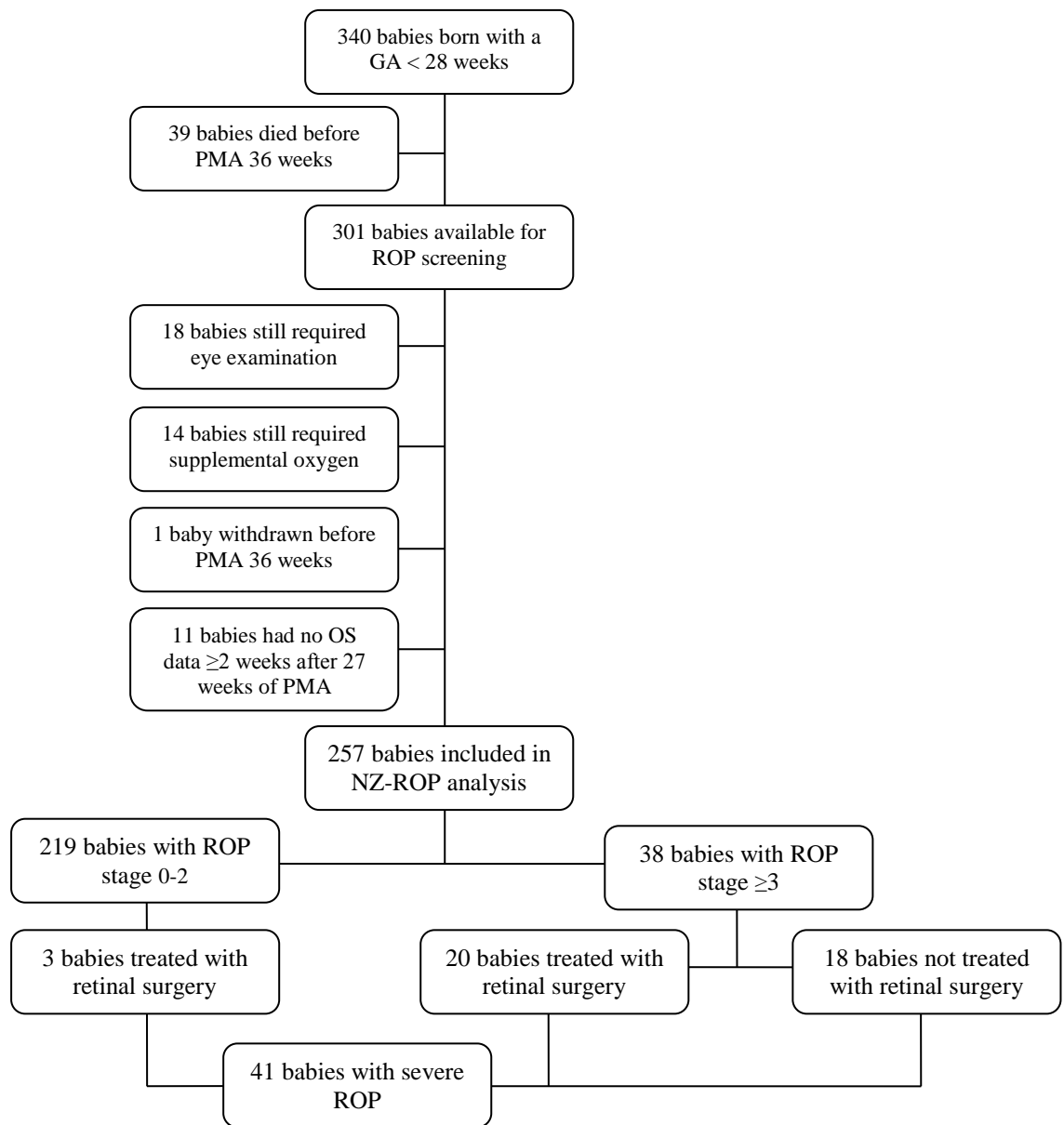


Figure 2.2: Flowchart of study population in NZ-ROP and ROP outcome

Oxygen saturation

The BOOST-NZ trial used specially adjusted and masked Masimo SET pulse oximeters (Masimo Corporation, Irvine, CA, United States) for comparing the effects of OS targeting in two groups of preterm babies. Within 24 hours after a baby was born and parental consent was obtained, randomisation of oximeter was performed and the oximeter with labelled serial number was attached to the baby. Data

acquisition commenced once a good waveform was obtained and OS was recorded once every 10 seconds for a continuous duration until 36 weeks PMA or discharged, whichever occurred earlier. The data was downloaded at least every fortnightly. The allocated oximeter must remain with the baby even when the baby is transferred to another NICU while still in oxygen.

The OS readings from the oximeter were used as the basis of computation for four OS measures that have been chosen as the severe ROP predictors: mean, SD, CV and desaturation of OS levels. These measures were computed for each baby after adjustments were made on the OS readings to remove the effects of the trial-imposed offsets (refer Section 3.4), and a suitable time interval was chosen (refer to Section 4.2 and Section 5.2). All subsequent analyses for NZ-ROP were based on the adjusted data.

Status of ROP

The babies' eyes were examined according to the BOOST-NZ protocol by experienced ophthalmologists to monitor and detect for any signs of ROP. The stage of ROP was recorded during these eye examinations, which started when the babies were at the age of four to six weeks after birth, and thereafter fortnightly until the eye was considered to be no longer in danger of developing severe ROP. These are when the eyes are fully vascularised or the vessels have reached zone III. For this study, the status of ROP (denoted as severe or no severe ROP) for each baby was recorded according to the most advanced ROP stage observed throughout all visits. Severe ROP was defined as stage 3 and above or had undergone laser therapy whilst no severe ROP was defined as stage 2 and below including no diagnosis of ROP.

2.6 Summary

ROP continues to be a significant visual impairment threat in very preterm babies. Incidence of ROP is high among the most immature and smaller babies in developed nations, but has now emerged among older and bigger babies in middle-income countries and urban areas of low-income countries (Gilbert et al., 2005). The use of supplemental oxygen that was initially implicated as the primary cause of ROP is now being closely monitored whilst prematurity and birth weight remain as the strongest risk factors of ROP (McColm and Fleck, 2001). There is also a growing

evidence of close relationships between ROP and male gender, poor postnatal weight gain, prenatal growth retardation, IGF-1 and VEGF (Hellstrom et al., 2003; Hellstrom et al., 2009). Guidelines for recommended screening criteria and timing of examination according to babies' birth weight and GA are adopted in various parts of the world to detect high-risk babies. Thereby, this would enable them to receive appropriate and timely treatment as needed. Eyes that progressed to serious ROP are treated with retinal surgery such as cryotherapy or laser therapy (American Association for Pediatric Ophthalmology, 2006; Wilkinson and Royal College of Ophthalmologists of the United Kingdom, 2008).

Recent developments have indicated that stable IGF-1 and VEGF (Hellstrom et al., 2003; Hellstrom et al., 2009), reduced fluctuations of blood oxygen levels (Cole, 2010), and close monitoring of oxygen therapy (Schmidt et al., 2013; BOOST II Group, 2013) in the early neonatal life are important for the development of normal retinal formation. Any slight perturbations in oxygenation will alter the delicate balance of the growth factors' regulation and inevitably leads to the progression of clinical retinopathy. In animal models of induced retinopathy, varying concentrations of oxygen for continuous or intermittent periods in the form of alternating, clusters, wider ranges, small increments and repeated episodes exacerbate the severity of ROP. However, there is still no clear answer of what range of OS is optimal for newborn babies receiving supplemental oxygen. This issue is currently being studied by a group of international trials (Askie et al., 2011).

Further studies are also needed to shed light on the link between OS fluctuations and ROP. A study (NZ-ROP) was conducted to determine whether measures of variability and saturation levels of OS are related to the development of severe ROP. This investigation involved 257 babies born <28 weeks GA that were enrolled in the BOOST-NZ trial (Darlow et al., 2014). Data consisted of ROP status, OS data and postnatal information. The latter includes the age and weight of the babies at birth, gender, and details of some treatment for oxygen therapy. The eyes examinations were conducted by trained ophthalmologists until the eyes were fully vascularised. The status of ROP for each baby was classified as severe ROP or no severe ROP, as indicated by the worst stage of ROP observed throughout all visits. The OS readings were recorded continuously within 24 hours after birth until 36

weeks PMA from Masimo pulse oximeters. The BOOST-NZ, which has its own separate objectives from NZ-ROP, had used specially programmed oximeters for studying the effects of two types of oxygen intervention. These modifications had resulted in the OS readings between certain ranges to be displayed higher or lower than the actual saturation. As such, some mechanism is needed to render the data suitable for use in NZ-ROP.

High incidence of severe ROP, associated with disability and rising cost of health services, represents a significant health impact in all countries. ROP affects the quality of life of the children and inflict a long-term burden on their families, society, and health care resources (Saigal and Doyle, 2008). Various measures have been developed such as providing better antenatal care to reduce the rate of preterm birth. There have also been improvements in the neonatal care practices to reduce the risk of ROP blindness through adequate human and equipment resources, timely diagnostic screening and examinations with appropriate treatment to prevent progression to blindness, and identification of criteria for determining babies at high risk of severe ROP (Bleincowe et al., 2013; Gilbert et al., 2005). Such interventions in managing these most vulnerable patients can contribute significantly to their visual diagnosis.

CHAPTER 3

PRE-PROCESSING OF OXYGEN SATURATION DATA FOR NZ-ROP

3.1 Introduction

Recruitment of babies for the BOOST-NZ trial was completed at the end of 2009. The data, which were collected from five hospitals, were received in two batches - November 2009 and August 2010. There were three main components (refer to Chapter 2, Section 2.6) of the data: the clinical data, the ROP status, and the OS data from the oximeter. These data were processed before they could be used for further analysis. Data checking and verification were conducted for the first two components.

The OS data were processed in three stages: data clean-up, data adjustments and data summary. The first stage mainly dealt with the inspection and organisation of the data, including accounting for signal quality. Data adjustments stage was a crucial stage involving the identification of high and low saturation oximeters used for the BOOST-NZ trial, removal of the trial-imposed offsets and correction of artefact from the Masimo oximeters. In the final stage, summary statistics were computed, which formed part of the predictors for the regression analysis in Chapter 5.

3.2 Data clean-up

The readings from the oximeter were recorded within 24 hours after birth until 36 weeks PMA or early discharge from the hospital. They were uploaded about once a fortnight to a file, which was named according to the baby's case number and the upload date. The files were checked for consistency in the filename and content. Whenever necessary, re-labelling of filenames, organisation of data for each baby, and systematic removal of mismatched or duplicated readings were performed. After the database was cleaned and properly organised, the quality of the OS values was accounted for. Readings with low signal quality indicated by the oximeter, such as 'low indicator quality signal', 'low perfusion', or 'sensor off' were discarded. Several factors could have contributed to the low signal quality, for instance, malpositioning

of oximeter's probe due to a patient movement, low blood flow to the measurement site, or unreliable readings during the initial start-up of the oximeter. Upon advice from a pediatrician, readings less than 50% were also omitted, as they are unrealistic because no preterm baby would normally survive with such low OS. An algorithm that removed the noise in the OS readings and assembled all data is provided in Appendix A.1. The following flowchart summarises the clean-up process performed for each baby.

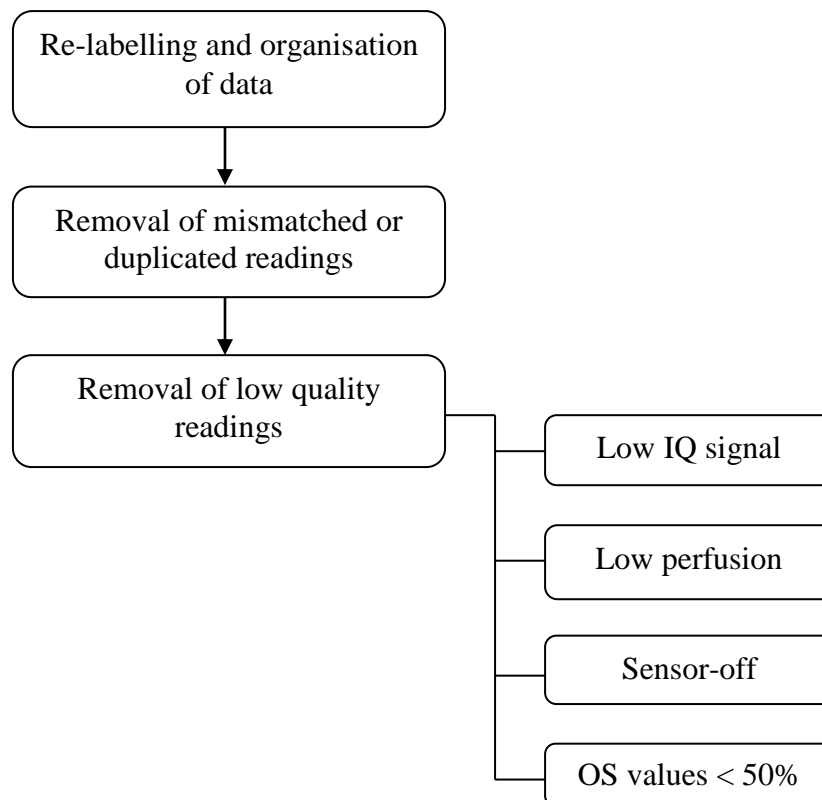


Figure 3.1: Flowchart of data clean-up procedure

3.3 Operation of Masimo oximeters for BOOST-NZ

The oximeters used for the BOOST-NZ trial had two features that were addressed before the OS data could be used for the ROP-NZ study. The first was the trial-imposed offsets for the OS readings. The oximeters used in the BOOST-NZ trial had been altered to read either higher or lower than the true saturation for readings between 85-95%. A low saturation oximeter displayed 88-96% when the saturation was actually lower at 85-95%, that is, the displayed reading is the true reading plus 3% but bounded above at 96%. Similarly, a high saturation oximeter displayed 84-92% when the saturation was actually higher at 85-95%, that is, the displayed reading

is the true reading minus 3% but bounded below at 84%. To prevent unmasking, the oximeters displayed a transition from true to offset values and vice versa. The transition always occurred incrementally over several seconds, showing intermediate values between 85-87% for a low saturation oximeter and between 93-95% for a high saturation oximeter.

A second feature of the BOOST-NZ oximeters was an algorithmic artefact that made it less likely for OS values between 87-90% to be displayed. Instead, they were displayed either as higher or lower (mainly higher) values surrounding the region. As a consequence, the true values in the region were moved to either side of the region. This, in turn, created a 'dip' in the frequencies for the affected region.

As the offsets and the dip region masked the true OS values, it is important that the affected values be adjusted to reflect what would have been actually read. These adjustments must be made before further analysis can be carried out.

3.4 Adjustments for offsets and dip artefact

The algorithms to implement these adjustments were developed in two stages. The first stage entailed the identification of the oximeter type (high or low saturation) for each baby. In the subsequent stage, adjustments were made for the offsets and the dip artefact according to the type of the oximeter. The algorithms for the adjustments are described in the next two sub-sections, 3.4.1 and 3.4.2.

3.4.1 Procedure for identifying high or low saturation oximeters

To unmask the oximeter's type, the OS readings were summarised as relative frequencies and plotted for each baby. Upon careful inspection of the relative frequency plots, two distinct patterns for the two types of oximeters were identified. For the first type, there was an almost-flat region at saturations 85-87% caused by the transition mechanism (Section 3.3), and a notable peak at 96% (Figure 3.2).

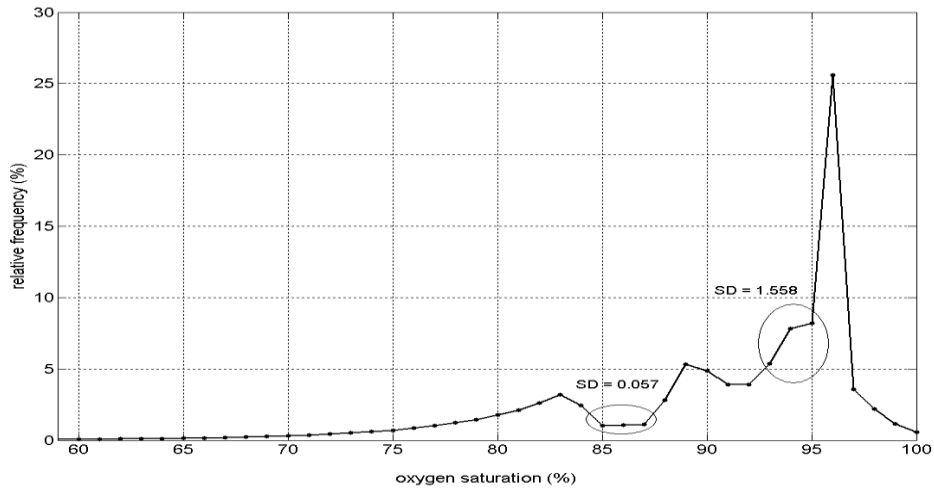


Figure 3.2: A typical OS pattern for low saturation oximeters: SD(85-87%) is smaller than SD(93-95%) and a peak occurs at 96%

For the second type, on the other hand, the almost-flat region occurred at 93-95% and a peak at 84% (Figure 3.3). From these patterns, it was apparent that the first corresponded to low saturation oximeters, and the second to high saturation oximeters. These patterns were used to distinguish the oximeter’s type. The resulting offset identifications were further verified by checking the serial number of the oximeter for each baby. To be consistent, babies using the same oximeter should be identified as having the same offset.

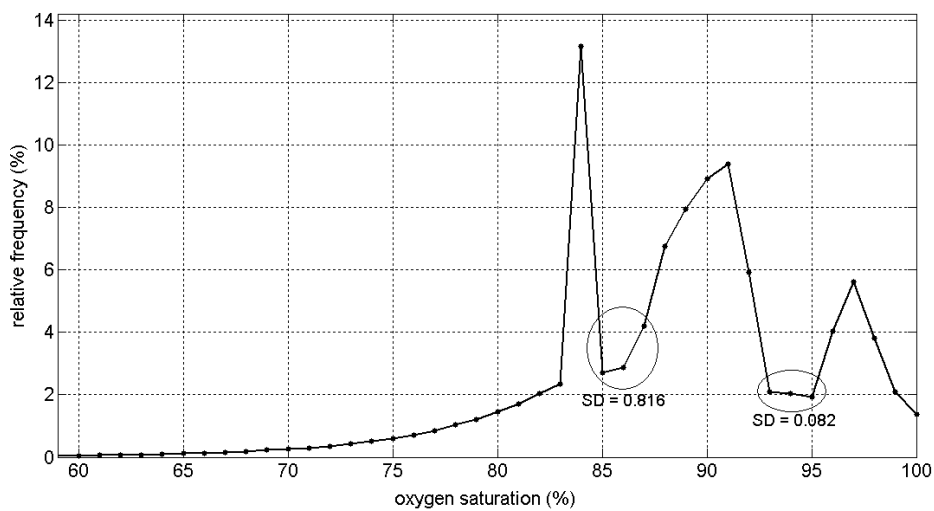


Figure 3.3: A typical OS pattern for high saturation oximeters: SD(93-95%) is smaller than SD(85-87%) and a peak occurs at 84%

The procedure to identify the offsets was summarised as follows:

- Calculate the standard deviation (SD) of the relative frequencies at saturations 85-87% and 93-95%.
- Assign a high saturation oximeter if a peak occurs at 84% saturation and $SD(85-87\%) > SD(93-95\%)$; else assign a low saturation oximeter if a peak occurs at 96% saturation and $SD(85-87\%) < SD(93-95\%)$; else treat as unassigned.
- Ensure consistency of the assigned cases by checking that all babies using the same oximeter have the same offset.
- Assign each of the unassigned cases as high or low by matching its serial number with those in the assigned group. Note that there were no unassigned cases without matching serial numbers in the assigned group.

3.4.2 Algorithm for adjusting offsets and dip artefact

Having identified the type of the oximeter, the next step was to develop and implement algorithms for adjusting the offsets and the ‘dip’ artefact. Separate algorithms were needed depending on the oximeter type. The details for the two algorithms are as follows:

(i) Algorithm for low saturation oximeters (offset is +3%)

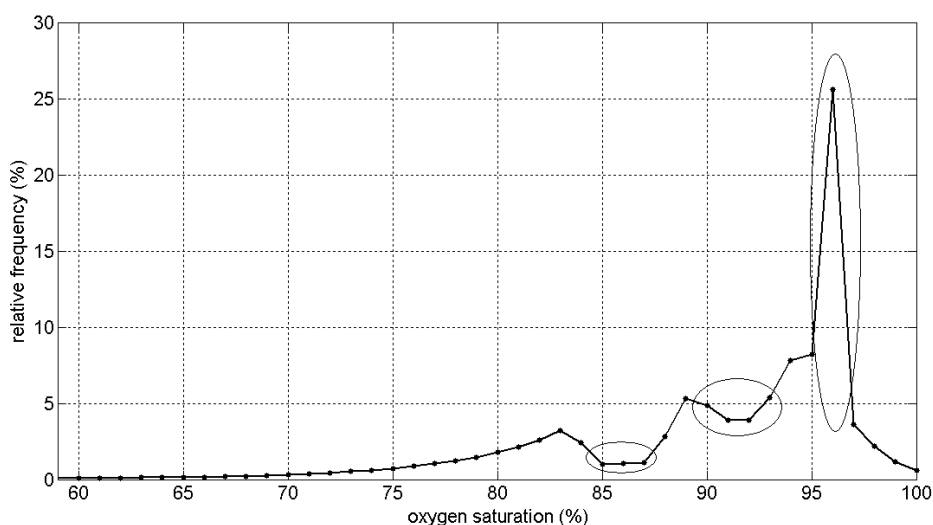


Figure 3.4: Effects of offset and ‘dip’ artefact, as indicated by the circles, for low saturation oximeters

Figure 3.4 depicts the distribution of the OS readings from a baby that was assigned to a low saturation oximeter. Recalling how the offset for a low saturation oximeter was achieved as well as the ‘dip’ artefact presence in the oximeter, the following features are relevant. Firstly, there was a three-point flat region at saturations 85-87% due to the transition mechanism (described in Section 3.3). Secondly, an unusual peak at 96% was also observed, resulting from true saturations of 93-95% being returned as 96% along with the true saturation of 96%. Thirdly, the ‘dip’ region at 87-90% now occurred at 90-93% because of the offset. Therefore, a three-step algorithm was developed to estimate the true readings. Each step consists of a linear or quadratic interpolation constrained by two reliable end points, and by an appropriate conservation of relative frequencies. Using r_k and \tilde{r}_k to denote the displayed and adjusted relative frequencies, respectively for saturation k . The algorithm for low saturation oximeters is outlined as follows:

Step 1: Adjust the three-point flat region.

Estimate \tilde{r}_{84} and \tilde{r}_{85} by quadratic interpolation constrained by true relative frequencies at saturations 83% and 86% (i.e. r_{83} and r_{89} before the offset removal), and redistributing the relative frequencies in $r_{84:88}$ to $r_{84:85}$ (see Figure 3.5(a)). Hence, solve

$$\begin{aligned} r_{83} &= 83^2 a + 83b + c, \\ r_{89} &= 86^2 a + 86b + c, \\ \sum_{k=84}^{88} r_k &= (84^2 + 85^2)a + (84 + 85)b + 2c, \end{aligned} \tag{3.1}$$

simultaneously for a , b and c , and use them to estimate r_{84} and r_{85} :

$$\tilde{r}_k = k^2 \hat{a} + k\hat{b} + \hat{c}, \quad k = 84, 85. \tag{3.2}$$

Next, remove the +3% offset to recover the relative frequencies at saturations 86% to 92% by setting $\tilde{r}_{86:92} = r_{89:95}$, (Figure 3.5(b)).

Step 2: Estimate $\tilde{r}_{87:90}$ to adjust for the ‘dip’ artefact. Solve for a and b from equations

$$\begin{aligned} \tilde{r}_{86} &= 86a + b \\ \tilde{r}_{91} &= 91a + b \end{aligned} \tag{3.3}$$

and use these estimates to interpolate linearly the relative frequencies in the ‘dip’ region (Figure 3.5(c)),

$$\tilde{r}_k = k\hat{a} + \hat{b}, \quad k = 87, \dots, 90. \quad (3.4)$$

Step 3: Adjust the region near the peak at 96%.

Estimate $\tilde{r}_{93,96}$ by quadratic interpolation constrained by true relative frequencies at saturations 92% and 97%, and redistributing the relative frequency in r_{96} to $\tilde{r}_{93,96}$ (see Figure 3.5(a)). Hence, solve

$$\begin{aligned} \tilde{r}_{92} &= 92^2 a + 92b + c, \\ r_{97} &= 97^2 a + 97b + c, \\ r_{96} &= \left(\sum_{k=93}^{96} k^2 \right) a + \left(\sum_{k=93}^{96} k \right) b + 4c, \end{aligned} \quad (3.5)$$

simultaneously. Use the estimates of a , b and c from the three equations to interpolate for $\tilde{r}_{93,96}$ (Figure 3.5(d) and Figure 3.5(e)),

$$\tilde{r}_k = k^2 \hat{a} + k\hat{b} + \hat{c}, \quad k = 93, \dots, 96. \quad (3.6)$$

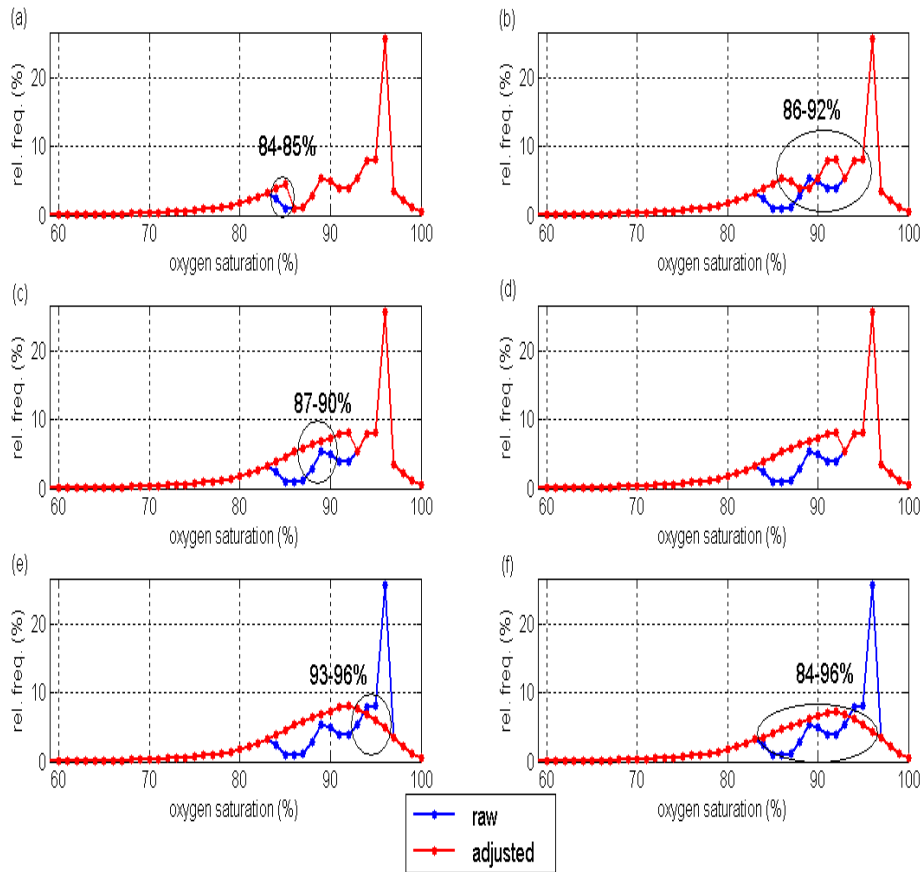


Figure 3.5 Corrections for offsets and dip artefact for low saturation oximeters

(ii) Algorithm for high saturation oximeters (offset is -3%).

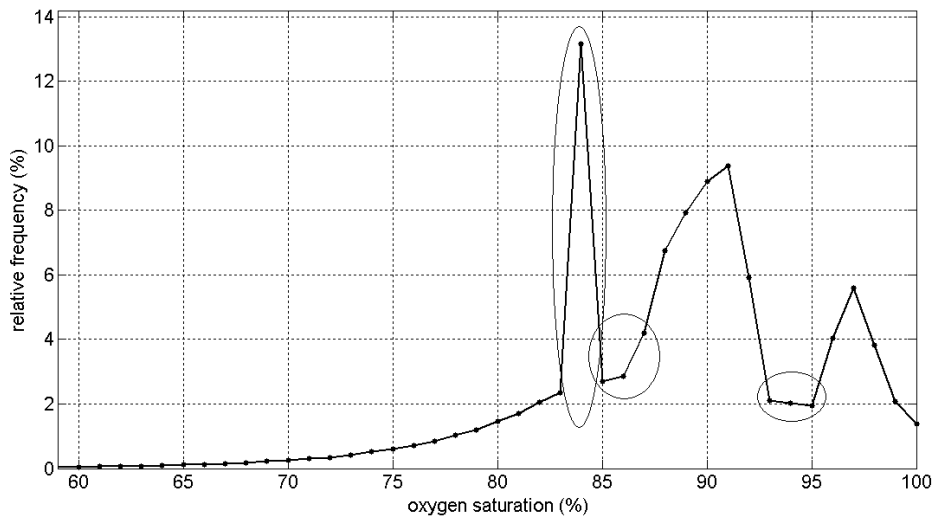


Figure 3.6: Effects of offsets and 'dip' artefact, as indicated by the circles, for high saturation oximeters

Similar to the low saturation oximeter, the high saturation oximeter also has the peak, ‘dip’ artefact, and the three-point flat region. However, the three features are located at different saturations (Figure 3.6). A peak was noted at 84%, resulting from true saturations of 85-87% being returned as 84% along with the true saturation of 84%. This peak corresponds to the true reading of 87%; therefore, comprising part of the ‘dip’ region (87-90%). The three-point flat region, which is the effect of masking the true saturations, is depicted by the circle at 93-95%. An algorithm to adjust for these features is described below along with the results of the implementation (Figure 3.7).

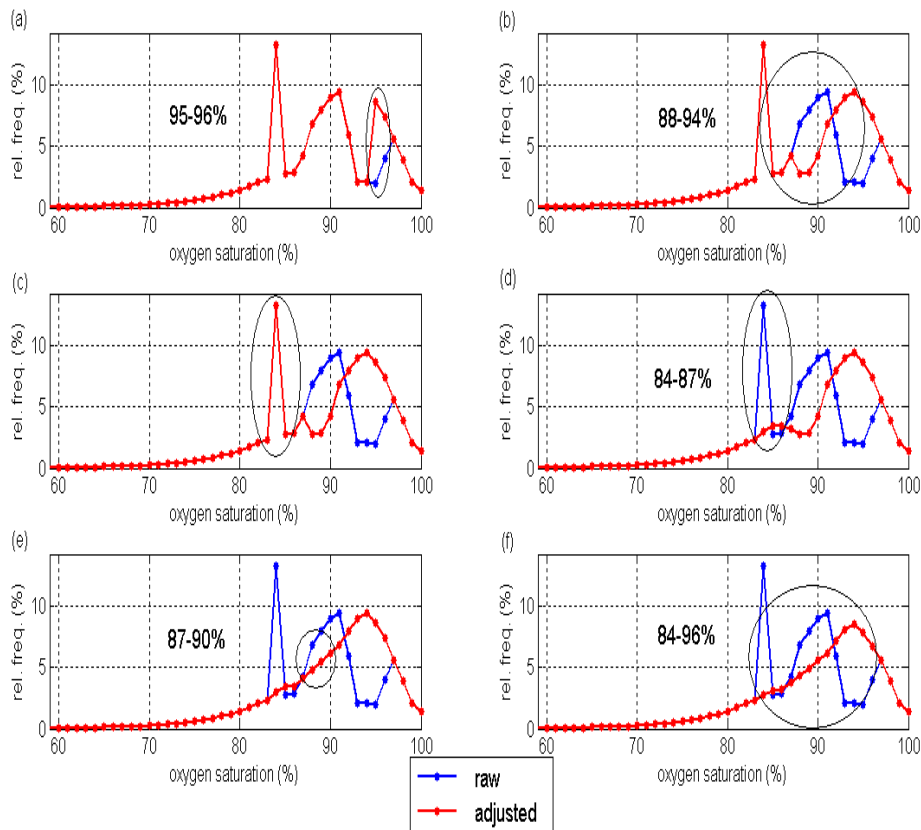


Figure 3.7: Results of corrections for offsets and dip artefact for high saturation oximeters

Step 1: Adjust the three-point flat region.

Adjustment for the region within the three-point values at 93-95% involves estimating \tilde{r}_{95} and \tilde{r}_{96} by quadratic interpolation constrained by true relative frequencies at saturations 94% and 97% (i.e. r_{91} and r_{97} before the offset removal), and redistributing $r_{92:96}$ to $\tilde{r}_{95:96}$ (see Figure 3.7(a)). Hence, solve

$$\begin{aligned}
r_{91} &= 94^2 a + 94b + c, \\
r_{97} &= 97^2 a + 97b + c, \\
\sum_{k=93}^{96} r_k &= (95^2 + 96^2)a + (95 + 96)b + 2c,
\end{aligned} \tag{3.7}$$

simultaneously for a , b and c , and use them to estimate \tilde{r}_{95} and \tilde{r}_{96} ,

$$\tilde{r}_k = k^2 \hat{a} + k\hat{b} + c, \quad k = 95, 96. \tag{3.8}$$

Next, remove the -3% offset to retrieve the true relative frequencies at saturations 88% to 94% by setting $\tilde{r}_{88:94} = r_{85:91}$, (Figure 3.7(b)).

Step 2: Adjust the region near the peak at 84%.

Estimate $r_{84:87}$ by quadratic interpolation constrained by true relative frequencies at saturations 83% and 88%, and redistributing the relative frequency in r_{84} to $\tilde{r}_{84:87}$.

Hence, solve equations

$$\begin{aligned}
r_{83} &= 83^2 a + 83b + c, \\
\tilde{r}_{88} &= 88^2 a + 88b + c, \\
r_{84} &= \left(\sum_{k=84}^{87} k^2 \right) a + \left(\sum_{k=84}^{87} k \right) b + 4c,
\end{aligned} \tag{3.9}$$

to estimate a , b and c . Use them to interpolate for $\tilde{r}_{84:87}$ (Figure 3.7(c) and Figure 3.7(d)),

$$\tilde{r}_k = k^2 \hat{a} + k\hat{b} + c, \quad k = 84, \dots, 87. \tag{3.10}$$

Step 3: Repeat step 2 from the algorithm for low saturation oximeters to adjust for the relative frequencies in ‘dip’ region, i.e. $\tilde{r}_{87:90}$ (Figure 3.7(e)).

Finally, for both algorithms, normalise all relative frequencies from 84% to 96% (Figure 3.5(f)) and Figure 3.7(f)),

$$\tilde{r}_k = \tilde{r}_k \cdot \frac{\sum_j r_j}{\sum_j \tilde{r}_j}, k = 84, \dots, 96. \quad (3.11)$$

The MATLAB program for the implementation of the two algorithms is provided in Appendix A.2.

3.4.3 Validation of the adjustment method

The algorithms were validated using data sets obtained from the United Kingdom. These data were OS readings from an extremely preterm baby, who received ventilation and supplemental oxygen during the study. The baby was monitored for two consecutive periods totaling about 24 hours. In the first period, OS readings were taken using a low saturation Masimo oximeter, while in the second period, a high saturation Masimo oximeter was used. For both periods, a non-offset Siemens oximeter was also used as benchmark.

Each algorithm (refer Sections 3.4.1 and 3.4.2) was then implemented on the corresponding offset Masimo readings, and the results were compared with the Siemens recordings. Figures 3.8 and 3.9 show the unadjusted (blue line) and adjusted data (red line) alongside the Siemens data (brown line). It can be seen that the adjusted data are remarkably close to the Siemens data. Using the algorithms, all of the three features – the peak, ‘dip’ artefact, and three-point flat region – that appear in the lower and higher saturation target oximeters have been adjusted successfully to determine the true OS recordings.

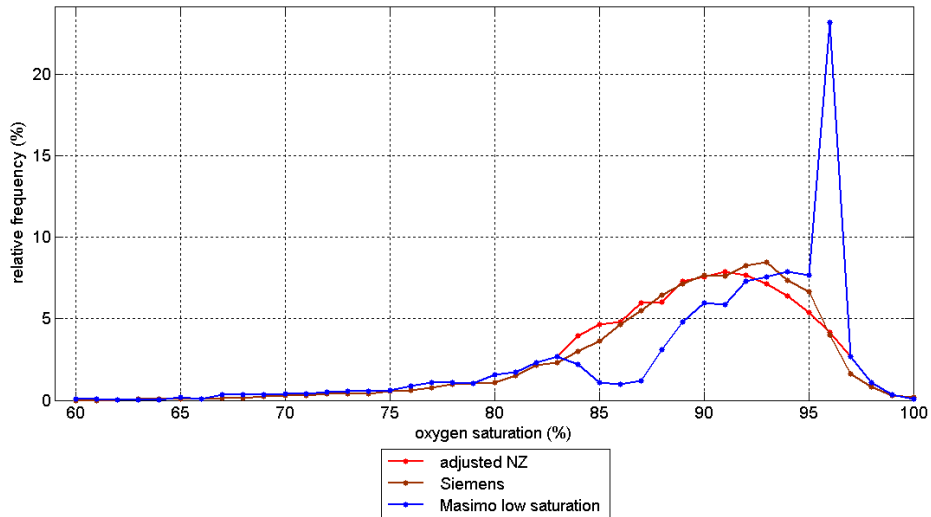


Figure 3.8: Comparison of unadjusted and adjusted readings from a Masimo low saturation oximeter and a Siemens oximeter

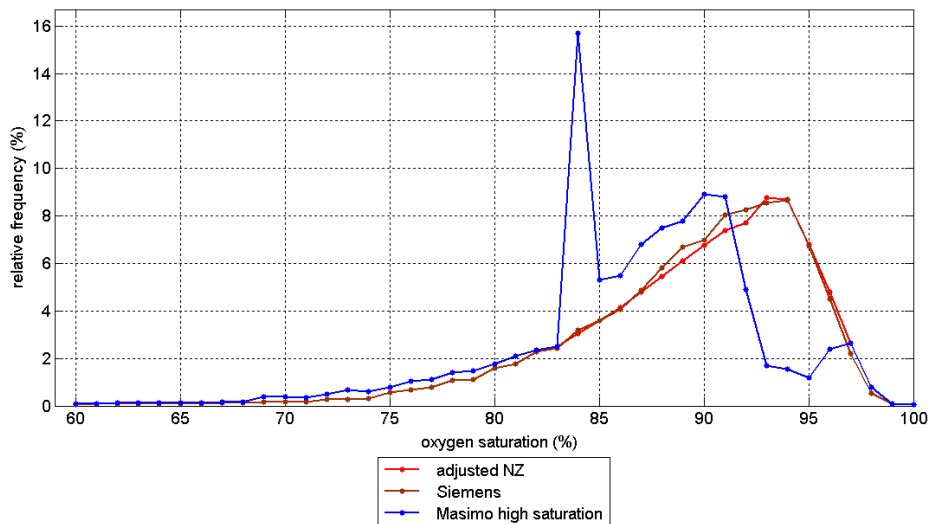


Figure 3.9: Comparison of unadjusted and adjusted readings from a Masimo high saturation oximeter and a Siemens oximeter

3.5 Summary statistics for the OS data

After the preceding pre-processing stages, the adjusted OS readings for each baby were checked to ensure that there is adequate data for obtaining OS summary statistics.

3.5.1 Checking for incomplete data

Readings of OS of each baby were recorded once every 10 seconds by a Masimo pulse oximeter. Therefore, a complete recording for a week for each baby should contain 60,480 data points. However, in reality some data points may be missing; this is indeed the case for the NZ-ROP data. Figure 3.10 illustrates that about 65-80% of the babies have almost complete weekly data (99%), whilst only 5-10% of babies have about 50% of weekly data. These results suggest that the occurrence of missing data is low and hence should not affect the computation of OS summary statistics.

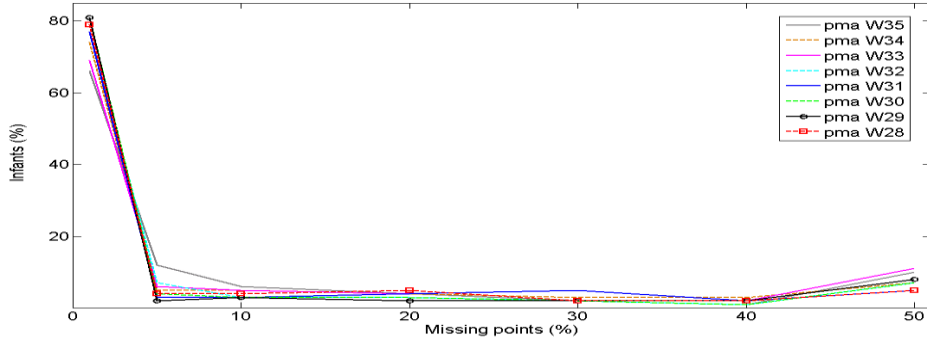


Figure 3.10: The relative frequency of infants with percentage of missing points in NZ-ROP data

3.5.2 Data summary

The last step in the pre-processing procedure was to define statistical summaries for the OS readings. Four summary statistics were selected to measure the variability and level of OS: mean (meanOS), standard deviation (sdOS), coefficient of variation (cvOS), and desaturation (dstOS) of OS. The meanOS served as a measure of the OS level, sdOS and dstOS were measures of the OS variability, and cvOS measured the OS variability relative to its mean. These statistics are defined as follows:

$$\text{meanOS} = \sum_{k=50}^{100} k \tilde{r}_k, \quad (3.12)$$

$$\text{sdOS} = \sqrt{\sum_{k=50}^{100} k^2 \tilde{r}_k - \text{meanOS}^2}, \quad (3.13)$$

$$\text{cvOS} = \text{sdOS}/\text{meanOS}, \quad (3.14)$$

and

$$\text{dstOS} = \sum_{k=50}^{80} r_k, \quad (3.15)$$

where \tilde{r}_k is assumed to be r_k for $k < 84\%$ or $k > 96\%$.

3.6 Summary

The OS data were obtained from a multi-centre randomised control trial; they were recorded using Masimo oximeters. However, these data cannot be used directly for the NZ-ROP project due to some features that resulted from the design of the oximeters. Therefore, the data were pre-processed in several stages, which included data adjustments and data summary. The oximeters had trial-imposed offsets and machine artefact, both of which affected the readings displayed on the oximeters; thus, resulting in higher or lower saturations readings between 85%-95%. To adjust for these two features, algorithms based on linear and quadratic interpolations were developed. Subsequently, these algorithms were validated using data sets of an extremely preterm baby from the United Kingdom, which consisted of offsets and non-offset OS data. The result was satisfactory; the adjusted data were very close to the non-offset data; hence, suggesting that the features had been carefully adjusted to reflect the true OS readings.

The final pre-processing stage was the computation of the OS summary statistics. However, to ensure adequacy of data for the computation, the total record of available OS readings in each baby was checked. A close inspection on the data for the period of 28 to 35 weeks PMA revealed that at least two-thirds of the babies had nearly 99% of weekly data. Nevertheless, there was a small percentage (5-10%) of babies, of whom half of their weekly data were missing. The low percentage of missing weekly data for each baby should not affect the computation of the OS statistics. Four OS statistics were then defined to measure the variability (cvOS, sdOS, and dstOS) and level (meanOS) of OS; these statistics were part of the predictors for the regression analysis in Chapter 5.

CHAPTER 4

METHODOLOGY FOR NZ-ROP

4.1 Introduction

The first objective of the NZ-ROP project is to establish whether four OS summary statistics are associated with the risk of extremely preterm babies developing severe ROP. These measures, as defined in Chapter 3 (Section 3.5.2), however, need to be computed upon specification of a time interval. In this chapter, the methods used to analyse the NZ-ROP are described, which include the techniques for identifying relevant time interval to summarise the OS data, strategy for developing predictive models, statistical modelling using logistic regression and validation procedure for the models. The chapter is concluded with a detailed description of methods for examining the degree of relationship of each predictor with severe ROP.

4.2 Time interval to summarise the OS statistics

The OS data were electronically sampled every 10 seconds. There were two issues with these data that must be addressed before they can be used for the calculation of the OS measures: specifying sufficient amount of data and identifying time period that could detect the fluctuations of neonatal oxygenation. Previous studies were examined to investigate the effects of oxygen fluctuations on preterm babies. It was noted that the data were aggregated: five and 10 days of receiving oxygen therapy (York, et al., 2004); weekly for three (Cuninngnam at al., 1995); four (Flynn et al., 1992); and eight (De Fiore at al., 2010; Saito et al., 1993) consecutive weeks of postnatal age. Aggregation of one week data is in accord with the guidelines for retinal screening of premature babies (Fierson et al., 2013) with the recommendation that babies need to be examined from as much as once in three weeks down to within one week depending on the categories of risk. This information lends support that a week of data was sufficient for quantifying the OS data. In Section 3.5.1, it has been demonstrated that the total recording of OS measurements for each baby was adequate for weekly data summaries.

For the second issue that is to determine the appropriate time period for the calculations of the OS measures, a pediatrician was consulted and the factor of time throughout the course of ROP was considered. Over the range of birth weight up to 1250 gram, ROP has been shown to develop from about 32 to approximately 41.5 weeks PMA (Good et al., 2005), and progressed to severe ROP at a median of 35 weeks PMA in preterm babies born less than 25 weeks GA (Coats et al., 2000). The development of ROP is characterised by growth of retinal blood vessels (see Section 2.2) and recent research data have suggested that this disease progresses in two distinct phases: the cessation of vessels growth after premature birth and uncontrolled growth of vessels that may lead to retinal detachments and blindness (Chen and Smith, 2007; Hellstrom et al., 2003). The first phase advances to the second phase at about 30–32 weeks PMA. As the timing to intervene medically in the development of retinopathy is critical, the pediatrician then recommended that the association between the level and variability of OS, and ROP be identified in the earlier weeks of PMA so as to curb and prevent more severe ROP cases in very young babies.

Consequently, potential time periods were investigated, which included visualising the trend of the group means of the OS measures from 28 until 35 weeks PMA. It was necessary to start at 28 weeks PMA since the oldest baby recruited for the BOOST-NZ was born at 27 weeks GA. Further, the following was looked into: (i) any significant differences between the severe and no severe ROP groups using the two-sample t-test, and (ii) the biggest drop or highest peak of the group means in a particular week. After evaluating the results across the eight successive weeks, 29 weeks PMA was used. The justification for choosing this particular PMA is discussed in Chapter 5 together with the results of the logistic regression when applied to the chosen PMA.

4.3 Logistic regression modelling

4.3.1 *The model*

The logistic regression model is one case of the broader class of generalised linear models (GLMs) introduced by Nelder and Wedderburn (1972). A GLM is specified by two components:

- a response variable y that follows one of the exponential family distributions, and

- a link function, g , which links the conditional mean of the response to a linear combination of predictors, that is $g(\pi) = X\beta$, where X is a row vector of predictors (with 1 as the first element of X when an intercept is included), β is a column vector of parameters, and $\pi = E(y|X)$.

For a binary logistic regression model (Cox, 1958), the response variable is $y = 1$ or 0 , which corresponds to the presence or absence of an event. The link function in this case is $g(\pi) = \text{logit}(\pi) = \log(\pi/1 - \pi)$; hence, the model can be written as

$$\pi = \frac{\exp(X\beta)}{1 + \exp(X\beta)}. \quad (4.1)$$

Parameter estimation

Logistic regression models were fitted in MATLAB (The MathWorks, Inc 1984-2010) using the ‘glmfit’ function. The function employs the iterative reweighted least squares (IRWLS) and produces the estimates for the regression coefficient $\hat{\beta}_j$ of j -th predictors, standard error $SE(\hat{\beta}_j)$ and p -value.

4.3.2 The response and predictors of the logistic regression

The records of all eye examination visits of 257 babies were examined (i) to confirm the end point of the eye examination, and (ii) to determine the binary status of ROP stage. For each baby, an ophthalmologist verified the most advanced ROP stage and classified the status as either presence (‘1’) or absence (‘0’) of severe ROP. These binary judgements were the response variable ($y = 1$ or 0) in the logistic regression models.

There were 11 predictors considered in the analyses,

- the four OS measures: *cvOS*, *sdOS*, *meanOS*, and *dstOS*,
- weight at birth (*weight*), which is recorded in grams,
- age at birth (*GA*), which is measured in days; *GA* was not categorised by week (23 to 27 weeks *GA*) because of the small number of severe ROP babies in three of the *GA* categories (23, 26, and 27 weeks *GA*),

- *gender* with ‘1’ as male and ‘0’ as female,
- number of days receiving ventilation during hospitalisation (*daysVentilate*),
- number of days receiving supplemental oxygen during hospitalisation (*daysO2*),
- dependency on supplemental oxygen at 36 weeks PMA (*O2W36*) with ‘1’ as presence of dependency and ‘0’ as absence of dependency,
- weight-for-GA (*wGA*), which was formed from the information of weight combined with GA. It has been suggested that monitoring of babies classified by birth weight alone may introduce bias in identifying infants who are at risk of a poor outcome as there is no distinction of the size for age and maturity (Roberts and Lancaster, 1999). *wGA* was categorised according to the Australian national birth-weight-for-GA percentile charts (Roberts and Lancaster, 1999). To create the categories for the percentiles, the sparseness of the data in the large-for-GA (>75th percentile) and small-for-GA (<25th percentile) groupings was considered. Therefore, the data were divided into three percentile groups and were coded as follows: 1, >75th (reference group); 2, 25th to 75th; and 3, <25th.

The set of the 11 predictors consists of eight continuous predictors, two binary predictors, and one categorical predictor. To gain a better understanding of each predictor, their profile based on kernel densities and box plots were constructed separately for the severe ROP and no severe ROP groups. These series of plots are useful in discovering the structure of the data, in that potential distinct patterns between the two groups and outlying observations are highlighted.

To complement the graphing of the data, the independent two-sample *t*-test, was applied to test for a difference in the group means between the severe and no severe ROP groups for the continuous predictors. Likewise, for the binary and categorical predictors, the chi-squared test was used to examine if there exists a difference in ROP status due to these predictors: *gender*, *O2W36*, or *wGA*.

4.3.3 Collinearity

Regression methods are sensitive to correlated predictors. When one or more predictors are linearly related with other predictors in the data set, this could potentially produce unreliable parameter estimates with inflated standard errors; thus, reducing the power of testing the affected coefficients (Belsley et al., 1980). To detect collinearity, one could examine the pairwise correlation coefficient, together with the 95% confidence interval (CI). Alternatively, we could identify collinearity by computing the variance inflation factor (VIF) (Marquardt, 1970) computed from $1/(1-R_j^2)$, R_j^2 is the square of the multiple correlation of the j -th predictor regressed on the other predictors in the model.

To handle collinearity, the predictors were separated into groups as follows:

- predictors with pairwise correlations of about 0.8 and above (ignoring the sign), and with extremely high VIF were grouped together, and
- predictors not correlated with non-group members were grouped individually.

In any given model, the combinations could consist of at most one predictor from each group of the strongly correlated predictors.

4.3.4 Predictor selection

Predictor selection is a difficult but an important task, which requires one to consider carefully depending on the inferential goals and the nature of the data set. Of the various procedures for entering predictors into the model, the commonly used procedure is stepwise selection (Miller, 2002). For studies with a large number of predictors (say, more than 50), the stepwise selection provides a practical way of dropping predictors that do not contribute to the model. Its usage, however, has not been without criticism. It can only produce a single model with no guarantee of being a good one as it is capable of producing poor parameter estimates (Chatfield, 1995). The resulting model can be misleading as some authentic predictors may have been omitted, or a number of predictors with deleterious effects may have gained entry (Derksen and Keselman, 1992).

The alternative procedure that was used in this project was all-subsets regression, which produced single- and multi-predictor models for all possible combinations of predictors (Miller, 1990). There was different combination of predictors in each model; thus, the predictions for babies at risk of developing severe ROP differ as well. Additionally, an approach based on a weighted average of the models was employed to circumvent the need to base inference on only one or several models. Single- and multi-predictor models, and model averaging approaches were then used to analyse potential association between the predictors and severe ROP. The following sections provide the outlines for the three analyses.

4.3.5 *Single- and multi-predictor analyses*

We examined all models that were obtained from the all-subsets regression. The results for single-predictor models were discussed in terms of the odds ratio (OR), p -value, as well as fitted probabilities of severe ROP and no severe ROP for selected predictors. The unadjusted OR is estimated as

$$\text{OR} = \exp(\Delta X \hat{\beta}) \quad (4.2)$$

and the corresponding 95% confidence limits for the OR are $\exp\left\{\Delta X \left[\hat{\beta} \pm 1.96SE(\hat{\beta})\right]\right\}$, where ΔX = interquartile range of predictor for continuous predictors, and $\Delta X = 1$ for categorical predictors. Note that the ORs for continuous variables depict the odds of severe ROP for patients at the 75th percentile of the distribution of the variable versus patients at the 25th percentile.

For the analysis of multi-predictor models, the models were ranked according to their respective AUC (refer to Section 4.5.2 and Section 4.6.3), and the number of occurrences of each predictor in the list of the first ten top models was recorded. The fitted probabilities of severe ROP and no severe ROP also were plotted to assess the prediction performance of several top models.

4.3.6 Model averaging analysis

The degree of relationship between the predictors and severe ROP is likely to vary in the single- and multi-predictor models. Thus, a model averaging approach was undertaken to determine the degree of relationship between each predictor with severe ROP. This approach takes into account all models, hence all predictors, and weights the predictors accordingly. The analysis was based on three model-averaged measures: the p -values and ORs (with 95% CI) for each predictor that were computed by averaging over all models that contain the specific predictor, and fitted probabilities for severe ROP and no severe ROP groups that were computed by averaging over all models.

Let J be the set of indices of models that contain predictor j . The model-averaged p -value and model-averaged OR would have the form

$$\sum_{m \in J} w_{m,j} f(x_{m,j}), \quad (4.3)$$

where f is a function of x_{mj} for predictor j in model m , and

$$w_{m,j} = \frac{\text{AUC}_m}{\sum_{u \in J} \text{AUC}_u} \quad (4.4)$$

i.e. $w_{m,j}$ are the normalised AUCs over all u models that contain predictor j . By definition, $0 < w_{m,j} < 1$ and $\sum_{m \in J} w_{m,j} = 1$ for fixed j .

Following from equation (4.3), each measure that is constructed in this form will be a weighted average of some function f of predictor j in model m . For model averaged p -value, $f(x_{m,j}) = P_{m,j}$, where $P_{m,j}$ denotes the p -value of predictor j in model m and for model-averaged OR, $f(x_{m,j}) = OR_{m,j}$, where $OR_{m,j}$ denotes the OR of predictor j in model m , as obtained from equation (4.2). The latter is accompanied

with a model-averaged 95% CI, $\left(\sum_{m \in J} w_{m,j} L_{m,j}, \sum_{m \in J} w_{m,j} U_{m,j} \right)$, where L_{mj} and U_{mj} are, respectively, the lower and upper confidence limits for OR_{mj} .

The model-averaged fitted probabilities are computed as $\sum_{m \in M} w_m \hat{\boldsymbol{\pi}}_m$, where $\hat{\boldsymbol{\pi}}_m$ is a vector of fitted probabilities of severe ROP for the observations in model m , and $w_m = \frac{AUC_m}{\sum_{m \in M} AUC}$ are the normalised AUCs over all M models, with $0 < w_m < 1$ and $\sum_{m \in M} w_m = 1$.

In the model-averaged analysis, the degree of relationship of each predictor with severe ROP was ranked according to its respective model-averaged p -value. Accordingly, the OR of the predictor was evaluated, and the model-averaged fitted probabilities were summarised in the form of boxplots. These results were then compared with those obtained from the single- and multi-predictor models, and the relations between these predictors and severe ROP were summarised.

4.4 Diagnostics

Careful fitting of the statistical models during the model building is important to avoid losing clinically important predictors, especially after adjusting for other predictors. Before these models can be concluded as being adequate for the data, they need to be checked for an appropriate functional form for each (continuous) variable, and for any outlying observation in the data.

Several diagnostic tools were used to accomplish the two aspects. For the first part, a direct way of checking was to plot the smoothed partial residuals from the logistic models and the Generalised Additive Models. For detecting outlying points, approximated jackknife residuals (Williams, 1987), Cook's statistic (Cook and Weisberg, 1982), hat matrix (Belsley et al., 1980) and a diagnostic measure from the robust logistic regression (Cantoni and Ronchetti, 2001) were visually and jointly assessed.

It is important to identify outlying points as their presence not only affects the parameter estimates but also other statistics computed from the model. As a result, this causes the regression relationship to appear to look stronger than it is, or non-existent leading one to abandon the model. Once the outlying points were detected, the sources of the errors were identified and the degree of their discrepancy on the data was assessed. This included analysing the data with and without the suspect observations.

4.4.1 Linearity assumption for continuous predictors

At the predictor selection stage, it is quite common to assume that each predictor is linearly related to the log odds of the response. This assumption, however, may not always hold after the initial modeling of the predictors. The subject matter knowledge was taken into account to assess the appropriate scale of the predictors and to guide the transformation of the predictors. Increased variability in oxygenations (York et al., 2004), desaturation episodes (De Fiore et al., 2010), and duration of oxygen therapy (Gallo et al., 1993) have been reported to be associated with the development of ROP. On the other hand, GA and weight have been found to have the opposite effect on the risk of severe ROP (Darlow et al., 2005).

Logistic regression partial residuals

The first approach to assess the parametric scale for each continuous predictor was to use the partial residuals from the logistic model. A plot of the partial residuals against a predictor is helpful, in that the linearity of a predictor can be examined on the basis of other predictors. Along with the plot, a lowess smoother (Cleveland, 1993) was applied on the residuals to create a clearer visual of the functional shape. The partial residual z_{ij} from a logistic regression model for the i -th observation of the j -th predictor, x_{ij} , as given by Landwehr, et al. (1984) is

$$z_{ij} = \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)} + x_{ij}\hat{\beta}_j, \quad (4.5)$$

where the first term corresponds to the variance-adjusted residual, $\hat{\beta}_j$ is the parameter estimate for the j -th predictor, y_i is the i -th response, and $\hat{\pi}_i$ is the predicted probability of severe ROP for the i -th observation.

Generalised additive logistic model

As an alternative, the structure of a continuous predictor can be evaluated by fitting a generalised additive logistic model (GAM), which is defined as

$$\text{logit}(\pi) = \beta_0 + \sum_{j=1}^p f_j(x_{\cdot j}), \quad (4.6)$$

where β_0 is the intercept parameter, $f_j(\cdot)$ is a smooth nonparametric function of the j -th predictor, and p is the number of predictors.

Additive models were enhanced by Hastie and Tibshirani (1990), the former author who developed the *gam* library in R. The library offers two types of smoother, cubic spline and lowess. The degree of smoothing depends on the smoothing parameter; higher penalties are imposed for more complex nonparametric fitted curves. *gam* uses the back-fitting algorithm, which is well-discussed in Hastie and Tibshirani (1990). The fitting process for a GAM involves two levels of iteration. The first iteration entails the GLM fitting using IRWLS procedure and the second iteration implements the back-fitting algorithm to estimate the f_j s for the additive model (Faraway, 2006: Chapters 6 and 12).

We experimented with both cubic spline and loess smoothers with several values of smoothing parameter on the eight continuous predictors. The differences in the results were very minimal between the two smoothers; for cubic spline, there was not much improvement when applying a smoothing parameter greater than three. Consequently, a cubic spline with a smoothing parameter equals to three was employed.

Piecewise linear regression model

The fit of a GAM provides a smoothed estimate of the predictor effect which can then be plotted to check for nonlinearity. In the case when the linearity assumption is violated, it could indicate an appropriate transformation required for a predictor. For strong violations of the assumption, for instance a U-shaped relationship, an extra (or multiple) term can be flexibly added to the predictor. Polynomial and piecewise linear regressions (also known as linear splines) are among the common choices for the expansion. The latter (Smith, 1979) generally consists of at least two segments of linear fit with a continuous joint between the segments; this joint is also known as the knot. With two line segments and a knot, a single-predictor piecewise model for a logistic regression can be written as

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_j + \beta_2 (x_j - k) I(x_j > k), \quad (4.7)$$

where k is the position of the knot, and $I(\cdot)$ is the binary indicator variable of the j -th predictor, that is, $I(A) = 1$ if A is true and 0 otherwise.

4.4.2 Logistic regression diagnostics

The discussions of outliers in regression normally involve detecting and handling several types of outliers; these include identification of outliers from the influential observations and large leverage values. Leverages are not necessarily influential points but have the potential to affect the fit. A leverage h_i of the i -th observation can be extracted from the diagonal elements of the hat matrix \mathbf{H} (Pregibon, 1981),

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}, \quad (4.8)$$

where \mathbf{W} is a $n \times n$ matrix of weights from the IRWLS procedure for fitting GLMs, and \mathbf{X} is the $n \times (p+1)$ design matrix of predictors.

For detection of the influential observations, we used the Cook's distance (Pregibon, 1981), which measures the overall influence of each observation on the parameter estimates. The influence statistic is

$$C_i = \frac{(\hat{\beta}_{-i} - \hat{\beta})^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) (\hat{\beta}_{-i} - \hat{\beta})}{p+1}, \quad (4.9)$$

where $\hat{\beta}_{-i}$ denotes the estimates of the parameters with the deletion of the i -th observation from the data.

The third diagnostic measure to check for the unusual points is a half-normal plot of approximated jackknife residuals. In the plot, the absolute residuals for n observations were sorted and compared with the quantiles of the half-normal distribution, $\Phi^{-1}\left(\frac{n+i}{2n+1}\right)$, for $i = 1, \dots, n$. With all points in one tail, we did not anticipate a straight line trend as the residuals were not expected to be normally distributed. Instead, we looked for those points that deviate from the trend. The residuals developed by Williams (1987), with the deletion of i -th observation, represent the approximation to the jackknife residuals:

$$z_{L,i} = (1-h_i)z_{SD,i}^2 + h_i z_{SP,i}^2, \quad (4.10)$$

where $z_{SP,i} = (y_i - \hat{\pi}_i) / \sqrt{\hat{\pi}_i(1-\hat{\pi}_i)(1-h_i)}$ and $z_{SD,i} = \text{sign}(y_i - \hat{\pi}_i) \sqrt{d_i / (1-h_i)}$. The numerator for the former is known as the Pearson residual, and for the latter the deviance residual. To define d_i , let $f(y | \pi) = \pi^y (1-\pi)^{1-y}$ denote the Bernoulli likelihood function for π evaluated at y , then $d_i = 2 \ln[f(y_i | y_i) / f(y_i | \hat{\pi}_i)]$.

There is, however, no exact threshold value as to what constitutes a large leverage and Cook's statistic values. Several rule-of-thumbs have been suggested from the literature; for example, a leverage is suspected of being an outlier when $h_i > 2(p+1)/n$ (Hoaglin and Welsch, 1978). However, Belsley et al. (1980) warned that in general this cut-off detects too many high-leverage values. The same issue occurs for the Cook's statistic. To be stringent with the cut-off, we fixed the threshold to $2(p+2)/n$ for leverage and $8/(n-2p-2)$ for C_i (Belio, 2005).

Detection for outliers in multiple regressions based on conventional procedures can frequently be a difficult task, especially if multiple extreme values exist in the data set or the number of predictors is large. Observations with high leverage may sometimes fail to stand out in the diagnostic plots, and so the regression analysis will produce poor results. In the view of this, the robust regression was considered and used in tandem with the conventional procedures.

4.4.3 Robust logistic regression diagnostics

The application of the robust methods is not limited to estimation and hypotheses testing in regression. Because the estimators are developed so that they are resistant to the outliers, the procedure can also be used to screen for outlying points. There are three main approaches toward the robustness problem in regression: the Huber-type M -estimation (Huber, 1981) for detecting outliers in the y -direction; the Mallows-type bounded influence estimation (Mallows, 1973) for a moderate percentage of outliers in the x -direction, or high leverage points; and the high breakdown point estimation for a high percentage (at most 50%) of outliers.

Several methods had been developed for the class of generalised linear models, for instance the Bianco and Yohai estimator (Bianco and Yohai, 1996; Croux and Haesbroeck, 2003), and the Cantoni and Ronchetti estimator (Cantoni and Ronchetti, 2001). The latter was used in this analysis. It is an M -estimator of Mallows-type, of which the effect of outlying values in the y -direction is controlled by placing a bound on the influence function, and in the x -direction by setting weights on the observations in the predictors. A suggested bounded function to ensure robustness

in the y -space is $\frac{\psi_t(z_i)}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$ (see equation 4.11), whilst a choice of weights to down-

weight high leverage values is the diagonal elements of the hat matrix.

The Cantoni and Ronchetti estimator is defined for a class of robust estimators in the general setup of generalised estimating equations. The j -th estimating equation can be written as

$$\sum_{i=1}^n \left[\psi_t \left(\frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \right) \sqrt{\frac{(1-h_i^*)}{\hat{\pi}_i(1-\hat{\pi}_i)}} \frac{\partial \hat{\pi}_i}{\partial \beta_j} - \gamma(\beta) \right] = 0, \quad (4.11)$$

where

$$\psi_t(z) = \begin{cases} z, & |z| \leq t, \\ \text{sign}(z)t, & |z| > t, \end{cases} \quad (4.12)$$

is the Huber function (Huber, 1981) with a tuning constant t , h_i^* is the i -th diagonal

entry of $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, and $\gamma(\beta) = \frac{1}{n} \sum_{i=1}^n E \left[\psi_t \left(\frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \right) \middle| X_i \right] \sqrt{\frac{(1-h_i^*)}{\hat{\pi}_i(1-\hat{\pi}_i)}} \frac{\partial \hat{\pi}_i}{\partial \beta_j}$. The

conditional expectation can be written explicitly as

$$\hat{\pi}_i \psi_t \left(\frac{1 - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \right) + (1 - \hat{\pi}_i) \psi_t \left(\frac{-\hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \right) = \hat{\pi}_i \psi_t \left(\sqrt{\frac{1-\hat{\pi}_i}{\hat{\pi}_i}} \right) + (1 - \hat{\pi}_i) \psi_t \left(-\sqrt{\frac{\hat{\pi}_i}{1-\hat{\pi}_i}} \right).$$

The method described by equation (4.11) has been implemented in the *robustbase* library in R via the ‘glmrob’ function. It yields the estimates of the parameters along with their standard error and p -value. The function also returns a set of weights with respect to their residual for each observation; these are denoted as the

Mallows robust weights, $\psi_t \left(\frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \right) / \left(\frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} \right)$.

4.4.4 Outliers detection

Identification of multiple outlying points involves several stages. First, the sign of the parameter estimates in all models was examined whether it was similar to that in the single-predictor analysis and generally indicated in the literature. Models that contained parameters estimates with the opposite sign were listed. Next, the conventional and robust diagnostic measures for these models were computed and plotted concurrently to detect all candidate outliers. The occurrences of suspected observations that exceeded the suggested thresholds of the diagnostic measures were recorded. The listed models were then re-fitted with the exclusion of the suspected observations, and their impact on the sign and standard error of the parameter estimates was assessed. Identification of outliers was finalised using the results from

these stages. To determine on how best to handle the outliers, we examined the effect of the outliers on the results obtained from the logistic regression and robust logistic regression performed on the full data set, and logistic regression with the outliers omitted.

4.5 Assessment of models

4.5.1 *Fitting of Model*

In the previous section (Section 4.4.1), the preliminary assessment of the linearity in structure of the continuous predictors were described. The partial residuals from the logistic regression models and GAM were plotted and the patterns were noted. Next, a test of adequacy of a linear relationship via a likelihood ratio (LR) test was performed to verify whether a nonlinear term in the form of cubic spline, piecewise or polynomial was worth considering for each predictor.

The LR test was used to evaluate the change in the deviance between two models: a model that was fitted with a linear term and another model that was fitted with a nonlinear term. A significant change in the deviance signifies that the model with the nonlinear term is significantly different from the linear model, and therefore, the nonlinear term is favourable for the predictor. If there are indications from the results that a nonlinear term could be needed for the predictors, simpler nonlinear parametric models such as linear piecewise or quadratic models are suggested as the alternatives to the cubic spline as they are more parsimonious and easier to interpret. Separate comparisons were then made between a linear model and a piecewise linear model, or between a linear model and a quadratic model. Using the results from the linear and nonlinear model fittings, the proposed functional form for each predictor was finalised.

Throughout the comparisons between models to verify suitable functional forms of the predictors, the LR test was used in conjunction with the Akaike Information Criterion (AIC). When two models are compared, a model with a lower AIC value is preferred. However, there is often no one single best model when the difference in AICs between two models is assessed. Therefore, consideration must then be given to other models whose AIC values are close to the minimal AIC (Hilbe, 2009). The AIC is defined as

$$\text{AIC} = -2LL + 2q, \quad (4.13)$$

where LL is the model log-likelihood and q is the number of unknown parameters in the model.

4.5.2 *Measuring the performance of model*

The error rate, ROC curve and the associated AUC are popular measures to estimate the performance of binary classifiers such as the logistic regression model in the medical research. However, the AUC has been shown to exhibit several desirable properties over the error rate including decision threshold independent (Bradley, 1997). The AUC summarises an ROC curve; its value reflects the overall performance of the model in discriminating between babies who will or will not develop severe ROP. A large value of an AUC indicates better discriminative ability of a model. An assignment to a class by random chance has an AUC of 0.5 whilst a perfect classification corresponds to a maximum AUC value of 1.0, in which all subjects are classified correctly.

A common approach to compute the AUC is to construct the ROC curve from the sensitivity (the proportion of correctly identified of diseased subjects) and specificity (the proportion of correctly identified of non-diseased subjects), as calculated from different threshold values, and then to estimate the area under it by numerical integration. An easier alternative that does not require explicitly constructing the ROC curve is to use the Wilcoxon statistic (Hanley and McNeil, 1982) to estimate the AUC (θ), that is

$$\hat{\theta} = \frac{1}{v(n-v)} \sum_{i=1}^v \sum_{l=1}^{n-v} \alpha(\hat{\pi}_{D,i}, \hat{\pi}_{ND,l}), \quad (4.14)$$

where $\hat{\pi}_{D,i}, i = 1, \dots, v$ and $\hat{\pi}_{ND,l}, l = 1, \dots, (n-v)$ are the predicted probabilities for the diseased and non-diseased subjects, respectively, and

$$\alpha(\hat{\pi}_{D,i}, \hat{\pi}_{ND,l}) = \begin{cases} 1 & \text{if } \hat{\pi}_{D,i} > \hat{\pi}_{ND,l} \\ \frac{1}{2} & \text{if } \hat{\pi}_{D,i} = \hat{\pi}_{ND,l} \\ 0 & \text{if } \hat{\pi}_{D,i} < \hat{\pi}_{ND,l} \end{cases} \quad (4.15)$$

The estimate is more meaningful when accompanied with a CI. A convenient choice would be the approximate 95% CI given by

$$\hat{\theta} \pm 1.96\sqrt{\text{var}(\hat{\theta})}, \quad (4.16)$$

provided that the sample size is sufficiently large, and θ is not near the boundaries. An estimate of the variance of $\hat{\theta}$ is given analytically by the DeLong estimate (DeLong et al., 1988), which is computed as follows.

For each diseased subject i , compute the quantities

$$\bar{\alpha}_{D,i} = \frac{1}{n-v} \sum_{l=1}^{n-v} \alpha(\hat{\pi}_{D,i}, \hat{\pi}_{ND,l}) \text{ and } V_D = \frac{1}{v-1} \sum_{i=1}^v (\bar{\alpha}_{D,i} - \hat{\theta})^2. \quad (4.17)$$

Similarly, for each non-diseased subject j , compute the quantities

$$\bar{\alpha}_{ND,l} = \frac{1}{v} \sum_{i=1}^v \alpha(\hat{\pi}_{D,i}, \hat{\pi}_{ND,l}) \text{ and } V_{ND} = \frac{1}{n-v-1} \sum_{j=1}^{n-v} (\bar{\alpha}_{ND,l} - \hat{\theta})^2. \quad (4.18)$$

The variance for $\hat{\theta}$ is then estimated as

$$\text{var}(\hat{\theta}) \approx \frac{1}{v} V_D + \frac{1}{n-v} V_{ND} \quad (4.19)$$

The all-subsets logistic models were re-fitted to the NZ-ROP data after diagnostics of model fits were undertaken. The fitted regression models were able to generate a fitted probability of severe ROP for each baby, and this was used to calculate the AUC for each model. The performance of the individual models were

then assessed by analysing the AUCs. Note that both the AUC and the fitted probabilities were computed via a cross-validation technique (Section 4.6.3).

4.6 Validation of model

It is important to examine if a developed model will yield a good predictive performance in other data sets. A good calibration of the model requires a validation on at least one independent data set. Nevertheless, if this is not feasible, one could withhold a percentage of the data from the model fitting process. An ideal procedure when the data set is large would be to divide the data into three subsamples: the training set for fitting the model, the test set for generating the predictive probability, and the validation set for validating the probability. However, when the amount of data is limited, sample-reuse validation schemes such as holdout validation and cross-validation could be employed. These techniques use sampling without replacement, in which data are partitioned into two subsets: one serves as the training set and the other as the test set.

4.6.1 Basic idea of cross-validation

The basic form of cross-validation is K -fold cross-validation. First, the data are partitioned into K equally (or almost equally) sized folds. One fold is used for testing and the remainder for training. To ensure equal chance of validation for each data point, the sets are rotated in K successive iterations such that within each iteration, a different fold of the data is held-out for testing and different $K - 1$ folds are used for training. Figure 4.1 demonstrates an example of K -fold cross-validation procedure with $K = 3$. The darker sections depict the subset of data that are used for training while the lighter sections represent the subset for testing. Occasionally, the subsets are stratified before the validation starts, and the cross-validation is repeated for several rounds, hence, yielding a *repeat-stratified K -fold cross-validation* scheme.

To compute an overall estimate of a performance metric such as AUC, two approaches that utilise the results of the folds during the cross-validation can be used. One is the pooling strategy, which involves collecting the model predictions from all folds and calculating the AUC on these combined outputs. Another is the averaging strategy, in which an AUC is computed separately for each fold and subsequently, these AUCs are averaged to produce an overall AUC.

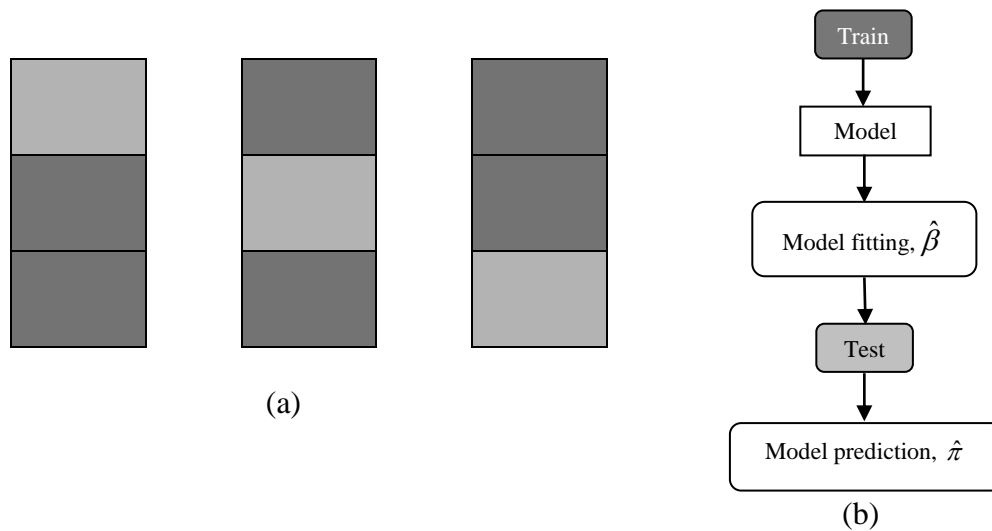


Figure 4.1: Procedure of K -fold cross-validation for $K = 3$, which displays (a) partition of data and rotation of folds for each of the three successive iterations, and (b) training and testing of data within each iteration.

4.6.2 Assessment of several validation schemes

Several studies have been carried out to review (Refaeilzadeh et al., 2009) and assess various cross-validation and bootstrap procedures (Kim, 2009; Parker et al., 2007; Kohavi, 1995). A brief summary of the comparison is presented in Table 4.1.

The studies compared a variety of simulation settings and real-world data with differing characteristics to evaluate the performance of most of the procedures (Table 4.1). Many interesting results were demonstrated, including the 0.632+ bootstrap (a variant of the bootstrap) which displays superior performance for small sample case. However, it suffers from a bias problem for small and even more for large data, hence performing poorly as compared to a repeated cross-validation with ten folds. A repeat hold-out scheme faces the same fate as the 0.0632+ bootstrap, in that it also has a bias problem, particularly so for small data. Overall, both schemes show less stable performance for small and large sample cases.

Table 4.1: Evaluation of different validation methods

Validation method	Advantage	Limitation
Resubstitution validation	All available data are used	Over-fitting
Holdout validation	No overlap between training and test set	Inefficient use of data; Results are highly affected by the choice of partition size for training and test sets
K -fold cross-validation	Available data are fully utilised	High dependency on the choice of K ; Higher bias for smaller K and data
Leave-one-out cross-validation	Nearly unbiased performance estimation	Very large variance and computationally expensive for large data
Repeated K -fold cross-validation	Reduced variance	Computationally expensive
Bootstrap	Low variance; best for small data	Very large bias

The repeated-stratified 10-fold cross-validation was unanimously recommended for general use due to its most stable performance as compared to its contenders. Its attractive features such as stratification and repeated runs of the validation have been identified as being extremely useful for reducing both the bias and variability. The choice of $K = 10$ is often suggested as 90% of the data are used for model fitting, hence, allowing generalisability of the results for the case when all data are used. Additionally, the cross-validation based on ten folds has less biased estimation of model performance; the bias is more apparent with smaller K (Kohavi, 1995).

4.6.3 Validation for NZ-ROP data

For the NZ-ROP data, we chose to implement a repeat-stratified 10-fold cross-validation (RS10CV) based on the following grounds.

1. It was not possible to get an external data set from the same population. The BOOST-NZ trial is part of an International Network, which consists of several

countries such as Australia, United Kingdom, Canada, Europe, and USA. The OS data from these trials were recorded from the Masimo oximeters, which employed different algorithms from that of the New Zealand study oximeters.

2. It was not realistic to perform hold-out validation because of the small sample size ($v = 41$) in the severe ROP group.

3. RS10CV appears to be the best method among its competitors. Results from the simulation studies (Kim, 2009; Parker et al., 2007; Kohavi, 1995) have shown that across the variants of internal validation (resubstitution, hold-out, leave-one-out, K -fold, repeated- K -fold, and 0.632+ bootstrap), RS10CV has low estimation bias and excellent trade-off between bias and variance. It was, therefore recommended for various usages such as selection of models, estimation of AUC and error rate, and assessment of multiple classifiers.

Cross-validation procedure for AUC and predicted probability

To examine the performance of a model, and to verify that it could provide good prediction for new data sets, the AUC and its 95% CI were computed from $N = 100,000$ repeats of RS10CV as follows:

1. Stratify the data according to their diseased status, yielding two groups of sizes v and $(n - v)$ for the diseased and non-diseased groups, respectively.
2. Randomly divide each of the groups into ten mutually exclusive folds of equal (or almost equal) size.
3. Combine the diseased and non-diseased groups for each folds.
4. Treat the first fold as the test set and the remaining nine folds as the training set. Fit the regression model to the training set, and obtain the regression estimates.
5. Apply the regression estimates to the test data, and calculate the predicted probability for each observation in the test set.
6. Rotate all folds and repeat step 4 to step 5 for all ten folds.
7. Repeat step 2 to step 6 N times. For each repeat, store the predicted probability for each observation, that is, $\hat{\pi}_{B,i}$, for $i = 1, \dots, n$ and $B = 1, \dots, N$.

8. Estimate an overall predicted probability for each observation by averaging across all N repeats, that is,

$$\tilde{\pi}_i = \frac{1}{N} \sum_{B=1}^N \hat{\pi}_{B,i}, i = 1, \dots, n \quad (4.20)$$

9. Calculate an AUC and its 95% CI using the estimated probabilities from step 8 in equations (4.14) to (4.19).

4.7 Computational aspects

All analyses were conducted in MATLAB (v. 2010a) and R x64 2.12.0. Several packages that were used in R were *gam*, *stats*, *faraway*, *splines*, and *robustbase*.

4.8 Summary

The methodological background for the NZ-ROP analysis was addressed in this chapter. There were two stages of analysis involved: establishing a suitable time interval to summarise four OS statistics, and modelling of these measures and several additional ROP predictors aiming to investigate the relations between 11 predictors and severe ROP, and to identify the key risk predictors for the disease.

The OS data were recorded at every 10 seconds within 24 hours after birth until 36 weeks PMA. To determine a suitable interval to compute the summary statistics for the OS data, an approach using an analysis of eight PMA weeks was used. It took into account the gestational age at birth of the oldest baby recruited for the NZ-ROP study. A time interval of 29 weeks PMA was finally chosen as it was able to detect the most distinct patterns that help to distinguish between babies with or without severe ROP.

Logistic regression models were developed to estimate the probability that a baby is at risk of developing severe ROP. The predictors for the models were weight at birth, GA, gender, weight-for-GA, duration of receiving ventilation, duration of receiving supplemental oxygen, and dependency on supplemental oxygen at 36 weeks PMA. Four new measures of pulse oximeter OS, whose relationship with ROP needed

to be determined, were also added to the list of predictors. The response variable was defined as the presence or absence of severe ROP.

Several predictors were identified to be strongly correlated, hence the predictors were separated into correlated and uncorrelated groups. Consequently, combinations of at most six predictors (from each group) were to be used in any one of the models. The models' adequacy was examined using the conventional and robust diagnostic tools, and generalised additive modelling. All-subsets regression, which produced single- and multi-predictor models for all possible combinations of predictors, was performed to identify predictors that can best describe the relationships with severe ROP. The performance of these models was measured by their respective AUC, which was validated using a repeat-stratified 10-fold cross-validation technique. Additionally, an approach based on the weighted average of the models was employed to determine the degree of relationship of the predictors with severe ROP. The analysis was based on three model-averaged measures: the *p*-values and ORs (with 95% CI) for each predictor, and fitted probabilities for severe ROP and no severe ROP babies, which were computed by averaging over all models.

CHAPTER 5

RESULTS AND DISCUSSION FOR NZ-ROP

5.1 Introduction

Continuous advances in neonatal care have resulted in increased survival of babies at risk of developing ROP. It is a vision-threatening complication that could progress to visual impairment and blindness in preterm babies when treatment is not provided in time. Each year, it is estimated that over 200 extremely preterm babies in New Zealand are at high risk of developing severe ROP (Cust et al., 2003) and endure ongoing morbidity such as visual deficits (Darlow et al., 1997).

Neonatal examination programmes are currently in practice in many countries to enable early diagnose of ROP among high-risk babies. Upon detection, the choices on specific treatments or decisions are based on an assessment of patient risk. Careful attention to pivotal factors that increase the risk of development of ROP might enlighten the role of interventions that could further lower the morbidity rate of this eye disease.

A number of studies have been carried out to understand the underlying risk factors of ROP with the aim to prevent and control the disease. Among those that have been reported to be associated with an increased risk to develop ROP are low GA (Darlow et al., 2004), low birth weight (McColm and Fleck, 2001), use of oxygen therapy (Lala-Gitteau et al., 2007; Kinsey et al., 1956), and gender (Jacobson et al., 2009). These factors are widely studied and more understood than that of oxygen fluctuations. Further, simultaneous analyses involving many key factors to unveil the most prominent ones are limited. Using data of preterm babies admitted to five neonatal NICUs in New Zealand, this study aims to provide a comprehensive analysis of relations between multiple clinical and postnatal factors and severe ROP. This chapter is focused on determining the usefulness of OS summary statistics for discriminating babies at risk of developing ROP, and the degree of relationship of these statistics and other factors with severity of ROP. Section 5.2 addresses the examination on identifying a suitable time interval for OS statistics. This is followed

by the details of the regression predictors according to ROP status in Section 5.3. The analysis from model fittings and diagnostics are shown in Section 5.4, while the results from using some common regression techniques and model averaging approach are presented in Section 5.5. The implications and limitations of the study are discussed in Section 5.6. All conclusions are summarised in the final section, Section 5.7.

5.2 Time interval for computing OS summary statistics

As described in Chapter 3, four OS summary statistics were defined to be used for the regression analysis to investigate the relationship between measures of OS features and severe ROP. The OS data are time series data; hence, the time interval for computing the statistics must be determined. To facilitate the analysis, we looked for patterns that help distinguish babies with or without severe ROP in order to locate a pattern that indicates a good separation between the two groups at a particular time interval.

5.2.1 Determining the time interval

A simple analysis of eight weeks of data from 28 to 35 weeks PMA was performed to determine the best time interval for summarising the OS data. The following steps were performed for each of the severe and no severe ROP groups:

- calculate a weekly value of meanOS from 28 to 35 weeks PMA for each baby;
- calculate a group mean for each individual week; and
- repeat the above steps for sdOS, cvOS, and dstOS.

Figure 5.1 shows the group means of the four statistics with 95% CI for the eight-week time period. Looking closely across all statistics, it can be seen that there is good separation between the two groups of babies, as depicted by the non-overlapping CIs of the mean values, for nearly all PMAs. To verify whether these differences are statistically significant, a two-sample t -test was applied to compare the group means of each statistic at each PMA.

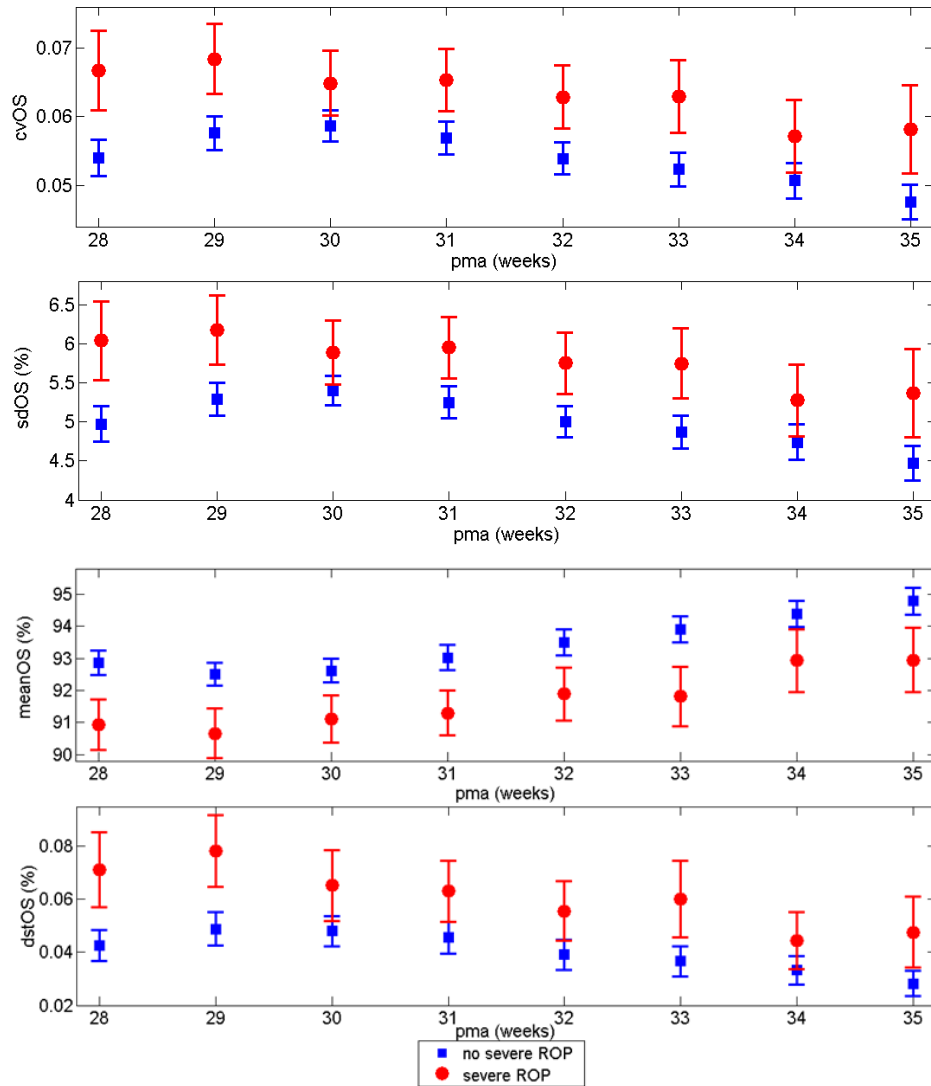


Figure 5.1: Group means (with 95% CIs) of four OS statistics for 28 to 35 weeks PMA: first panel, cvOS; second panel, sdOS; third panel, meanOS; and fourth panel, dstOS

Table 5.1 presents the results from the t -test, which reveals that there are significant differences in the mean values of the statistics between the two groups of babies at nearly all weeks of PMA. This finding suggests that there are significant differences in the oxygenation profiles between the two groups of babies. To determine a particular time period of PMA(s) that would be appropriate for constructing the OS summary statistics, the eight-week OS data were scrutinised more closely.

Table 5.1: Group means of (a) cvOS, (b) sdOS, (c) meanOS, and (d) dstOS with respect to ROP status and PMA

PMA (weeks)	Mean of cvOS (95% CI)				<i>p</i> -value
	No severe ROP		Severe ROP		
28	0.054	(0.051,0.057)	0.067	(0.061,0.073)	3.0E-04
29	0.058	(0.055,0.060)	0.068	(0.063,0.073)	1.2E-03
30	0.059	(0.056,0.061)	0.065	(0.060,0.070)	3.1E-02
31	0.057	(0.055,0.059)	0.065	(0.061,0.070)	5.2E-03
32	0.054	(0.052,0.056)	0.063	(0.058,0.068)	3.0E-03
33	0.052	(0.050,0.055)	0.063	(0.058,0.068)	8.6E-04
34	0.051	(0.048,0.053)	0.057	(0.052,0.062)	5.0E-02
35	0.048	(0.045,0.050)	0.058	(0.052,0.065)	1.4E-03

(a)

PMA (weeks)	Mean of sdOS (95% CI)				<i>p</i> -value
	No severe ROP		Severe ROP		
28	4.969	(4.743,5.196)	6.037	(5.531,6.542)	4.1E-04
29	5.320	(5.078,5.503)	6.178	(5.737,6.619)	1.3E-03
30	5.398	(5.207,5.589)	5.888	(5.480,6.297)	4.6E-02
31	5.250	(5.049,5.452)	5.950	(5.555,6.345)	6.9E-03
32	5.001	(4.803,5.199)	5.752	(5.357,6.146)	3.5E-03
33	4.869	(4.656,5.083)	5.747	(5.296,6.198)	1.3E-03
34	4.740	(4.512,4.968)	5.274	(4.812,5.736)	6.3E-02*
35	4.469	(4.244,4.695)	5.362	(4.797,5.926)	2.1E-03

(b)

PMA (weeks)	Mean of meanOS (95% CI)				<i>p</i> -value
	No severe ROP		Severe ROP		
28	92.87	(92.48,93.26)	90.93	(90.15,91.72)	1.6E-04
29	92.51	(92.14,92.87)	90.66	(89.89,91.43)	1.3E-04
30	92.62	(92.26,92.98)	91.12	(90.38,91.85)	1.3E-03
31	93.03	(92.63,93.42)	91.30	(90.60,92.00)	6.6E-04
32	93.50	(93.10,93.90)	91.89	(91.06,92.72)	2.0E-03
33	93.91	(93.49,94.32)	91.82	(90.89,92.74)	1.0E-04
34	94.39	(93.98,94.81)	92.93	(91.96,93.91)	6.5E-03
35	94.79	(94.38,95.20)	92.95	(91.94,93.96)	5.0E-04

(c)

PMA (weeks)	Mean of dstOS (95% CI)				<i>p</i> -value
	No severe ROP		Severe ROP		
28	0.043	(0.037,0.048)	0.071	(0.057,0.085)	3.4E-04
29	0.049	(0.042,0.055)	0.078	(0.065,0.092)	5.0E-04
30	0.048	(0.042,0.054)	0.065	(0.052,0.079)	2.3E-02
31	0.046	(0.040,0.052)	0.063	(0.052,0.075)	2.3E-02
32	0.039	(0.034,0.045)	0.056	(0.044,0.067)	2.3E-02
33	0.037	(0.031,0.042)	0.060	(0.046,0.074)	1.7E-03
34	0.033	(0.028,0.039)	0.045	(0.034,0.055)	9.9E-02*
35	0.028	(0.023,0.033)	0.048	(0.034,0.061)	2.2E-03

(d)

p-values are from *t*-test for comparing differences in the mean values of OS statistics between severe ROP and no severe ROP babies. **p*-values > 0.05.

Figure 5.1 illustrates that the group means of cvOS, sdOS, and dstOS exhibit similar patterns over time: (i) an increasing trend followed by a decreasing trend with a maximum occurring at 29 weeks PMA, and (ii) higher mean values for severe ROP babies. The statistic meanOS displays patterns that are opposite to these. Taken together, these indicate that OS has highest variability and lowest level at 29 weeks PMA, thus, making it a candidate time interval for summarising the OS data. A

further observation from Table 5.1 is that the largest differences (indicated by $P < 0.0005$) occur at 28 weeks PMA for cvOS, sdOS, and dstOS; and at 28, 29, and 33 weeks PMA for meanOS. These suggest that 28 weeks PMA should also be considered for summarising the OS data.

Subsequently, in addition to 28 and 29 weeks PMA considered separately, a combined data set from both weeks was also assessed. These results are shown in Table 5.2. Comparing them with the earlier results for 28 and 29 weeks PMA in Table 5.1, we see that the p -values for the combined data remain significant for all four statistics, and are consistently smaller than those for 29 weeks PMA. Three pairwise t -tests were then conducted for babies with ROP for the three periods: between 28 and 29 weeks PMA, 28 weeks PMA and the combined data, and 29 weeks PMA and the combined data. Likewise, these tests were also performed for babies without ROP. Table 5.3 presents the p -values from the six tests. The results indicate no significant difference in the means of OS statistics between the pairwise time periods (the smallest p -value, $P > 0.05$). Therefore, any of these three time periods could be used to obtain the OS summary statistics for the regression analysis for investigating the influence of risk factors for severe ROP.

Table 5.2: Group means of cvOS, sdOS, meanOS, and dstOS, with respect to ROP status, based on a combined data set of 28 and 29 weeks PMA

OS statistic	Mean (95% CI)		p -value
	No severe ROP	Severe ROP	
cvOS	0.057 (0.055,0.059)	0.068 (0.063,0.073)	5.3E-04
sdOS	5.242 (5.035,5.448)	6.146 (5.704,6.587)	8.4E-04
meanOS	92.65 (92.28,93.10)	90.80 (90.03,91.57)	1.1E-04
dstOS	0.046 (0.040,0.052)	0.074 (0.061,0.088)	1.9E-04

p -values are from t -test for comparing differences in the mean values of OS statistics between severe ROP and no severe ROP babies; the differences between the two groups are significant.

Table 5.3: *P*-values from *t*-test for comparing differences in means of OS statistics between pairwise time periods in severe ROP and no severe ROP babies

OS statistic	No severe ROP			Severe ROP		
	T1 vs. T2	T1 vs. T3	T2 vs. T3	T1 vs. T2	T1 vs. T3	T2 vs. T3
cvOS	0.0637	0.1245	0.7296	0.6599	0.7478	0.9005
sdOS	0.0531	0.1018	0.7431	0.6801	0.7510	0.9198
meanOS	0.2268	0.4342	0.6589	0.6264	0.8057	0.8073
dstOS	0.2158	0.4897	0.5613	0.4798	0.7331	0.6981

Time periods are indicated as T1: 28 weeks PMA; T2: 29 weeks PMA; and T3: combined 28 and 29 weeks PMA.

Hence, a separate all-subsets regression was conducted involving OS statistics derived from each time period. The results from the three regression analyses were quite similar, in that, the predictor variables that were significantly associated with the ROP status were the same – albeit a slight difference in the regression coefficients. However, the AUC of the models were slightly higher and the *p*-values of the regression coefficients were much smaller (by at most 86%) for 29 weeks PMA. Hence, only the results based on 29 weeks PMA will be presented in this chapter in Section 5.5.

5.2.2 Discussion

Recent studies have shown that mature babies tend to have increased and more stable OS levels (Hofstetter et al., 2007), whilst babies with complications of prematurity such as chronic lung disease are more likely to exhibit lower OS levels, leading to greater frequency of desaturation episodes (McEvoy et al., 1993). Using continuous OS monitoring by oximeters, we observed higher OS variability (cvOS, sdOS, and dstOS), and lower levels (meanOS) in babies with stage 3 ROP and above, or requiring laser therapy, which are in accordance with the findings from the previous studies.

Previous findings that increased oxygen fluctuations occur in the earlier weeks of life rather than later (York et al., 2004; Cunningham et al., 1995; Saito et al., 1993) are quite compatible with our observations at 28 and 29 weeks PMA for babies with

severe ROP. In those studies, which used intermittent sampling of oxygenation, the CV of arterial oxygen (York et al., 2004; Saito et al., 1993) and SD of oxygen tension (Cunningham et al., 1995) were significantly elevated in babies with threshold ROP (York et al., 2004) or with stage 3 or higher ROP (Cunningham et al., 1995; Saito et al., 1993) within the first two weeks after birth.

In an early study of ROP (Pierce et al., 1996), it was suggested that irregular hypoxic events in oxygenation may induce abnormal retinal development in the later phase of ROP. This was then supported by Di Fiore et al. (2010), who reported that babies who went on to develop severe ROP requiring laser therapy had higher hypoxic events, with an increasing trend during the first four to five weeks of life. Their data were recorded at 2-second sampling rate, and a desaturation or hypoxic event was defined as a fall in OS below 80% for duration between 10 seconds and 3 minutes. The babies were born at 24 to 27 weeks GA. In our study, we used a 10-second sampling rate, and defined dstOS as the percentage of OS readings below 80%. The babies had almost similar ages as those in the Di Fiore et al. (2010) study. Our data show higher dstOS in babies with severe ROP than in the other group of babies with an increasing trend observed during 28 to 29 weeks PMA (refer Table 5.1), which agree well with Di Fiore et al.'s (2010) results.

There have been numerous threads of evidence pointing out the importance of time course, starting from 30 weeks PMA, in relation to the progress of ROP. One study (Good et al., 2005) observed that the vision threatening disease started to develop from about 32 to approximately 41.5 weeks PMA. Another (Coats et al., 2000) found that progression to severe ROP occurred at a median of 35 weeks PMA in preterm babies born less than 25 weeks GA and of birth weight up to 1250 gram. The progress of ROP from phase one to phase two (as described in Section 4.2) has been suggested to be associated with oxygen-regulated growth factors, VEGF and IGF-1 (Smith, 2005; Chen and Smith, 2007). Hellstrom and colleagues (2003) measured the mean of IGF-1 prospectively at a weekly interval; the growth factor was most deficient at 30-33 weeks PMA among severe ROP babies compared to less severe or no ROP babies. They concluded that low IGF-1 levels after preterm birth appeared to inhibit retinal vascular development. Smith (2005) further showed that when the cessation of blood vessel development persisted until 30-32 weeks PMA,

hypoxia that precipitates the second phase of retinal detachments and loss of vision will occur. This second destructive phase can be avoided if some preventive measures are taken during the chain of events in the first phase.

These studies reiterated that the timing of changes in the course of ROP development is important. They showed that some of the factors contributing to the first phase of ROP were observed from 30 weeks PMA in a premature child. In our data, we noted that there were marked differences in the OS statistics between babies with and without severe ROP, with higher variability and lower levels of OS in severe ROP babies in the earlier weeks of PMA, particularly 28 and 29 weeks. It may therefore be possible to examine and hence detect the retinal disorder as early as 28 weeks PMA.

5.3 The response and predictors for logistic regression

The study population included 257 preterm babies from 23-27 weeks GA. They were admitted to five hospitals: Christchurch Women's Hospital, National Women's Hospital, Middlemore Hospital, Wellington Hospital, and Dunedin Hospital, between September 2006 and December 2009. This cohort represented 85.4% of the 301 BOOST-NZ babies that survived until 36 weeks PMA. The 257 babies were divided into two groups based on their individual highest stage of ROP. 41 babies were categorised into the severe ROP group ('1'), comprising stage 3 and above or requiring laser surgery. The remaining 216 were placed into the no severe ROP group ('0'), comprising babies with no ROP (115, 44.7%), ROP stage 1 (60, 23.3%), and ROP stage 2 (41, 15.9%), all of whom did not require laser surgery.

The percentage of severe ROP in extremely premature babies in New Zealand for the 40-month period was 13.6% (41/301), which was comparable with 15.0% of the Australian (Gunn et al., 2012) and 16.0% of the American (SUPPORT Group, 2010) studies. Severe ROP was also seen in 14.8% in a German (Schwarz et al., 2011) study. These figures have been summarised in Chapter 2, Table 2.1. When comparing the incidence of severe ROP in New Zealand over a 12-year period, it was noted that the incidence based on BOOST-NZ data was higher ($P=0.0079$) than the rate 6.9% (20/288) seen in the 1998-1999 study (Cust et al., 2003). Overall, premature babies survival improved slightly, from 79% in 1998-1999 to 89% in 2006-2009 ($P=0.0004$).

With relatively small sample size in both studies, the observed increased (from 42% to 65%) survival of 23 weeks GA babies was not significant ($P=0.2021$). However, there was a particular rise (from 68% to 84%) in survival of 24-25 weeks GA babies ($P=0.0014$).

Recent reports have indicated that over the past two decades, incidences of certain prematurity problems including respiratory and ROP illness have been rising in developed countries (Slidsborg et al., 2008). Such complications, however, are in accordance with the improved survival rate of very premature babies due to improvement in fetal monitoring technology and neonatal intensive care (Horbar et al., 2012). Our results support these trends, showing more premature babies survived from 1998 to 2009.

5.3.1 Characteristics of severe ROP and no severe ROP babies

Table 5.4 summarises seven postnatal variables, and four OS statistics according to the ROP status; these variables are the predictors for the regression models. The entries for weight and GA in Table 5.4 show that babies in the severe ROP group tend to be smaller ($P=3.2E-08$) and younger ($P=1.8E-07$). This is also observed in the composite variable wGA, where the percentage of severe ROP babies is higher in the <25th percentile category (34% vs. 16%, $P=0.012$). Additionally, they have significantly higher duration of ventilation (mean 28 days vs. 9 days, $P=2.2E-11$; median 23 vs. 4 days, $P=6.5E-07$) and supplemental oxygen (mean 102 days vs. 52 days, $P=1.7E-09$; median 100 vs. 43, $P=2.3E-06$). Since the severe ROP babies require extended intensive care due to late recovery from respiratory problems, this might perhaps explain for their dependency on oxygen at 36 weeks PMA ($P=4.3E-05$). Prematurity, weight, and prolonged exposure to oxygen therapy have been consistently identified as the risk factors for ROP (Wilkstrand et al., 2011; Di Fiore et al., 2010).

Overall, there are more male babies in the study (139, 54%), but there is no significant difference between the severe ROP and no severe ROP babies ($P=0.688$). As discussed previously in Section 5.2, the mean values (at 29 weeks PMA) between the two groups differ for cvOS, sdOS, dstOS, and meanOS.

Table 5.4: Summary of severe ROP and no severe ROP babies

Variable	Severe ROP (n = 41) ^a	No severe ROP (n = 216) ^a	<i>p</i> -value ^b
cvOS	0.068 (0.017)	0.058 (0.019)	1.2E-03
sdOS (%)	6.18 (1.44)	5.32 (1.64)	1.9E-03
meanOS (%)	90.7 (2.53)	92.5 (2.82)	1.3E-04
dstOS	0.078 (0.044)	0.049 (0.049)	5.0E-04
	[0.079 (0.048, 0.110)]	[0.035 (0.013, 0.072)]	5.2E-05
weight (g)	730 (157)	917 (199)	3.2E-08
GA (weeks)	25.3 (1.2)	26.3 (1.1)	1.8E-07
daysVentilate (days)	28 (24)	9 (13)	2.2E-11
	[23 (4, 41)]	[4 (1, 11)]	6.5E-07
daysO2 (days)	102 (65)	52 (43)	1.7E-09
	[100 (45, 143)]	[43 (17, 79)]	2.3E-06
O2W36	26 (63%)	65 (30%)	4.3E-05
wGA (percentile)			0.012
>75th ^c	8 (20%)	76 (35%)	
25th to 75th	19 (46%)	106 (49%)	
<25th	14 (34%)	34 (16%)	
gender	21 (51%)	118 (55%)	0.688

^aEntries in table are mean(SD) and [median(interquartile range)] for continuous data; and frequency count (%) for categorical data.

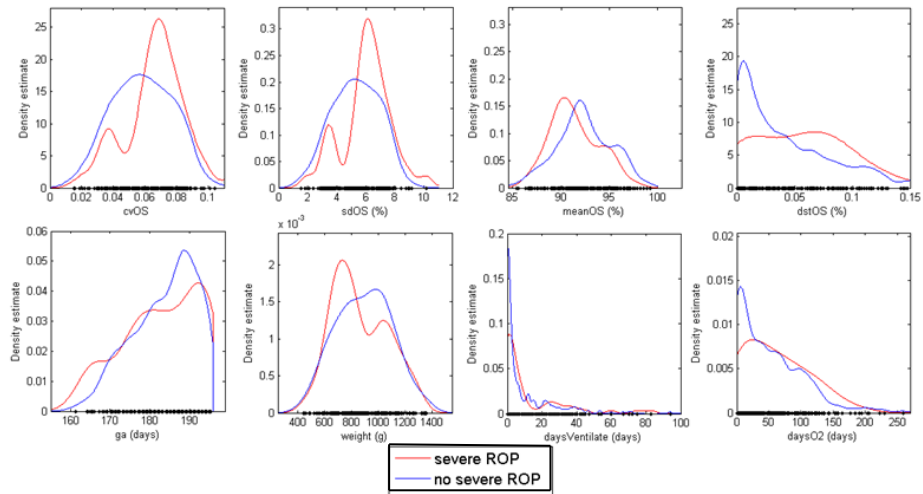
^b*p*-values are from *t*-test for difference in means, Mann-Whitney test for difference in medians, and chi-squared test for difference in frequency counts.

^cReference group for wGA

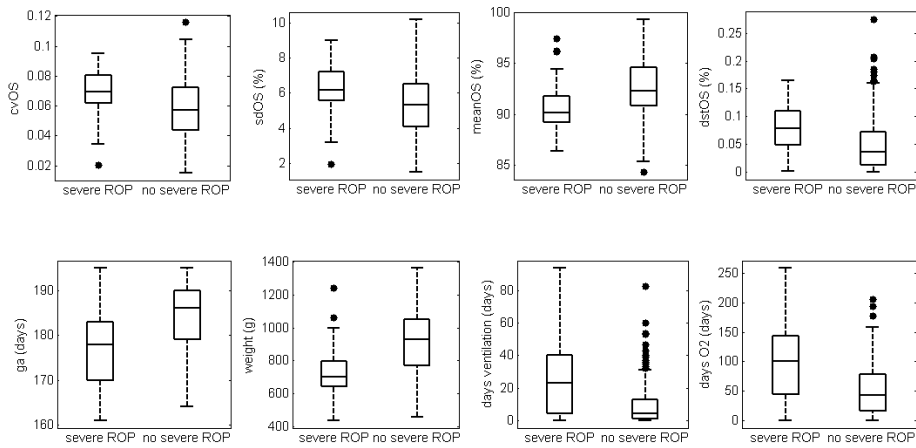
The characteristics of the continuous variables were further explored by examining their densities, using the kernel density estimators and box plots. Figure 5.2(a) illustrates the density of each variable according to ROP status. The red line represents the density for the severe ROP group, and the blue line for the no severe ROP group. The density for dstOS, daysVentilate and daysO2 is notably skewed, which lend support for using the Mann-Whitney test, median, and interquartile summaries in Table 5.4 above.

There is a higher probability mass in regions corresponding to greater variability in severe ROP babies at 29 weeks PMA (refer Figure 5.2(a) for cvOS, sdOS and dstOS), which is consistent with increased variability of OS in severe ROP cases observed in a number of studies (Di Fiore et al., 2010; York et al., 2004; Cunningham et al., 1995; Saito et al., 1993). Additionally, it was observed that babies who developed severe ROP also had lower OS levels as can be seen from the mode for severe ROP, which is to the left of the mode for no severe ROP (refer from Figure 5.2(a) for meanOS). The OS level is still being actively studied to determine its influence on the progression of ROP. Previous trials (Anderson et al., 2004; Sun et al., 2002; Tin et al., 2001) focused on assessing the effects of lower (85-89%) and higher ranges (91-95%) of OS on the occurrence of ROP (refer Section 2.3.1). A policy of lower OS less than 93% had been suggested to control severe cases (Chow et al., 2003). However, a recent trial (Support Study Group, 2010) has reported a higher rate of death in the lower OS group; therefore, expressing concern for the increasing support of lower range of OS. Note that these trials have not considered the issue of whether optimum OS would differ according to PMA.

The pairwise box plots in Figure 5.2(b) depict that there are potentially outlying points occurring in all variables except GA. Upon closer inspection, most of these potential outliers correspond to the same babies. The impacts of these observations will be further investigated and discussed in Section 5.4.2.



(a)



(b)

Figure 5.2: Kernel density plots (a), and box plots (b) for continuous variables of NZ-ROP according to ROP status

5.3.2 Collinearity among the predictors

The variables were checked before regression models were fitted in order to detect whether collinearity exists among them. Figure 5.3 displays scatterplots for each pair of continuous variables with binary ROP status encoded in red and blue colours. In addition, estimate of the Pearson correlation (95% CI) for each pair of variables are shown in Table 5.5. From both Figure 5.3 and Table 5.5, we immediately observe that all OS statistics (cvOS, sdOS, meanOS, and dstOS) exhibit strong correlation with absolute value greater than about 0.8, and that the correlation for daysVentilate and daysO2 is also quite strong being close to 0.8.

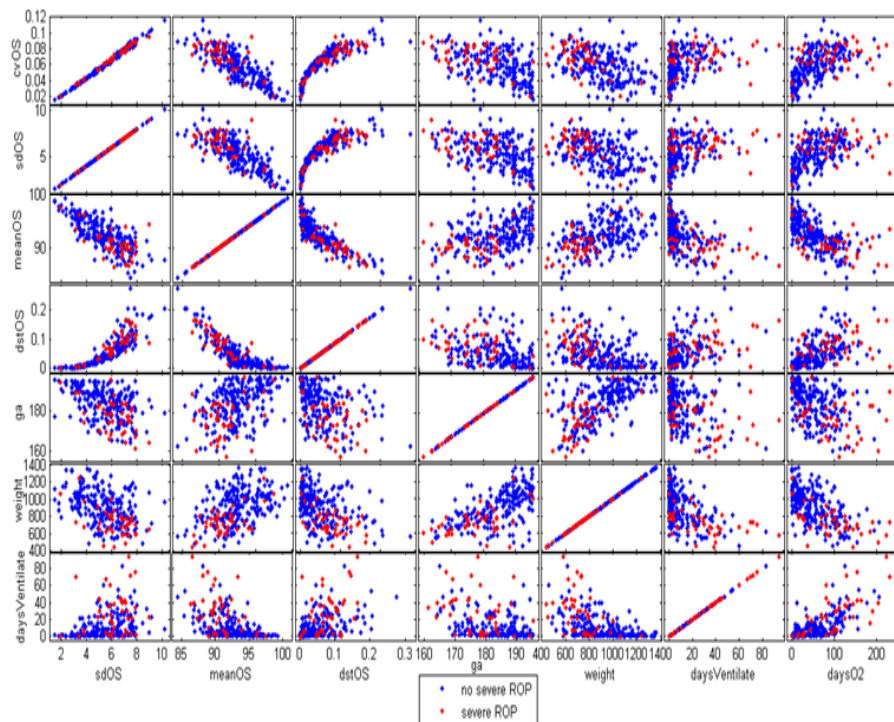


Figure 5.3: Scatterplots for each pair of predictors in NZ-ROP with binary ROP status encoded in red and blue (red dots indicate severe ROP cases and blue dots indicate no severe ROP cases)

Table 5.5: Estimates of Pearson correlation (with 95% CI) for each pair of predictors

Variable	sdOS	meanOS	dstOS	GA	weight	days Ventilate	daysO2
cvOS	0.997 (0.997, 0.998)	-0.788 (-0.830, -0.736)	0.881 (0.850, 0.906)	-0.469 (-0.560, -0.368)	-0.430 (-0.525, -0.325)	0.385 (0.275, 0.484)	0.501 (0.403, 0.587)
sdOS		-0.746 (-0.796, -0.686)	0.849 (0.811, 0.880)	-0.465 (-0.556, -0.363)	-0.415 (-0.512, -0.309)	0.361 (0.250, 0.463)	0.476 (0.376, 0.566)
meanOS			-0.850 (-0.881, -0.812)	0.411 (0.304, 0.508)	0.475 (0.374, 0.564)	-0.485 (-0.573, -0.386)	-0.617 (-0.687, -0.535)
dstOS				-0.432 (-0.526, -0.327)	-0.442 (-0.535, -0.338)	0.463 (0.361, 0.554)	0.543 (0.450, 0.624)
GA					0.623 (0.542, 0.692)	-0.438 (-0.532, -0.334)	-0.457 (-0.549, -0.355)
weight						-0.512 (-0.597, -0.415)	-0.590 (-0.664, -0.504)
days Ventilate							0.738 (0.677, 0.789)

Table 5.6: VIF for eight continuous variables

Variable	VIF
weight	5.8
GA	2.9
gender	1.1
wGA	3.1
daysVentilate	2.4
daysO2	4.0
O2W36	2.0
cvOS	5851
sdOS	4616
meanOS	27
dstOS	39

Collinearity was also checked using VIF (Table 5.6). An inspection for each variable shows that the VIFs for cvOS, sdOS, meanOS, and dstOS, ranging from 27 to 5851, were high compared to the rule-of-thumb threshold of 10 suggested by Marquardt (1970). These outcomes were consistent with the earlier results from the estimates of the correlation.

Further examinations revealed that (i) there were problems with the sign of the parameter estimates i.e. whenever a regression model contained at least two of the OS statistics, (ii) only one parameter estimate was significant for models containing both daysVentilate and daysO2, and (iii) only one parameter estimate was significant for models containing both weight and wGA. All of these variables are of great interest to the clinicians for identifying patients who will require treatment for severe ROP.

These results indicate that substantial correlation existed between certain groups of variables: the four OS statistics, two oxygen therapy variables, weight and size for GA. All-subsets regression was employed, in which each regression model was fitted with a combination of at most six predictors, with at most one variable from each group of the strongly correlated variables as follows.

- GA
- gender
- O2W36
- one of cvOS, sdOS, meanOS, or dstOS
- one of daysVentilate or daysO2
- one of weight or wGA.

This restriction on the occurrence of predictors in each model was not far from the rule-of-thumb suggested by Harrell et al. (1984) that not more than about $\nu/10$ predictors should be considered in a model. Here, ν is the smaller of the sample sizes for the two groups of babies, which in our case is 41 babies for the severe ROP group. The resulting total number of possible models is 359.

5.4 Logistic regression model fitting and diagnostics

Results from diagnostics of model fits are presented in two parts. Firstly, verification of linear structure of continuous predictors is presented in Section 5.4.1. For predictors that appear to have nonlinear structure, fitting based on nonlinear models were considered. Secondly, the outcomes from detecting potential outlying observations are explained in Section 5.4.2.

5.4.1 Examination of structure of continuous predictors

The linearity in the structure of the continuous predictors was first checked using the partial residuals from the logistic regression, and was again examined using generalised logistic additive models (GAMs). Figure 5.4 shows the partial residual plots for the eight continuous variables. As it was difficult to discern a pattern in the plots of the residuals, a lowess curve was fitted through the points of each plot. The smooth for the eight predictors are relatively flat. Additionally, note that the residuals that occur mostly above the lowess smooth belong to the severe ROP group. A more careful look at the residuals indicates that there appear to be several strongly influential observations. Most of them are from severe ROP babies, which correspond to the same observations that were identified by the box plots in Section 5.3.1. The effects of the outliers are revisited in Section 5.4.2.

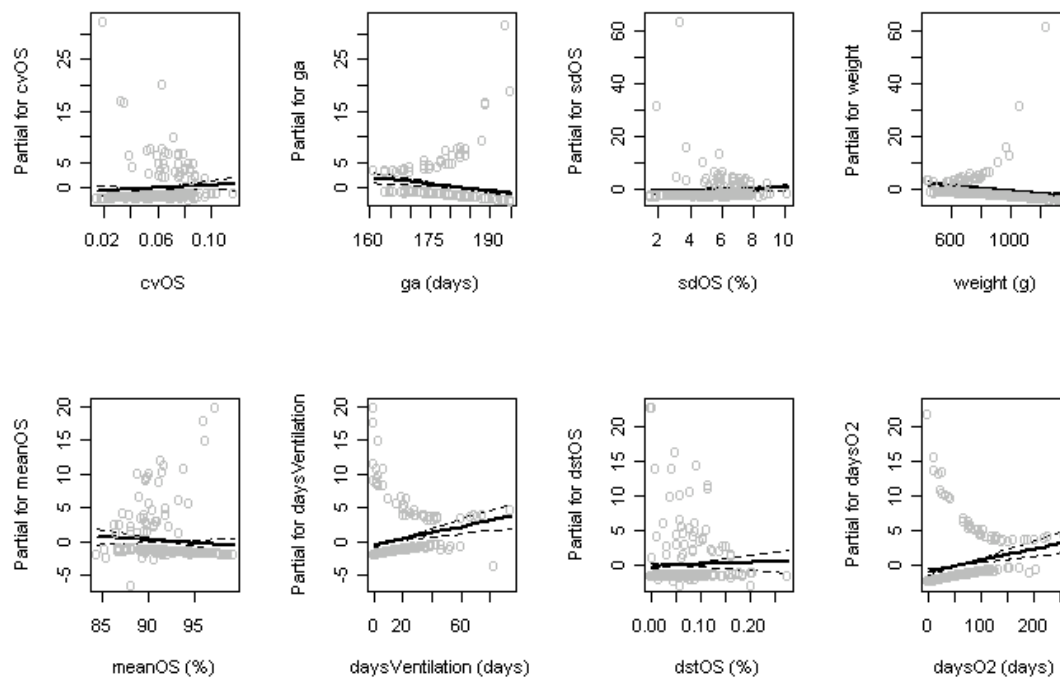


Figure 5.4: Partial residual plots of eight continuous predictors. The circles denote the residuals and the solid line depicts the smoothed residuals using a lowess smoother.

Additionally, GAM was used to check the linearity assumption for the predictors. Recall that the nonlinear term that we are using in GAM is based on cubic spline (refer Section 4.4.1). To check the structure of each predictor, several models based on the combinations of at least two predictors were fitted, and the GAM fit for each predictor was plotted separately; the structure for individual predictors was found to be quite similar among the models. Consequently, only results from models with two predictors are presented in this section.

Each two-predictor model contains a predictor and GA, which is well-known as an important risk factor for severe ROP. An examination to determine the appropriate fit for GA was then carried out. A GAM consisting of a nonlinear term for GA and a linear term for weight was fitted; weight was chosen as it is another important risk factor for severe ROP. Through the GAM fitting for GA (Figure 5.5), it is noted that the curve has an approximately monotonically decreasing pattern as GA increases. It is noted that the curve decreases steeply in two periods of GA: from 23 weeks GA (160 days) to about 24 weeks GA (170 days), and from 26 weeks to 27

weeks GA. To determine whether a linear structure is adequate for GA, a linear model (consisting of a linear term for GA and weight) and GAM were compared. The first two entries (denoted by Model 1) in Table 5.7 present the results of the fittings. A smaller AIC value for the linear model and a large p -value from the LR test ($P=0.3577$) lend a support on using a linear term for GA. Recall that the null hypothesis for the test states that simpler model holds.

Using this result, the structure for the remaining seven predictors was examined by fitting a GAM that contained each predictor fitted as a nonlinear term together with GA fitted as a linear term. The resulting curves of the predictors with the structures chosen by the GAM fits are displayed in Figure 5.5. For most of the predictors, there are wider ranges at the extremes of the standard-error curves, which are due to fewer data points at the extreme regions. The smooth appears to be linear for weight, daysVentilate, and daysO2. Risk appears to be decreasing monotonically as weight increases, with babies of weights less than 750 g facing higher risk of severe ROP. In contrast, daysVentilate and daysOS produce the opposite effect on this retinal disorder. The data show that the risk escalated rapidly after the babies were exposed to more than 20 days of ventilation, and 50 days of supplemental oxygen. The trends for the four OS statistics, however, warrant for a closer inspection. It can be seen that a straight line could fit between the pointwise standard error curves for cvOS and sdOS. On the other hand, the GAM fits of meanOS and dstOS seem to exhibit some form of nonlinear characteristics. To ascertain whether these patterns are the appropriate structure for the seven predictors, a further evaluation was undertaken by comparing a linear model (consisting of linear terms for GA and the predictor) against the GAM.

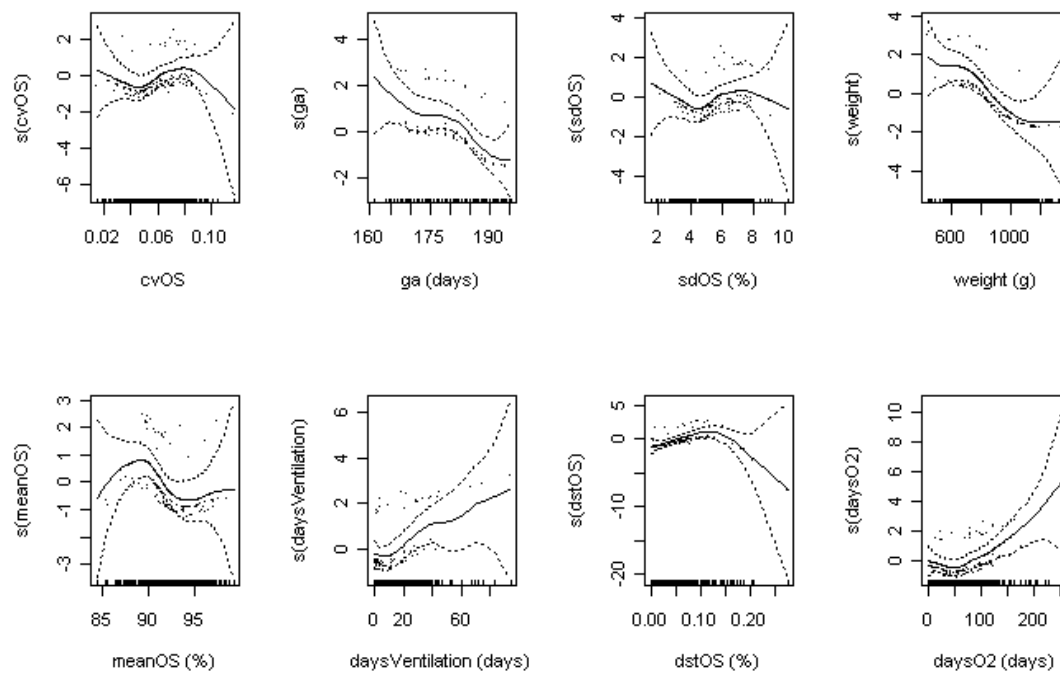


Figure 5.5: Plots of GAM fit for eight continuous predictors, with pointwise standard errors and partial residuals included. The black solid line represents the fitted cubic spline $s(\cdot)$, the dotted lines represent the upper and lower pointwise twice-standard-error curves, and the dots represent the residuals. The rug plots at the bottom of each graph show the occurrences of the data values.

The results from fitting linear model and GAM for each of the seven predictors together with GA (denoted by Model 2 to Model 8) are summarised in Table 5.7. Notice that the AIC for the linear model is no larger than the one for the GAM for weight, sdOS, daysVentilate, daysO2, and cvOS, which suggests that a linear term would suffice for these five predictors. An examination of the p -values of the LR test, which yielded large values ($P > 0.05$) re-affirm these results. Opposite results, however, were obtained for dstOS and meanOS. The AIC based on GAM is smaller, with significant p -value from the LR test ($P = 0.0209$) for dstOS, and marginally significant ($P = 0.0677$) for meanOS. These results indicate that a nonlinear term could be needed for the two predictors. From the shape of the GAM curves in Figure 5.5, it appears that linear piecewise or quadratic models could provide alternatives to the cubic spline. These simpler nonlinear parametric models are more parsimonious and easier to interpret.

Table 5.7: Comparison between models with linear term, and GAM with nonlinear term for continuous predictors

	Model ^b	AIC	<i>p</i> -value ^a
1.	weight+GA	194.5	0.3577
	weight+s(GA)	196.4	
2.	weight+GA	194.5	0.3590
	s(weight)+GA	196.5	
3.	cvOS+GA	204.8	0.1367
	s(cvOS)+GA	204.8	
4.	sdOS+GA	205.1	0.2272
	s(sdOS)+GA	206.1	
5.	meanOS+ GA	201.5	0.0689
	s(meanOS)+ GA	200.1	
6.	dstOS+ GA	204.3	0.0209
	s(dstOS)+ GA	200.6	
7.	daysVentilate+GA	190.1	0.5981
	s(daysVentilate)+GA	193.0	
8.	daysO2+GA	191.4	0.1798
	s(daysO2)+GA	191.9	

^a*p*-values are from the LR test for comparing a model with linear terms (first row) against a model with a nonlinear additive term (second row).

^b*s*(·) denotes a cubic spline.

Based on these analyses, it is concluded that six of the predictors would adequately be modelled as linear. The remaining two predictors, namely meanOS and dstOS required further investigation, which will be discussed next.

The fitted cubic spline for meanOS shows an interesting pattern of a steady increasing risk as the level of OS increases up to about 89% followed by a sharp decrease. Accordingly, the curve could be partitioned into two regions with a knot at 89% when fitted with a piecewise linear function. MeanOS was then examined

separately, using a piecewise linear model with a knot at k denoted as $\text{pw}(\cdot, k)$, and a quadratic model denoted as $\text{poly}(\cdot)$. To determine which model will be better for meanOS, a linear model (consisting of linear terms for meanOS and GA) and a piecewise linear model were compared using the LR test and AIC. Likewise, a separate comparison between a linear model and a quadratic model was performed. The results for these models (denoted by Model 1 and Model 2) are displayed in Table 5.8. No significant difference is found, either between a linear model and a two-segment piecewise linear model ($P=0.0987$), or between a linear model and a quadratic model ($P=0.5614$). These results are reflected in the models' AICs, which are very close together. The GAM fit that corresponds to meanOS also has a hint of a peak at 89% and a trough at 93%. Therefore, a three-segment piecewise linear model with knots at 89% and 93% was fitted (not shown in Table 5.8). This, however, yielded a similar outcome as obtained from the two-segment piecewise linear model. Therefore, we concluded that a linear model would suffice for meanOS.

Next, we investigated the nonlinear feature in dstOS. From Figure 5.5, it can be seen that there was a gradual increase in risk of severe ROP with increasing dstOS, and thereafter, a steady decrease as dstOS passes a threshold at about 0.15%. The decreasing trend in the upper range of dstOS is interesting as this suggests that the risk of developing severe ROP weakens gradually after dstOS increases to a certain level. We considered two nonlinear parametric models for dstOS: a piecewise linear model with a knot at 0.15% (thus having a positive slope in the left hand segment and a negative slope in the right hand segment), and a quadratic model. In both models, dstOS was fitted together with GA fitted as a linear term. Evaluation was made to determine which one of these models was most appropriate for dstOS. The results, denoted by Model 3 and Model 4, are represented in Table 5.8. The AIC for the two-segment piecewise linear model is much lower than that of the linear model; this is further supported by the small p -value ($P=0.0028$). An improved result, however, is not achieved from the quadratic model, which yields a higher AIC and a larger p -value ($P=0.0110$) as compared to the piecewise linear model; thus, justifying the preference for using a two-segment piecewise linear model for dstOS in the ROP analysis.

Table 5.8: Comparison between models with a linear term, and models with a nonlinear parametric term for meanOS and dstOS

	Model ^b	AIC	<i>p</i> -value ^a
1.	meanOS+GA	201.5	0.0987
	pw (meanOS, 89)+GA	200.7	
2.	meanOS+GA	201.5	0.5614
	poly (meanOS)+GA	203.1	
3.	dstOS+GA	204.3	0.0028
	pw(dstOS, 0.15)+GA	197.4	
4.	dstOS+GA	204.3	0.0110
	poly(dstOS)+GA	199.8	

^a*p*-values are from the LR test for comparing a model with linear terms (first row) against a model with a nonlinear parametric term for meanOS or dstOS (second row).

^bpoly(·) denotes a polynomial fit and pw(·, *k*) denotes a piecewise linear regression fit with a knot at *k*.

There appears to be a conflicting phenomenon on the contribution of dstOS to the risk, in which there is a decreasing trend as dstOS passes a threshold. This pattern seems to contradict with the clinicians' intuition that ROP risk should increase with increasing desaturation episodes of preterm babies (Di Fiore et al., 2010). Fetuses are regularly treated with prenatal steroids. These steroids have been suggested to provide a protective effect against the risk of ROP (Jagielska et al., 2008; Console et al., 1997). In this study, the medical history of the babies was examined; the clinical records of the 13 babies in the right-hand segment of dstOS showed that they had all received steroids during their prenatal period to promote fetal lung development. Hence, this provides the plausible explanation for the phenomenon of the negative slope of the right-hand segment.

5.4.2 *Detection of outlying observations*

The analyses that had been performed so far, had not taken into account the presence of outlying points. However, earlier on, we have noted the existence of several suspicious observations from the box plots (Figure 5.2(b) in Section 5.3.1) and partial residual plots (Figure 5.4 in Section 5.4.1). These observations could be candidate outliers, which are normally detected by identifying overly influential points and high leverage values.

In this section, results from the analysis of detecting outlying points are presented. The analysis was performed using a combination of conventional diagnostic measures and a robust measure (refer to Section 4.4.4). Identification of multiple outlying points involves a three-step approach. In the first exploratory step, all-subsets regression was fitted to the full data, and the sign of the parameter estimates in all models was examined. Results showed that “incorrect” sign in some of the estimates occurred in 21 models, each containing three or four predictors. Here, “incorrect” was defined as having the opposite sign to that indicated in the single-predictor analysis and generally indicated in the literature. In the second step, the conventional and robust diagnostic measures for these 21 models (with full data set) were fitted and plotted to detect all candidate outliers. The occurrences of suspected data points that exceeded the suggested thresholds of the diagnostic measures were recorded. In the third step, these observations were further examined by assessing their impact on the sign and the standard error of the parameter estimates. The list of outliers was then finalised using results from both steps two and three.

To identify data points that are candidate outliers, four diagnostic measures were used: approximated jackknife residual, Cook’s distance statistic, leverage from hat matrix, and Mallows robust weight, which is used in order to identify observations assigned with small weights. Recall in Section 4.4.2 that none of these diagnostic measures is reliable enough on its own to detect all outliers. This is illustrated in the following example (refer Figure 5.6 and Table 5.9) based on a model, with `dstOS`, `weight`, and `daysVentilate` as predictors. The jackknife residual plot shows that observations 130 and 157 are quite isolated from the main data. These observations, along with observations 204 and 207, also receive small Mallows weights, with values between 0.1 and 0.35. This indicates that the four observations are potentially

influential for both modelling and inference. Notice that two of them (observations 130, 207) attain the maximum values of the Cook statistic. In particular, there is a leverage point (observation 8), which is assigned with a high weight of one by the Mallows weight. Results from these plots suggest that there appear to be five candidate outliers, namely, observations 8, 130, 157, 204, and 207.

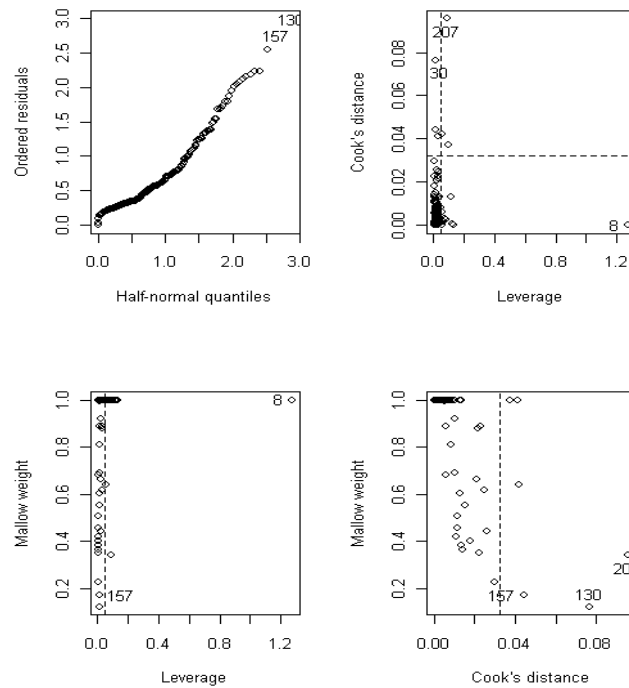


Figure 5.6: Diagnostic plots for model `dstOS+weight+daysVentilate`. Jackknife residuals (top left); Cook’s distance against leverage (top right); Marlow’s weight against leverage (bottom left); and Marlow’s weight against Cook’s distance (bottom right). The dotted lines denote the thresholds, $8/(n - 2p - 2)$ for Cook’s distance and $2(p + 2)/n$ for leverage. The numbers beside the circles correspond to the observation numbers.

Table 5.9: Identification of candidate outliers from diagnostic measures for model `dstOS+weight+daysVentilate`

Diagnostic measure	Candidate outliers
Jackknife residuals	130, 157
Cook’s distance	130, 207
Leverage	8
Mallows weight	130, 157, 204

A list of five potential outliers emerging from the 21 models in step two is produced in Table 5.10. Observations 130, 157, and 204 have the highest frequency among other observations, occurring in 20 of the 21 models (95%). Observations 8 and 207 are another two candidate outliers, which occur in 19 (90%) and 12 (57%) of the 21 models, respectively. This result provides evidence of aberrant observations that deserve further investigation.

Table 5.10: Candidate outliers and their occurrences in 21 models, as identified by four diagnostic measures

Candidate outlier	Occurrences in models (%)
130	20 (95)
157	20 (95)
204	20 (95)
8	19 (90)
207	12 (57)

It is known that atypical observations can have a large impact on the maximum likelihood estimates, leading to biased estimators and wrong interpretations of the fitted models (Rousseeuw and Leroy, 1987). In the next step, the effect of these candidate outliers on the sign and standard error of the parameter estimates was assessed. Using the list of potential outliers from step 2, the 21 models were re-fitted after removing the five observations. Some changes were observed in the models' estimates and standard errors, which resulted in appropriate signs for the estimates in all models. Taken together the results from steps two and three, observations 157, 130, 204, 8 and 207 were flagged as outliers.

The following analysis was carried out to determine on how best to handle the five observations. Three approaches were used: logistic regression on the full data set (denoted as MLE), robust logistic regression on the full data set (denoted as Robust), and logistic regression with the five outliers omitted (denoted as MLE2). The robust estimation of the parameters is via the Mallows estimator, as defined by equation (4.11) in Section 4.4.3, with tuning constant $t = 1.2$ and weight

function $w(X_i) = \sqrt{1-h_i}$. The various results yielded by these approaches are demonstrated in the following cases.

Case I

The results in Table 5.11 summarise a case of an incorrect sign of a parameter estimate for predictor meanOS, produced by MLE as shown in the first column. Nevertheless, the sign changes when the robust method is used or when the outliers are removed. The estimates from both MLE and robust methods are quite different, in that the estimated standard errors of the non-robust method are smaller, and the parameter estimates are somewhat larger than their robust counterparts. For the remaining three predictors, it can be seen that the parameter estimates of MLE2 are relatively close to the MLE estimates.

Table 5.11: Comparison of parameter estimates between logistic and robust logistic estimation for a model, meanOS+weight+daysVentilate+O2W36

Predictor	MLE ^a		Robust ^a		MLE2 ^b	
	Parameter estimate	SE	Parameter estimate	SE	Parameter estimate	SE
Constant	-0.625	7.859	14.605	11.063	8.507	8.739
meanOS	0.013	0.088	-0.149	0.123	-0.081	0.097
weight	-0.004	0.001	-0.004	0.002	-0.004	0.001
daysVentilate	0.030	0.012	0.012	0.017	0.034	0.014
O2W36	0.496	0.451	0.230	0.593	0.287	0.486

Model is estimated with ^afull data or with ^bexclusion of observations 130, 157, 8, 204, and 207.

Case II

A second case is illustrated based on a model, with sdOS, GA, daysO2, and wGA as predictors. Table 5.12 shows a surprising finding when the robust method estimates a negative sign for sdOS. A close inspection at Mallows robust weight showed that observation 204 was not assigned with a low weight. A positive sign is obtained only when observation 204 is removed from the analysis. The re-fitted

MLE2 model, which excludes the observation, takes into account the atypical nature of the observation in the estimation procedure.

Table 5.12: Comparison of parameter estimates between logistic and robust logistic estimation for a model, sdOS+GA+daysO2+wGA

Predictor	MLE ^a		Robust ^a		MLE2 ^b	
	Parameter estimate	SE	Parameter estimate	SE	Parameter estimate	SE
Constant	12.720	5.384	13.128	6.062	13.069	5.393
sdOS	0.004	0.146	-0.027	0.167	0.045	0.151
GA	-0.087	0.028	-0.089	0.031	-0.090	0.028
daysO2	0.011	0.004	0.013	0.005	0.010	0.004
wGA (pct):						
25th to 75th	0.366	0.491	0.099	0.554	0.391	0.491
<25th	1.040	0.575	0.962	0.625	1.037	0.576

Model is estimated with ^afull data or with ^bexclusion of observation 204.

Case III

A further example based on cvOS, weight, and gender as predictors (given in Table 5.13) shows the effect of observations 130, 157, 8, 204, and 207. These observations seem to increase the MLE estimate of gender. The negative sign for gender, which is correct for NZ-ROP data, is attained only when the five observations are removed from the data.

Table 5.13: Comparison of parameter estimates between MLE and robust estimation for a model, cvOS+weight+gender

Predictor	MLE ^a		Robust ^a		MLE2 ^b	
	Parameter estimate	SE	Parameter estimate	SE	Parameter estimate	SE
Constant	1.535	1.329	0.946	1.409	1.181	1.408
cvOS	13.468	11.411	18.496	12.414	28.290	12.750
weight	-0.005	0.001	-0.005	0.001	-0.005	0.001
gender	0.040	0.371	0.121	0.395	-0.0208	0.396

Model is estimated with ^afull data or with ^bexclusion of observations 130, 157, 8, 204, and 207.

With these results, the all-subsets logistic models were re-fitted to the NZ-ROP data. The model fittings excluded the previously identified five outlying observations. Diagnostic checking (as in Section 5.4.1) was again undertaken; the parametric forms of the predictors are almost similar to those obtained in the earlier analysis. Figure 5.7 depicts the GAM fits for the OS statistics. Overall, there is an increasing trend of risk that corresponds to cvOS and sdOS, and decreasing trend for meanOS, whilst the nonlinear form for dstOS appears to be clearer. The OS statistics for variability have a downward trend for larger values; this trend, however, is more prominent in dstOS. Additionally, there was little evidence of any distinct outlying observations in the reduced data set. The performance of these models was then assessed and the results are presented in the next section.

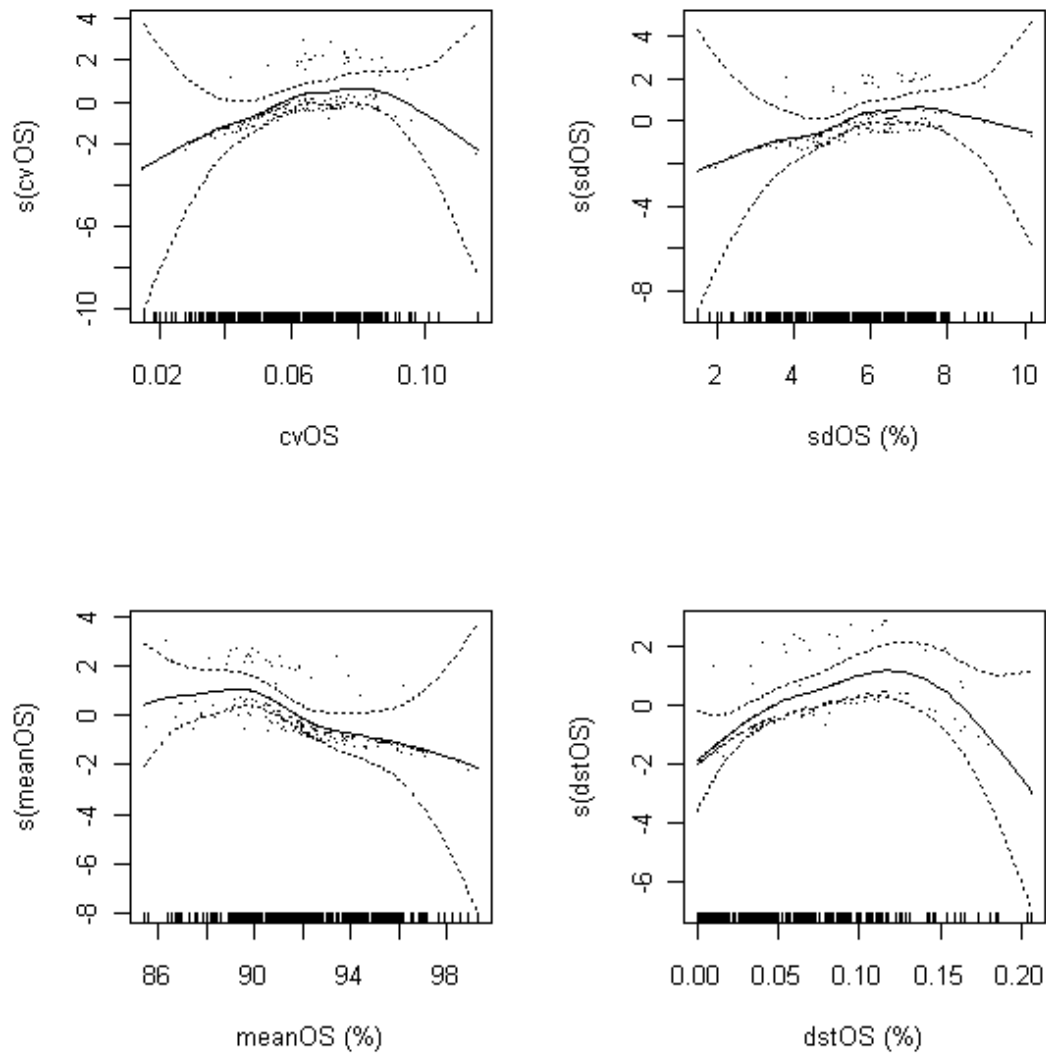


Figure 5.7: Plots of GAM fit (excluded observations 157, 130, 204, 8, and 207) for four statistics of OS (cvOS, sdOS, meanOS, and dstOS), with pointwise standard errors and partial residuals included. The black solid line represents the fitted cubic spline $s(\cdot)$, the dotted lines represent the upper and lower pointwise twice-standard-error curves, and the dots represent the residuals. The rug plots at the bottom of each graph show the occurrences of the data values.

5.5 Comparative performance of models

The logistic regression models were fitted using all-subsets regression, which performed predictor selection by fitting separate logistic regressions on the combinations of at most six predictors. In total, 359 models were fitted, which included the single-predictor and multi-predictor regression models. The predictors were seven clinical and postnatal factors and four OS statistics, whilst the response variable was severe ROP or no severe ROP. Predictors that were most influential in discriminating babies at risk of developing severe ROP were assessed and identified. These are discussed in the following sections.

5.5.1 Single-predictor models

The parameter estimates ($\hat{\beta}$) for single-predictor logistic regression models are reported for each of the eleven predictors in Table 5.14, accompanied by their respective SEs, p -values, and ORs (depicted in Figure 5.8). With the exception for gender ($P=0.595$), all predictors were significant for severe ROP ($P<0.05$), with meanOS, weight and GA showing protective associations. Strong relation ($P<1.0E-6$) with severe ROP was demonstrated by the following predictors: weight, GA, daysVentilate, daysO2, and dstOS. The first three variables are well known risk factors for severe ROP. Within the wGA category, babies born with lower weight for GA had a higher risk of ROP (OR = 4.15, 95% CI = (1.52, 11.31)). There were relatively few babies (Table 5.15) in this category (47 out of 252 or 18.7%), but their severe ROP rate was high (13 out of 47 or 27.7%).

Table 5.14: Single-predictor logistic regression models of NZ-ROP

Predictor	$\hat{\beta}$ (95% CI)		SE	<i>p</i> -value ^c	Unadjusted OR ^e (95% CI)
cvOS	43.00	(21.63, 64.37)	10.90	8.0E-5	3.406 (1.853, 6.261)
sdOS (%)	0.485	(0.236, 0.734)	0.127	1.3E-4	3.288 (1.784, 6.057)
meanOS (%)	-0.338	(-0.489, -0.187)	0.077	9.9E-6	0.286 (0.163, 0.500)
dstOS (%) ^a				1.6E-7 ^d	
< 0.15%	25.38	(15.76, 35.00)	4.909	2.3E-7	4.380 (2.502, 7.668)
≥ 0.15%	-128.7	(-232.1, -25.27)	52.77	0.015	0.031 (0.002, 0.505)
weight (g)	-0.007	(-0.009, -0.005)	0.001	8.7E-8	0.129 (0.070, 0.240)
GA (days)	-0.131	(-0.180, -0.082)	0.025	1.1E-7	0.194 (0.105, 0.358)
daysVentilate (days)	0.064	(0.040, 0.088)	0.012	2.5E-8	2.884 (1.957, 4.252)
daysO2 (days)	0.020	(0.012, 0.028)	0.004	1.2E-7	4.446 (2.499, 7.911)
O2W36	1.506	(0.775, 2.237)	0.373	5.5E-5	4.509 (2.170, 9.366)
wGA (percentile)				0.017 ^d	
> 75th ^b					
25th to 75th	0.631	(-0.290, 1.552)	0.470	0.180	1.879 (0.748, 4.721)
< 25th	1.423	(0.419, 2.427)	0.512	0.005	4.150 (1.521, 11.31)
gender	-0.187	(-0.877, 0.503)	0.352	0.595	0.829 (0.416, 1.653)

^a dstOS is fitted with a piecewise linear function with a knot at 0.15%.

^b Reference group

^c *p*-value is from the Wald's test.

^d *p*-value is from the LR test.

^e The unadjusted ORs for continuous variables depict the odds of severe ROP for patients at the 75th percentile of the distribution of the variable versus patients at the 25th percentile.

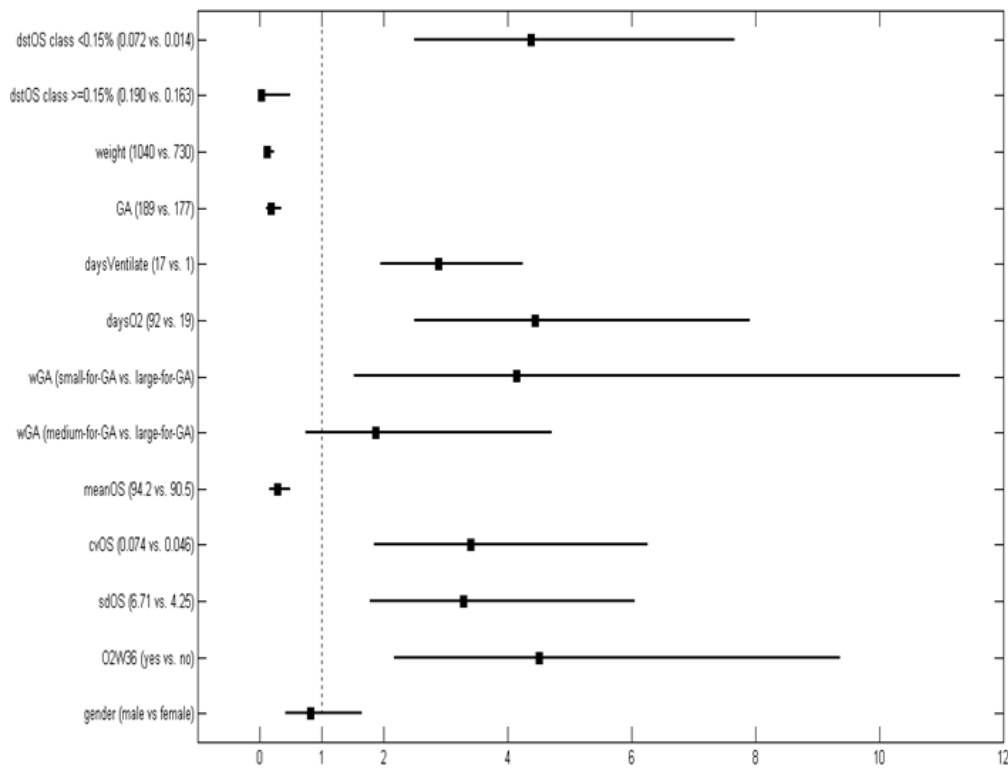


Figure 5.8: Unadjusted ORs (with 95% CIs) of severe ROP based on 252 babies in NZ-ROP for all predictors. For continuous variables, the value depicted reflects the odds of severe ROP for babies at the 75th percentile of the distribution of the variable versus patients at the 25th percentile.

Table 5.15: Predictor weight-for-GA and severe ROP

	Severe ROP no./ total (%)
wGA (percentile)	
> 75 th ^a	7/83 (8.4%)
25 th to 75 th	18/122 (14.8%)
< 25 th	13/47 (27.7%)

^a Reference group

The $\hat{\beta}$ for meanOS is negative, which suggests that severe ROP cases are less likely with higher OS levels at 29 weeks of PMA. For OS statistics for variability (cvOS, sdOS and the lower range of dstOS), their $\hat{\beta}$ s are all positive; thus indicating

that higher variability implies a higher probability of severe ROP. *dstOS* is the most dominant predictor ($P=2.3E-7$) among the four OS statistics, with the risk of severe ROP increased with increasing *dstOS* up to 0.15% (OR = 4.380, 95% CI = (2.502, 7.668)). Figure 5.9 shows the box plots of fitted probabilities of severe ROP and no severe ROP calculated from the model containing *dstOS*; it can be seen that the two groups of babies are well-separated. These results lend support to the clinicians' suspicion that OS fluctuations could be an early indicator of ROP.

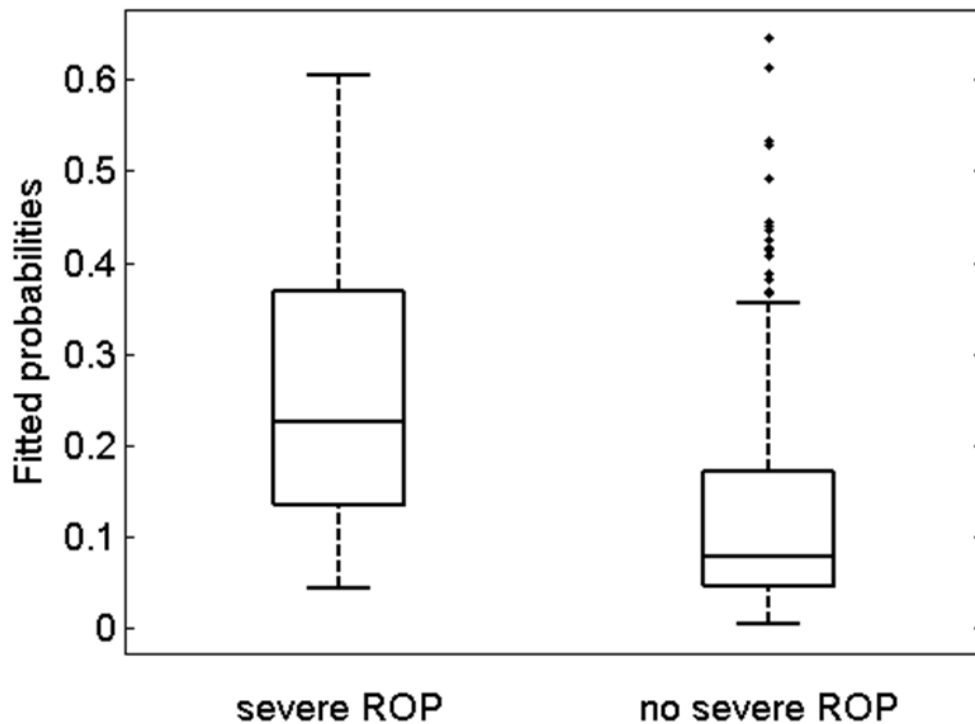


Figure 5.9: Box plots of fitted probabilities of severe ROP and no severe ROP for a single-predictor model, *dstOS*. The 25th, 50th, and 75th quartiles are 0.1361, 0.246, and 0.370 for severe ROP, and 0.047, 0.079, and 0.171 for no severe ROP, respectively.

5.5.2 Multi-predictor models

The all-subsets regression comprised of 359 models. Potential interaction was assessed among selective predictors such as between OS statistics and gender, and O2W36. However, no appreciable interaction effect was detected. Table 5.16 shows 15 models listed according to their AUCs; the first ten were the top models, whilst the remaining five contained several combinations of OS statistics and risk factors that clinicians felt were important for ROP. The predictors that were highly related to

severe ROP in the single-predictor analysis remained important here. After adjustment with other key clinical factors, strong relationships with severe ROP were shown by dstOS, GA, weight, and daysVentilate; dstOS appears in all the top ten models and each of the other three appears in eight of these models. Other variables such as gender, O2W36, wGA, and daysO2 appear less often (four times for gender, three times for O2W36, and only once for wGA and daysO2). dstOS was the most important predictor among the four OS statistics, as indicated by its presence in the top models.

Table 5.16: Selected multi-predictor regression models of NZ-ROP

Rank ^a	Models ^c	AUC (95% CI)
1 ^b	dstOS, weight, daysVentilate	0.853 (0.792, 0.914)
2	dstOS, GA, weight, daysVentilate	0.853 (0.792, 0.914)
3	dstOS, weight, gender, daysVentilate	0.852 (0.792, 0.912)
4 ^b	dstOS, GA, weight, O2W36	0.851 (0.787, 0.915)
5	dstOS, GA, weight, daysO2	0.851 (0.790, 0.912)
6	dstOS, GA, weight, daysVentilate, O2W36	0.851 (0.788, 0.914)
7	dstOS, GA, weight, gender, daysVentilate	0.849 (0.790, 0.909)
8	dstOS, GA, gender, daysVentilate, wGA	0.848 (0.786, 0.911)
9	dstOS, GA, weight, gender, daysVentilate, O2W36	0.848 (0.785, 0.911)
10 ^b	dstOS, GA, daysVentilate	0.847 (0.788, 0.905)
37 ^b	dstOS, GA, wGA	0.835 (0.771, 0.899)
63 ^b	meanOS, GA, weight	0.829 (0.766, 0.893)
107 ^b	cvOS, weight, O2W36	0.822 (0.753, 0.891)
115 ^b	sdOS, weight, O2W36	0.821 (0.753, 0.889)
176 ^b	meanOS, GA, wGA	0.815 (0.744, 0.886)

^a Actual rank from 359 models.

^b *p*-value of each predictor is less than 0.05.

^c dstOS is fitted with a piecewise linear function with a knot at 0.15%.

In Table 5.16, there are several models containing predictors with p -values less than 0.05 and AUCs well above 0.8. Note that the performance of individual models depends on the combination of predictors in each model. The box plots of the probabilities of severe ROP and no severe ROP predicted using the top model with `dstOS`, `weight`, and `daysVentilate` as predictors are shown in Figure 5.10.

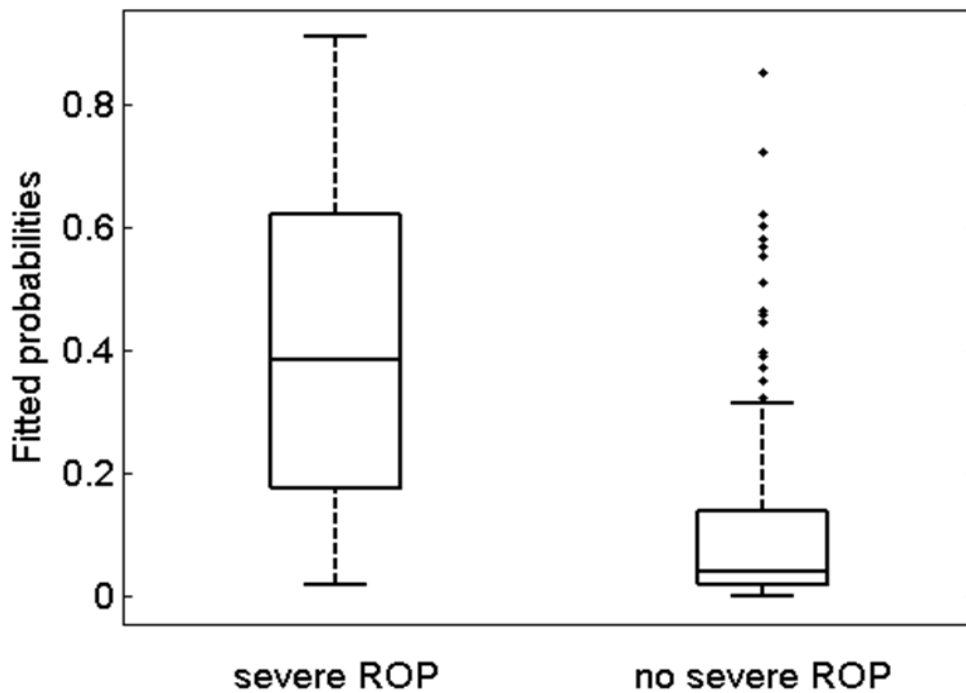


Figure 5.10: Box plots of fitted probabilities of severe ROP and no severe ROP for a multi-predictor model, `dstOS+weight+daysVentilate`. The 25th, 50th, and 75th quartiles are 0.174, 0.384, and 0.622 for severe ROP, and 0.016, 0.039, and 0.138 for no severe ROP, respectively.

With 11 predictors being investigated, we aim to pare down the variables to a smaller number that impacts the outcome the most, and hence, useful for discriminating between babies with good or poor outcome. One possible way is to consider a number of potential models, which differ in terms of combination of predictors and predictions for babies at risk of developing severe ROP, as seen from Table 5.16, Figure 5.9, and Figure 5.10. Alternatively, rather than to base inference on only one or several models, a weighted average of the models could be employed to circumvent the need for predictor or model selection. Using this approach, all models,

hence all predictors were taken into account and weighted accordingly. The results for NZ-ROP using a model averaging strategy are illustrated in the next section.

5.5.3 Model averaging approach

To determine predictors that have a strong relationship with severe ROP, a model averaging approach (refer Section 4.3.4) was adopted, in which model-averaged p -values were calculated. Additionally, the OR (with 95% CI) for each predictor, and fitted probabilities for the risk of severe ROP were also computed by averaging over all models. Together with the analyses of the single- and multi-predictor models, the results are discussed in Section 5.6.

Model-averaged p -values and ORs

With the model-averaged p -value, the relative dominance of the predictors with respect to their relationship with severe ROP can be quantified. Table 5.17 presents the strength of each predictor, as ranked according to their corresponding p -values. Highly influential variables in this group of predictors, with model-averaged p -values less than 0.01, were lower range of dstOS ($< 0.15\%$), weight, GA, and daysVentilate. Other variables such as the upper range of dstOS ($\geq 0.15\%$) and daysO2 were marginally associated with the risk of severe ROP. From Table 5.17, it can be seen that (i) dstOS was the strongest predictor among the four OS statistics, (ii) weight and GA were more important individually than when combined as wGA, and (iii) daysVentilate was the dominant predictor among the three oxygen therapy variables, i.e. daysO2 and O2W36. Overall, these predictors represent a subset of risk factors that could be clinically useful in identifying babies potentially at risk of severe ROP.

Table 5.17: Model-averaged p -values and ORs of severe ROP based on 252 babies in NZ-ROP

Predictor	p -value ^a	OR ^a (95% CI)
dstOS (< 0.15%) ^b	0.0073	2.800 (1.455, 5.399)
weight	0.0075	0.261 (0.102, 0.685)
GA	0.0088	0.318 (0.157, 0.646)
daysVentilate	0.0093	2.045 (1.330, 3.146)
dstOS (\geq 0.15%) ^b	0.0423	0.020 (0.001, 0.883)
daysO2	0.0816	2.527 (1.244, 5.170)
wGA (< 25th percentile)	0.1154	3.305 (1.011, 10.85)
meanOS	0.1371	0.563 (0.287, 1.107)
cvOS	0.1428	1.932 (0.941, 3.975)
sdOS	0.1457	1.926 (0.933, 3.986)
O2W36	0.2447	2.097 (0.849, 5.230)
wGA (25th to 75th percentile)	0.2568	1.847 (0.657, 5.198)
gender	0.3472	0.670 (0.298, 1.507)

^a p -values and ORs are obtained from model averaging.

^b dstOS is fitted with a piecewise linear function with a knot at 0.15%.

Table 5.17 also provides the model-averaged ORs for severe ROP for each predictor; the results are depicted in Figure 5.11. The OR for dstOS (< 0.15%), of which the predominant relation with severe ROP has been noted earlier (Table 5.16), was quite pronounced. Even after adjustment for all other clinical predictors, preterm babies at the upper quartile of the dstOS (< 0.15%) distribution were nearly three times more likely to develop severe ROP than babies at the lower quartile, (OR = 2.800, 95% CI = (1.455, 5.399)). The presence for high ROP risk, however, was not apparent in the other three OS statistics, or in O2W36 and gender.

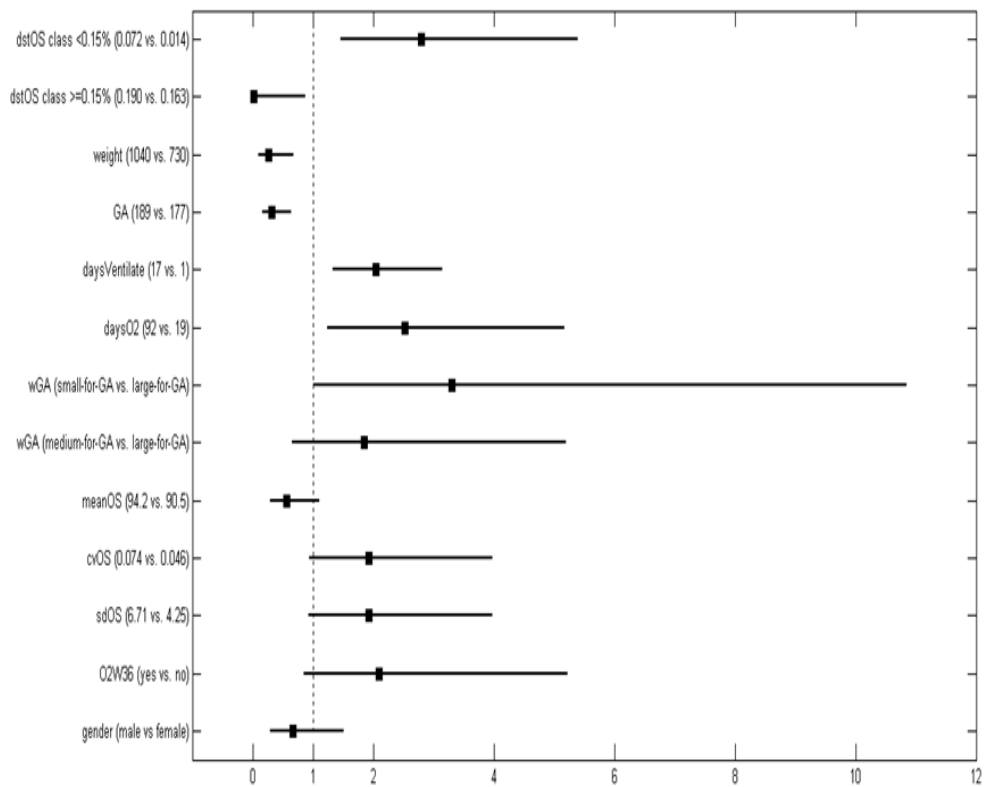


Figure 5.11: Model-averaged ORs (with 95% CIs) of severe ROP based on 252 babies in NZ-ROP for all predictors. For continuous variables, the value depicted reflects the odds of severe ROP for babies at the 75th percentile of the distribution of the variable versus patients at the 25th percentile.

Model-averaged fitted probabilities

A summary of the model-averaged fitted probabilities is presented in Figure 5.12. The upper quartile of the estimated probabilities for the severe ROP group was 0.426, whilst the lower quartile for the no severe ROP group was 0.035. Overall, a good separation exists between the two groups (Kolmorov-Smirnov two-sample test, $P=2.9E-8$). This suggests that the model averaging approach performs well in discriminating between babies who will or will not develop severe ROP.

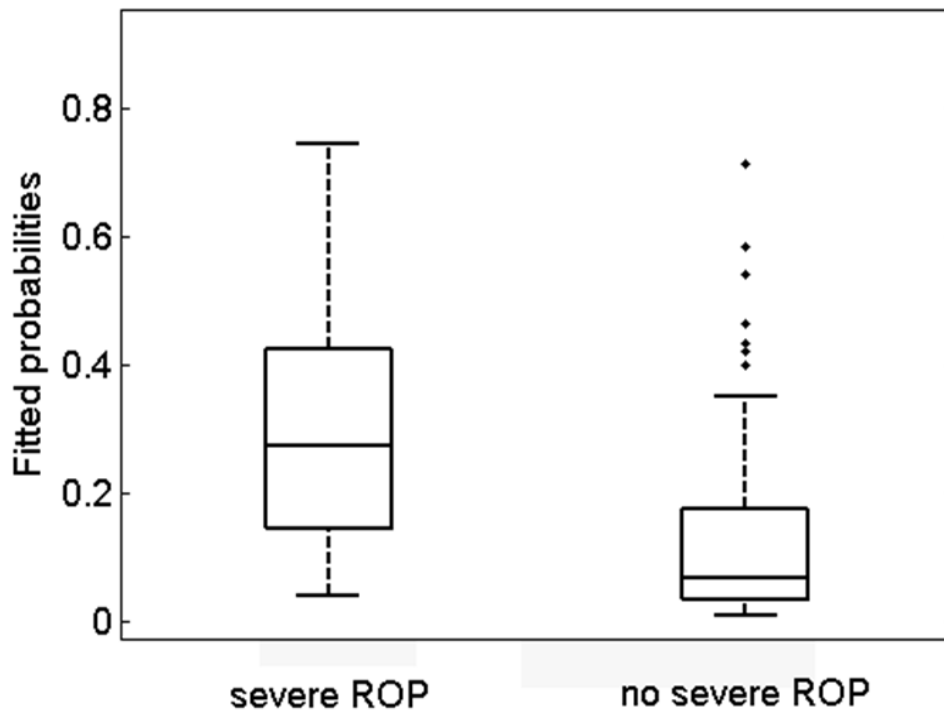


Figure 5.12: Box plots of fitted model-averaged probabilities of severe ROP and no severe ROP. The 25th, 50th, and 75th quartiles are 0.147, 0.275, and 0.426 for severe ROP, and 0.035, 0.069, and 0.176 for no severe ROP, respectively.

5.6 Discussion

The influence of clinical determinants of severe ROP in extremely preterm babies is multifaceted, involving many factors. In this study based on a cohort of babies less than 28 weeks GA, we attempted to provide an analysis of relationships between eleven postnatal predictors including OS statistics, and severe ROP.

The best subset of predictors

Single-, multi-predictor, and model averaging analyses were performed to identify predictors that can best describe the relationships with severe ROP. Some of the results from the three analyses were identical, whilst others differed in terms of the strength of association between predictors and outcome. We observed that these predictors could be categorised into three cases.

- Case I: dstOS, weight, GA, and daysVentilate emerged as the most highly

significant predictors in the single-predictor and model averaging analyses. Additionally, they appeared no less than eight times in the top ten models in the multi-predictor analysis.

- Case II: cvOS, sdOS, meanOS, wGA, and O2W36 were significantly associated to severe ROP in the single-predictor analysis. However, the individual effects of these variables were negated when adjusted for other variables. Gender was not significant in the single-predictor and model averaging analyses.
- Case III: Besides daysVentilate, daysO2 was another highly associated oxygen therapy variable in the single-predictor analysis. However, unlike daysVentilate, the strong effect of daysO2 did not persist in the multi-predictor analysis and was only marginally associated ($P=0.082$) with the adverse outcome of ROP in the model averaging.

Taking the results produced by these three analyses together, it appeared that dstOS, weight, GA, and daysVentilate were dominant risk factors for severe ROP. Therefore, these factors emerged as the best subset of predictors among the eleven initially considered. Note that this combination of four predictors appeared as the second top model in the multi-predictor analysis.

As we scrutinised closely the results from all three analyses, it can be seen that the results from the single- and multi-predictor analyses tended to be more liberal, whilst the model averaging approach was somewhat more conservative. We illustrate this in the following examples.

- In the single-predictor analysis, nearly all predictors were significantly associated with the outcome, with very small p -values. In contrast, many of them did not remain significant in the model averaging analysis.
- The fitted probabilities obtained from the top model (Figure 5.10) in the multi-predictor analysis and those from the model averaging analysis (Figure 5.12) were broadly similar, in that, discrimination between the two groups of babies with severe ROP and no severe ROP was very distinct. However, the former had more fitted probabilities that were closer to 1 for the severe ROP group.

We believe that the subtle differences in the three approaches indicate two extreme limits of performance; hence, the actual relationships could be somewhere between these two limits.

Oxygen fluctuations

Early studies on newborn rats exposed to oxygen-induced retinopathy showed that repeated hypoxic episodes (McColm et al., 2004; Coleman et al., 2008), alternating hypoxia and hyperoxia, small increments (Werdich et al., 2004), or wider range of oxygen fluctuations (Penn et al., 1995) caused severe retinal neovascularisation. In studies involving preterm babies, increased fluctuations of oxygen in the early weeks of life were reported as a critical factor for development of severe or threshold ROP (Saito et al., 1993; Cunningham et al., 1995; York et al., 2004). Except from York et al., (2004), information on the magnitude of the risk due to OS fluctuations is very limited as it has not been investigated in-depth and extensively. In our study of preterm babies at 29 weeks PMA, we examined the relation of variability and level of OS (using four OS statistics) on the development of severe ROP, and found that desaturation was the most significant OS predictor (model-averaged $P=0.007$) for severe ROP. This finding is in accord with De Fiore et al. (2010), who studied the association of severe ROP with multiple factors including desaturation episodes ($P<0.001$). Importantly, after adjustment with other clinical predictors, we noted that the relationship of desaturation with severe ROP risk was the strongest among all OS statistics (model-averaged OR = 2.800, 95% CI = (1.455, 5.399)). cvOS was another statistic that we had considered for OS variability; its contribution to the risk of developing severe ROP (model-averaged OR = 1.93, 95% CI = (0.94, 3.98)) was broadly similar to the study by York et al. (2004), which reported increased risk of severe ROP (OR = 1.69, 95% CI = (0.95, 2.99)) for preterm babies based on a 10% increment of CV of PaO₂ after receiving 30 days of oxygen therapy.

There have been previous studies that had investigated the impact of different OS levels on the morbidity of preterm babies including ROP. For instance, four studies (Tin et al., 2001; Anderson et al., 2004; Sun et al., 2002; Chow et al., 2003) have suggested that a OS target range of about 80-95% may reduce retinal surgery by

61-100%. Recent studies, however, showed that lower saturation targets reduce the incidence of severe ROP but may also increase risk of death (SUPPORT Group, 2010; Vaucher et al., 2012; Schmidt et al., 2013); BOOST II Group, 2013). The HOPE ROP study (McGregor et al., 2002) examined the progression rate to threshold (severe) ROP in two groups of preterm babies, one with median OS greater than 94%, and the other less than 94%, in room air when ROP was first diagnosed. They found that fewer babies progressed to threshold ROP in the higher OS group compared to the lower OS group (32% vs 46%, $P < 0.004$). Our data showed that with lower mean OS, there is a higher risk of severe ROP (model-averaged OR = 0.563, 95% CI = (0.287, 1.107)), which appeared to be consistent with the neonatal physiology at 29 weeks PMA as advised by a pediatrician. The basis for this phenomenon is not fully known but may be attributed to persistent hypoxia that occurs at the end of phase one ROP. This induces an overproduction of the VEGF growth factor and eventually leads to the second phase of retinal detachments (Smith, 2005).

Demographics

Lower GA and birth weight are the strongest and most consistently identified risk factors for ROP (McColm and Fleck, 2001; Good et al., 2005). We observed similar findings, wherein the risk increased for babies born earlier or with lighter weight. In the US, initial eye examinations have been recommended for babies born at less than 30 weeks GA, or with birth weight less than 1500 g, or between 1500 g to 2000 g in the case of high risk babies (American Association for Pediatric Ophthalmology, 2006). Our observations re-emphasised the importance of monitoring these babies with timely eye examinations and consistently scheduled ophthalmologic follow-ups.

Recent evidence has shown that the incidence of ROP cases has risen in the group of extremely premature babies in several developed countries, and that the increase was more notable in lower birth weight babies (Slidsborg et al., 2008). As GA and weight reflect different aspects of fetal wellbeing, i.e. maturity and growth, some studies on ROP started to use relative weight for GA (wGA). However, the adverse effects of low wGA on ROP among preterm babies are difficult to determine from published findings because these varied from one study to another. Some studies report that being born low wGA increases the risk of severe ROP (Qiu et al., 2012;

Slidsborg et al., 2008; Darlow et al., 2005), whilst others found no association between wGA and severe ROP (Austeng et al., 2010; Filho et al., 2009; Allegaert et al., 2009), which is in accordance to what we have obtained in our study. These conflicting results may be due to different ways of defining the low wGA group, such as (i) restriction on percentiles for GA (e.g. birth weight < 10th percentiles for GA or birth weight < 4th percentiles for GA); (ii) use of growth chart that differ according to countries (Australia, Canada, and Sweden); and (iii) use of single-centre versus multiple-centre data.

It has been suggested that the relationships between GA, weight and morbidity are less clear partly due to using cohorts of babies based on weight alone. In contrast, low wGA babies are shown to have higher rates of morbidity in cohorts selected on the basis of GA and weight (Qiu et al., 2012; Regev et al., 2003). This might explain why wGA is sometimes preferred over weight when assessing the impact of growth restriction on the development of ROP. Our data, however, showed that when weight and GA are considered individually, their relationship with the severity of ROP is stronger than when they are combined as wGA.

Gender has been increasingly recognised as a critical determinant in babies with ROP. Darlow et al. (2005) and Jacobson et al. (2009) have identified male gender to be significantly associated with an increased risk of ROP. In this study, it was not an important risk factor before and after adjustment for other factors. This was also observed in other comparable studies of ROP that involved a larger number of babies (Austeng et al., 2010; Gunn et al., 2012). More generally, whilst female babies were noted to have a survival advantage, male babies were found to be more susceptible to complications of prematurity and mortality (Synnes et al., 1994). The basis of the better outcomes for female babies is unclear, but may relate to the development of organ maturation during fetal growth (Ingemarsson, 2003).

Oxygen therapy

It has long been evident that ROP was observed more often among babies treated with extended duration of oxygen therapy during the perinatal period (Kinsey et al., 1956; Lala-Gitteau et al., 2007). This was further supported by Di Fiore et al. (2010), Cunningham et al. (1995), and our study, which found that babies who

developed severe ROP had received longer duration of ventilation and supplemental oxygen as compared to the no severe ROP group; these findings are summarised in Table 5.18.

Table 5.18: Findings on duration of oxygen therapy and ROP status

Oxygen therapy	Study	Severe ROP ^a	No severe ROP ^a	<i>p</i> -value ^b
Ventilation	Di Fiore et al., 2010	40 (18)	14 (15)	< 0.001
	Cunningham et al., 1995	50 (34)	29 (23)	< 0.01
	NZ-ROP	28 (24)	9 (13)	2.2E-11
Supplemental oxygen	Di Fiore et al., 2010	51(12)	34 (20)	< 0.001
	Cunningham et al., 1995	104(42)	70 (69)	< 0.001
	NZ-ROP	102(65)	52 (43)	1.7E-09

^aMean (SD)

^b*p*-values are from *t*-test for difference in means.

The impact of greater periods of assisted ventilation and supplemental oxygen on the health of young newborns is of great concern especially in extremely preterm babies as they are often given oxygen therapy due to their under-developed cardio-respiratory system. It is therefore relevant to ascertain which of the factors related to oxygen therapy has greater association with ROP. Our analysis of three factors - duration of ventilation, duration of supplemental oxygen, and use of supplemental oxygen until 36 weeks of PMA - disclosed that the first is the strongest predictor for severe ROP, rather than the other two. Recall that the risk of severe ROP increases with longer duration of ventilation (Table 5.15).

Apart from ROP, there have been numerous other reports on ventilation-related complications such as bronchopulmonary dysplasia, cancer, and cerebral palsy. Bronchopulmonary dysplasia that is in a form of inflammation of the lungs and airways is quite common in preterm neonates who are ventilated and exposed to oxygen (Askie et al., 2003). An increased risk of childhood leukemia has been reported for babies who were resuscitated with 100% oxygen at birth; the odds ratio was 2.6 when a facemask and bag were used, and increased to 3.5 for manual ventilation lasting 3 minutes or more (Naumburg et al., 2002). In a another study of

newborn babies exposed to prolonged sessions of oxygen therapy, the odds ratio of cerebral palsy was almost 3.0 in mechanically ventilated very low birth weight babies (Collins et al., 2001).

Limitations of study

There are several possible limitations to this study. First, the wGA categories were defined based on the Australian national growth charts (Roberts and Lancaster, 1999). The charts were chosen due to the absence of locally derived charts and close similarity in the demography and environmental factors of New Zealand and Australia. However, Indigenous babies (28230, 4%) were not included in the Australian charts as their correct GAs were difficult to be obtained (Roberts and Lancaster, 1999). This issue could cause a small number of babies in this study to be inappropriately classified for certain wGA categories. Second, inter and within centres variability in certain aspects of care and clinical monitoring of patients is inherent in this study. As trial protocols had been implemented in BOOST-NZ to ensure homogeneity among care providers when providing treatments to patients, the resulting variability is believed to be minimal and does not influence the conclusions of the study. The third and final limitation, the ROP incidence in this study pertains only to babies of birth weight < 1500g and < 28 weeks of GA. Previous studies have reported on severe ROP cases among low birth weight babies that were older than those in this analysis (Regev et al., 2003). There are possibilities that these older babies were inadvertently not screened for ROP screening, which could lead to under-estimation of ROP incidence for low birth weight babies.

Despite these limitations, this study was based on a comprehensive database collected prospectively and consistently throughout the study period. It clearly documents some important risk factors of severe ROP for extremely preterm babies born in several hospitals in New Zealand.

5.7 Summary

The OS data from an eight-week period (28 to 35 weeks PMA) were analysed to determine a suitable time period for computing OS statistics to be used in the logistic regression. We observed that (i) there were significant differences in the means of the OS statistics between the two groups of babies at nearly all weeks of

PMA, with the largest difference occurring at 28 or 29 weeks PMA; (ii) cvOS, sdOS, and dstOS were significantly higher in babies who develop severe ROP while meanOS was significantly lower, and (iii) the maximum value for variability and minimum value for level were attained at 29 weeks PMA. Taking into account the results from the 28 and 29 weeks PMA, and combined data, all findings for OS statistics of the NZ-ROP study were reported based on 29 weeks PMA.

Single-, multi-predictor, and model averaging analyses were performed to investigate the relationships between levels, variability of OS, and severe ROP; the relationships between multiple risk factors and severe ROP; and the respective degree of importance of these factors with the disease. We identified that there is a relationship between dstOS and severe ROP, and we confirmed weight, GA, and daysVentilate as significant risk factors for severe ROP. Of these variables, we found that (i) dstOS was the strongest predictor among the four OS statistics, (ii) weight and GA were more important individually than when combined as wGA, and (iii) daysVentilate was the most prominent predictor among the three oxygen therapy variables. Our analysis also identified four factors that formed a set of the most dominant factors of severe ROP in preterm babies: desaturation of OS levels, baby's age and weight at birth, and duration of exposure to oxygen therapy (ventilation support).

CHAPTER 6

NZ –LP STUDY: INVESTIGATION OF SOME STATISTICAL SUMMARIES OF OS DATA

6.1 Instability in neonatal oxygenation

Preterm infants display physiological instabilities such as low OS level and variation of heart rate and respiratory patterns.(Baird, 2004) These instabilities arise from a combination of immature organ development, neonatal illnesses (Hofstetter et al, 2007), clinical interventions (Murdoch and Darlow, 1984), and environmental factors (temperature, light and noise). This makes it difficult to distinguish instabilities that are merely due to prematurity from those partly caused by neonatal illnesses that affect cardiorespiratory functions. Determining the cause, so as to provide appropriate clinical care that avoids unnecessary and potentially harmful interventions often presents a challenge.

Various measures and techniques have been developed to detect instabilities in heart rate and respiratory (Seely et al., 2014). Fetal heart rate variability has been quantified to monitor fetal wellbeing in high-risk pregnancies and labour (Yum and Kim, 2003). Validated measures of prediction of neonatal illness have been developed based on heart rate characteristics (Malik et al., 1996; Griffin et al., 2007). Indices or scores have been calculated from respiratory support data to describe the severity of pulmonary status and respiratory failure in ventilated preterm babies (Madan et al., 2005; Subhedar et al., 2000).

However, a similar level of understanding has not been reached for OS. It is known that an inappropriate amount is detrimental for premature infants, with both hypoxaemia and hyperoxaemia capable of tissue injury in preterm babies (Silverman, 2004). But what the appropriate amount is and what features of OS are useful are still unclear (Cole et al, 2003). There is a considerable uncertainty regarding the OS threshold below which additional inspired oxygen should be given to babies (Cole et al, 2003).

In the absence of gold standard for code of practice for oxygen monitoring, many clinicians use clinical experience and limited information defined for term babies as surrogate values for preterm babies (Silvestri et al., 2002; Balfour-Lynn et al., 2009). Such information was often presented in the form of saturation level, i.e. mean, median, and median baseline (Hoppenbrouwers et al., 1992; Hunt et al., 1999), and the data were gathered under various settings using conventional devices. Hence, there were some variations in the definitions suggested for normal oxygenation for full term babies.

A convenient source of OS measurements is pulse oximeter, which is routinely used in many modern NICUs. A set of oximeter OS data from preterm babies admitted to an NICU was obtained. The data consisted of OS measurements recorded for ten babies at 34 weeks, nine babies at 35 weeks and 31 babies at 36 weeks PMA. All babies were healthy but two were considered to be mildly unstable and require greater attention. Due to the small sample size, a modest investigation was carried out at 36 weeks PMA to explore whether some common statistical summaries of the OS recordings are useful in the clinical care of late preterm babies. Three statistical measures: the mean, SD and CV, which is the ratio of the SD to the mean, were investigated. The potential usefulness of each measure was assessed by its ability to distinguish the two unstable babies from the others using box plots and a hierarchical clustering technique.

6.2 Studies on oxygenation in healthy preterm babies

One of the knowledge gaps still remains involving oxygenation in healthy preterm babies is the lack of a standard reference range of OS (Silvestri et al., 2002; Balfour-Lynn et al., 2009). These ranges, also referred to as reference ranges, reference intervals and normal values, are an important tool in medical area. They serve as a guideline to clinicians and medical practitioners in interpreting results obtained from the patients. In clinical term, a normal range consists of a range of values that typically represent the proportion of healthy individuals that take their values of a variable. Thus, values lying outside the range may suggest the presence of an illness or a disorder.

A clear definition of what constitutes normal oxygenation for preterm infants has been difficult to obtain. This could be partly due to the absence of clarity on how often and/or how long a baby could be allowed on low concentration of oxygen level, lack of both longitudinal data and knowledge on the dynamics of the preterm infants, and limited study on suitable measures that can adequately summarise the instabilities in OS. Previous studies on OS in term and preterm babies had used readings from the conventional oximeter, umbilical artery catheters and oxymonitor electrodes. These studies had focused on the saturation level, and the most common measure used was the median baseline. To compute this measure, the means of individual babies were first calculated using “clean” segments of data, and then the median of these means was obtained. A segment is classified as “clean” when the duration is at least ten seconds and at least 15 seconds away from any respiratory pause; the respiratory rate, amplitude and waveform are regular; and there is no occurrence of movement artefact (Hunt, et al., 1999).

6.2.1 Full term babies studies

Previous studies that contributed important information on the oxygenation in early life for healthy term babies have been extensive (Poets, 1998; Hunt, et al., 1999; Bakr and Habib, 2005; Samuel, et al., 2013; Rhein et al., 2012; Terrill et al., 2014). The studies involved between 34 and 4274 babies born more than 36 weeks GA. The investigators evaluated the impact of GA, activity state (sleeping, quiet, fussy, crying), altitude of hospitals, and type of monitoring device on OS at the time of measurement. They reported that the OS values increase to a small degree with increasing postnatal age, fussy/ crying newborns have lower OS values compared to quiet and sleeping newborns, babies born at high altitude have significantly lower OS values than those born at sea level, and the difference of OS values were about 2% between several oximeters that employ different algorithms. The OS values of the newborns were computed as mean, median or median baseline for several specific duration such as a 24- to 72-hour period from admission to discharge (O’Brien et al., 2000; Samuel et al., 2013), and the two to six months after birth (Terrill et al., 2014; Hunt et al., 1999; Poets et al., 1998; Stebbens et al., 1991; Poets et al., 1991). Median baseline of OS greater than 95% has been suggested as “normal” for healthy babies (Hunt et al., 1999; Balfour-Lynn et al., 2009), which indicates that the acceptable range of oxygenation in term healthy babies is relatively narrow.

6.2.2 Preterm babies studies

According to Rhein, et al. (2012) and our own findings, published data relating to oxygenation pattern and typical values in preterm babies have been limited. It was also noted that the number of babies in those studies was small, ranging from 28 to 55 babies (27 – 36 weeks GA). These babies were monitored under various conditions and settings: there were babies with periodic breathing (Razi et al., 2002) and lung complications (Poets, et al., 1992; Hofstetter et al., 2007), babies monitored within a laboratory setting (Hoppenbrouwers, et al., 1992), babies studied within their first week (Richard et al., 1993; Ng, et al., 1998) or first month of life (Harigopal et al., 2011; Rhein et al., 2012), and babies followed-up at 37 and 43 weeks PMA (Poets et al., 1992) and at three-monthly intervals in the first year of life (Beresford et al., 2005; Mok et al., 1988). These studies mostly used mean, median, median baseline, 5th and 95th percentiles (Rhein et al., 2012; Harigopal et al., 2011; Beresford et al., 2005; Ng et al., 1998) to investigate the oxygen values in healthy preterm babies. The median (or median baseline) of the OS levels for preterm babies in these studies were from 95% to 100%.

Late-preterm babies have received less attention than extremely preterm ones as they tend to be considered similar to term babies. However, there have been growing concerns that late-preterm babies are at an increased risk of morbidity, mortality and hospital readmission compared to term babies (Escobar et al., 2006; Engle et al., 2007). Guidelines for the evaluation and management of these babies have been recommended (Balfour-Lynn et al., 2009; Raju et al., 2006), addressing clinical care, knowledge gaps and pre-discharge of the babies. Two of the knowledge gaps that have been identified are the level of physiological maturity and tools to identify medical complications in late preterm babies. In the pre-discharge criteria (American Academy of Pediatrics, 2004), it was recommended that the “vital signs of the babies need to be within reference ranges and stable for the 12 hours preceding discharge, including a respiratory rate of less than 60 breaths per minute, a heart rate of 100 to 160 beats per minute, and axillary temperature of 36.5 to 37.4°C (97.7–99.3°F) measured in an open crib with appropriate clothing”. Note that OS levels have not been included in the criteria, reflecting the lack of information on this physiological variable. As preterm infants are often prone to life threatening events,

more oxygen studies are needed to aid clinicians in making health assessments or treatment decisions.

6.3 NZ-LP study

The typical measures to summarise OS values from the previous studies were based on the saturation level (mean, median percentiles, median of the means and median baseline). However, some clinicians also believe that the information about the health state of preterm babies is contained in the variability of physiological signals. Variability in oxygen has been suggested to be closely associated with ROP (Cole, 2010; Di Fiore et al., 2012). Studies have found that variability in oxygen in the first two weeks of life is a significant predictor of severe ROP (Cunningham, et al., 1995), and it can increase the severity of retinopathy in newborn rats (Penn, 1995). For very-low-birth-weight premature babies, increased in fluctuation of oxygen may impose a higher risk of developing threshold ROP (Saito, et al., 1993; York, et al., 2004). These studies show that variability derived from the OS measurements could provide some insight on the health conditions of babies.

We therefore aimed to explore the possible usefulness of several OS statistics in the clinical care of preterm babies. This investigation, New Zealand Late Preterm Babies (NZ-LP) study, was prompted by the availability of a set of oximeter OS data of late preterm babies obtained from an NICU in Christchurch, New Zealand. Since OS level has been of considerable interest (Poets, 1998; Hunt et al., 1999; Hoppenbrouwers et al., 1992; Beresford et al., 2005; Rhein et al., 2012; Terrill et al., 2014), one of the statistics investigated is the mean. Another feature of OS that has received some attention is variability; hence, the second statistic chosen is the SD. The third is the CV, which is the ratio of the SD to the mean; thus, providing a composite measure of variability relative to the mean level.

6.3.1 Patients

This study was an observational study that required no intervention. It was approved by the Upper South B Regional Ethics Committee in New Zealand. Babies were eligible for the study if they had been admitted to the neonatal intensive care unit at Christchurch Women's Hospital between November 2004 and January 2006 with gestation less than 36 weeks, had never received mechanical ventilation or had

required continuous positive airways pressure (CPAP) for less than twelve hours. Babies were also excluded if they had required treatment with opiates or sedative drugs, were born after maternal drug use in pregnancy, had known congenital malformations involving the cardio-respiratory system, or had conditions requiring major surgery. At the time of entry to the study, babies had to be well, not requiring supplemental oxygen with OS consistently $>92\%$ and not have been treated with methylxanthine therapy in the previous seven days. The study depended on the availability of a Masimo Radical pulse oximeter (Masimo Corporation, Irvine, CA, United States) and research nurse time; hence not all eligible babies were approached for consent.

After parental consent was obtained, an enrolled baby was connected to the pulse oximeter. A sensor was secured to the baby's finger or foot in a position that provided the most optimal signal. Data acquisition commenced once a good waveform was attained after the baby had been breathing spontaneously for at least 6 hours. OS was recorded once every two seconds for a continuous duration of at least 6 hours. Eventually, data were collected for two groups of babies: (1) Group P, comprising 10 babies, had OS measurements recorded once a week from 34 weeks PMA until discharge, there were 9 babies still in the study at 35 weeks, 6 at 36 weeks and 1 at 37 weeks.; and (2) Group N, consisting of 25 babies with recordings made only at 36 weeks PMA. For an investigation that was exclusively carried out for 36 weeks PMA, the subset of Group P's data taken at 36 weeks PMA was combined with the Group N's data to form a data set of 31 babies altogether).

Although all babies had to be clinically well to be eligible for study, two babies were identified as being clinically mildly unstable. Baby P17 was unstable at 34 - 36 weeks PMA. The day following enrolment at 34 weeks PMA, she had several episodes of apnea requiring stimulation and was placed on low-flow nasal oxygen, which was continued for 3 days. There was no evidence of sepsis or other problems. At 35 weeks PMA, her OS readings were less than 90% for 24% of the time, some of these episodes were associated with feeds. She was again placed on low-flow nasal oxygen, which was continued for 48 hours. At 36 weeks PMA, she had brief desaturations with feeds and the low-flow nasal oxygen was recommenced for a further 3 days. Baby N21 was unstable at 36 weeks PMA when saturations less than

90% were recorded for almost 30% of the monitoring period. Clinical review revealed no problems bar a soft systolic murmur and an echo-cardiogram was undertaken, which was normal but showed a small patent foramen ovale. The saturations were consistently 93% or greater over the next few days.

6.3.2 Pre-processing of OS data

The Masimo Radical has maximum sensitivity with OS values between 70-100%. Readings below 70% were less reliable as they could occur from movement or detachment of the sensor (Giuliano and Higgins, 2005). However, as a new generation oximeter the Masimo Radical reduced errors from these causes. Suspect readings that occurred from unclear causes were retained by introducing a weighting function to down-weight them. This weighted each measurement between 70-100% by 1, while measurements between 0-69% were linearly weighted between 0 and 1. All subsequent analyses were based on the weighted measurements. These down-weighted measurements constituted a very small part (about 0.3% were readings of 1-69% and 1.2% were readings of 0%) of the total measurements that were recorded. The formulation for the weighting function is described in the next passage.

Each measurement is assigned to a (normalised) weight according to three categories: the zeros, less reliable measurements (1-69%) and all measurements of 70-100%. For the first category, the weight for each measurement is assigned as zero. For the remaining two categories, assignment of weight involved three stages:

1. Sort all measurements 1-100% in ascending order, $x_{(1)}, \dots, x_{(m)}, x_{(m+1)}, \dots, x_{(n)}$ such that measurements $i = 1, \dots, m$ are the less reliable data and measurements $i = m+1, \dots, n$ are the measurements of 70-100%.
2. Assign weight to each of the sorted measurements. The less reliable data in some way reflect the unreliable level of the readings from the oximeter. Hence, the weight for the less reliable data is assigned as

$$w_{(i)} = x_{(i)}, \quad i = 1, \dots, m \quad (6.1)$$

while the weight for each measurement of 70-100% is assigned as

$$w_{(i)} = 100, \quad i = m + 1, \dots, n. \quad (6.2)$$

3. Redistribute the weight for all measurements in categories two and three so as to ensure that the weight of the measurements is proportionate within its own category. With normalisation of the weights, the weights of the zeros are redistributed to the non-zero measurements. The normalised weight for the less reliable data is given as

$$w_{(i)}^* = \left(\frac{w_{(i)}}{\sum_{j=1}^m w_{(j)}} \right) \frac{m}{n}, \quad (6.3)$$

and for measurements of 70-100% is

$$\begin{aligned} w_{(i)}^* &= \left(\frac{w_{(i)}}{\sum_{j=m+1}^n w_{(j)}} \right) \left(\frac{n-m}{n} \right) \\ &= \left(\frac{1}{n-m} \right) \left(\frac{n-m}{n} \right) \\ &= \frac{1}{n}. \end{aligned} \quad (6.4)$$

Note that by definition, $0 < w_{(i)}^* < 1$ and $\sum_{i=1}^n w_{(i)}^* = 1$.

The weighted measurements of OS for each baby were summarised for each available PMA using three statistical measures: the mean for measuring saturation level, SD for measuring saturation variability and CV (= SD/ mean) as a composite measure of variability relative to the mean level. The weighted mean and weighted SD, respectively, are given by

$$\bar{x}^* = \sum_{i=1}^n w_{(i)}^* x_i \quad (6.5)$$

and

$$s^* = \sqrt{\sum_{i=1}^n w_{(i)}^* x_{ii}^2 - (\bar{x}^*)^2}. \quad (6.6)$$

Using equations (4.25) and (4.26), the weighted CV was computed as

$$s^*/\bar{x}^* . \quad (6.7)$$

These statistics were then assessed whether they could be useful in distinguishing two clinically mildly unstable babies from a group of healthy babies.

6.4 Methodology

In this section, the approaches to identify unstable babies through the use of the three measures are outlined. First, a qualitative graphical assessment was made using the kernel density technique and box plots. The distribution of saturations at each PMA for stable and unstable babies were plotted and assessed. Subsequently, a quantitative appraisal was performed by using each measure in turn to group the babies into clusters, such that babies within a cluster looked more “similar” with respect to the measure than babies from different clusters. This was done using the hierarchical clustering (Sneath, 1973) technique together with a distance measure for determining the degree of similarity. The extent to which this is achieved was measured by an index of clustering performance. All computations were performed using MATLAB (v. 2010a) package.

6.4.1 Graphical assessment

Kernel density

Consider a random sample x_1, \dots, x_n for density f . The kernel density estimator for f is defined as

$$\hat{f}_K(x; h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \sum_{i=1}^n \phi(x; x_i, h^2), \quad (6.8)$$

where $K(x)$ is the kernel function and h is the bandwidth that controls the degree of smoothing applied to the data. The following conditions are required for the bandwidth and kernel function: $h > 0$, $K(x) \geq 0$ and $\int K(x)dx = 1$ (Wand and Jones, 1995).

The computational task of the kernel density estimate requires the specification for the bandwidth, kernel function and adjustment at the boundaries of the distribution. The direct plug-in method (Wand and Jones, 1995) was employed to

obtain a suitable bandwidth for the NZ-LP oxygen data. It has been reported as one of the most successful approaches among all bandwidth selection methods, both empirically and theoretically (Jones et al., 1996). The choice of kernel function has little impact on the performance of the kernel density estimator as the efficiency is close to unity for many distributions. The widely used Gaussian kernel function of zero mean and unit variance was chosen due to its simplicity. For the boundaries of the distribution, these were adjusted using the transformation method (Marron and Ruppert, 1994; Ruppert and Cline, 1994).

Box plots

The weighted CV, SD and mean were computed (using any available data sets of 34 weeks until 37 weeks PMA) for each baby, and the distributions of these measures were then assessed using box plots. In a box plot, the lower and upper edges of the box denote the locations of the first and third quartiles, respectively. The line within the box denotes the location of the median. The difference between the third and first quartiles, which corresponds to the length of the box, is the interquartile range (IQR). The lower tail of the box plot extends from the lower edge to the smallest point that is within 1.5 times the IQR below the lower edge. Likewise, the upper tail extends from the upper edge of the box to the largest point that is within 1.5 times the IQR above the upper edge. The common convention of labeling points that were more than 1.5 times the IQR beyond the first or third quartile as outliers was adopted. The actual number of times the IQR that an outlier fell beyond the first or third quartile was used to indicate the extreme level of that outlier, and will be referred to as the extremeness index.

6.4.2 Clustering analysis

At 36 weeks PMA, in which the sample size is larger (31 babies), a more detailed analysis for the usefulness of mean, SD and CV statistics was conducted using a hierarchical clustering (Sneath, 1973) method. The squared Euclidean distance was employed as the distance measure, where all values computed from this measure serve as inputs to the clustering algorithm. Clustering performance was measured using the Calinski-Harabasz (CH) index (Calinski and Harabasz, 1974), which has been in the ranking of top three indices among 30 indices that evaluate clustering results (Milligan and Cooper, 1985; Arbelaitz et al., 2013).

Hierarchical clustering

Hierarchical clustering is a method of grouping unlabelled objects into homogeneous groups such that the within-group similarity is maximised and the between-group similarity is minimised. Clustering of objects was performed through the use of pairwise distances of all observations via the agglomerative algorithm (Ward, 1963), the procedure is described below.

1. Compute $n(n-1)/2$ pairwise distances between n observations using the squared Euclidean distance. This is given as

$$d_E = (x_i - x_k)^2, i = 1, \dots, n-1 \text{ and } k = i+1, \dots, n. \quad (6.9)$$

2. Consider each observation as being apart and in its own cluster. Therefore, there are n clusters.

3. Compute between-cluster dissimilarity, i.e. compute the maximum distance between any two clusters. Here, all pairwise distances of observations between two clusters are considered.

4. Find the smallest between-cluster dissimilarity and merge the two clusters, there are now $n-1$ clusters.

5. Repeat step 3 to step 4 until only one cluster is left.

Dendrogram

The dendograms are used to illustrate the results from the hierarchical clustering using the mean, SD and CV. A dendrogram provides the overall insight of the cluster structure, i.e. informing the amalgamation process of clusters at each step. It shows possible clusters of objects labelled on the horizontal axis, with distances between clusters represented by the lengths of the vertical lines. For data set that consisted of 30 or fewer observations, each leaf in the dendrogram corresponds to one observation, whereas some leaves correspond to more than one observation for data set that consisted of more than 30 observations. This is due to the lower branches being collapsed to make the complete tree looking less crowded.

Measuring the clustering performance

The CH index was employed to assess the quality of the clustering outcomes. It quantifies the degree of separation among groups of data points, specifically, it measures the dispersion of the data points in a cluster and between clusters, respectively. A higher value of the index indicates that there is a better separation between clusters and greater homogeneity within each cluster, hence a stronger clustering result overall. The index is defined as

$$CH = \frac{B / (c-1)}{W / (n-c)}, \quad (6.10)$$

where n is the number of data points, c is the number of clusters, W is the within-cluster sum of squares, and B is the between-cluster sum of squares.

6.5 Results

Due to the small sample sizes at 34 and 35 weeks PMA, a limited longitudinal analysis was performed for 34 to 36 weeks PMA utilising box plots as well as summary statistics. Table 6.1 shows the basic demographic data of babies included in the study and Table 6.2 provides key summary statistics (mean, median, SD, and IQR) for all of the OS measurements taken at 34 to 36 weeks PMA of groups P and N. A pattern of increasing saturation level over time can be observed from the mean. There is also a pattern of decreasing variability with increasing PMA emerging from the SD. These trends are consistent with improving cardiorespiratory function as the babies matured. In these healthy babies, the improvement appears to be discernible at weekly intervals.

Table 6.1: Summary of GA, birth weight and gender of the babies included in study

	GA in weeks mean (range)	Birth weight in g mean (range)	Gender Male: female
Group P (<i>n</i> = 10)	31.5 (29-33)	1775 (1310 - 2630)	0:10
Group N (<i>n</i> = 25)	34.2 (33-35)	2240 (1415 - 2810)	12:13
Group P and N (pooled from 31 babies at 36 weeks PMA)	32.9 (29-35)	2008 (1310 - 2810)	12:19

Table 6.2: Summary of OS (%) data in healthy preterm babies at 34 to 36 weeks PMA

	34 weeks (<i>n</i> = 10)	35 weeks (<i>n</i> = 9)	36 weeks (<i>n</i> = 31)
Mean (SD)	95.4 (5.1)	96.6 (4.3)	96.9 (3.6)
Median (IQR*)	97 (94 - 98)	98 (96 - 99)	98 (96 - 99)
5 th , 95 th percentiles	81, 100	83, 100	90, 100

*IQR expressed as lower quartile to upper quartile

Figure 6.1 shows the distributions of OS for two babies – P02 and P17. OS was recorded from 34 weeks until 36 weeks PMA for P02 and until 37 weeks PMA for P17. P02 was clinically stable during this period but P17 was mildly unstable from 34 to 36 weeks and became stable only at 37 weeks. In both cases, the general trend of increasing OS over time is visible. More importantly, whilst P02 had consistently high OS at all three PMAs, the OS for P17 was persistently low at 34, 35 and 36 weeks and improved markedly only at 37 weeks. As can be seen from the plots, the densities for P17 became wider and the concentration gradually shifted toward higher percentages with increasing PMAs.

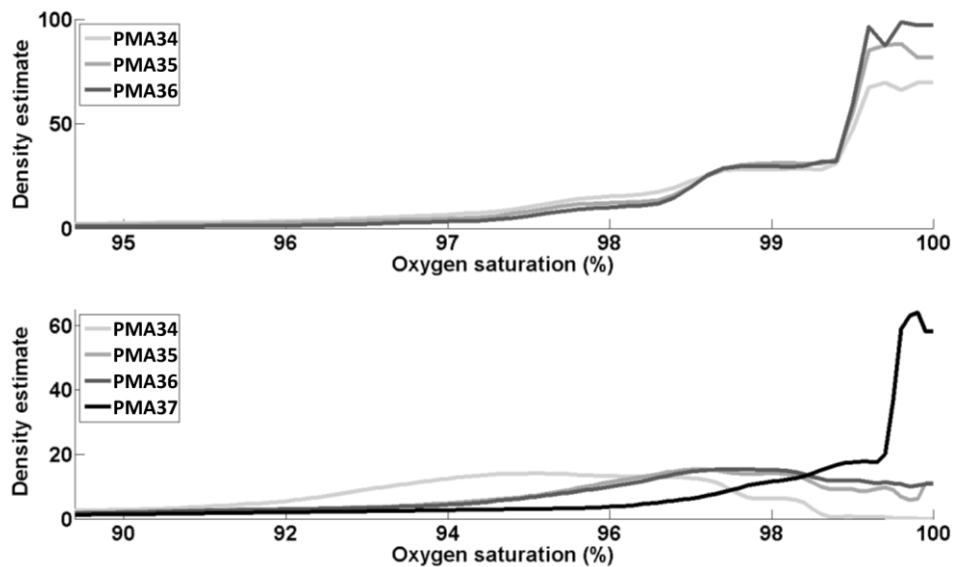


Figure 6.1: Kernel densities of OS for two babies, P02 (up) and P17 (down)

Table 6.3: CV, SD and mean of OS for individual babies at 34, 35 and 36 weeks PMA

34 weeks				35 weeks			
Case no.	CV	SD	mean	Case no.	CV	SD	mean
P02	0.036	3.5	98.0	P02	0.029	2.9	98.5
P03	0.024	2.3	97.3	P03	0.021	2.0	98.2
P06	0.053	5.1	96.0	P06	0.038	3.6	96.8
P10	0.014	1.4	97.9	P10	0.020	2.0	96.4
P11	0.014	1.4	97.2	P11	0.016	1.6	97.9
P14	0.052	4.8	92.3	P14	0.025	2.4	95.0
P17	0.062	5.8	92.4	P17	0.087	8.1	92.4
P18	0.020	1.9	97.7	P18	0.018	1.8	98.4
P19	0.091	8.1	89.0	P19	0.044	4.2	96.7
P20	0.027	2.5	95.3				

36 weeks							
Case no.	CV	SD	mean	Case no.	CV	SD	mean
P02	0.014	1.4	99.2	N24	0.022	2.1	98.2
P03	0.023	2.3	98.2	N26	0.033	3.1	96.0
P06	0.045	4.3	97.2	N28	0.027	2.5	94.2
P10	0.014	1.4	97.7	N36	0.044	4.2	96.5
P11	0.014	1.4	97.4	N37	0.037	3.6	95.8
P17	0.079	7.4	93.1	N52	0.024	2.3	97.6
N01	0.030	2.9	96.9	N56	0.010	1.0	97.2
N02	0.032	3.2	97.1	N57	0.020	2.0	96.8
N04	0.021	2.1	98.9	N58	0.020	1.9	96.4
N05	0.020	1.9	99.0	N66	0.016	1.5	97.4
N06	0.022	2.2	98.2	N76	0.015	1.5	98.6
N13	0.038	3.6	97.1	N77	0.017	1.6	97.4
N16	0.033	3.1	95.8	N79	0.018	1.7	97.4
N17	0.032	3.1	96.4	N81	0.024	2.4	97.5
N19	0.024	2.3	98.2	N82	0.016	1.5	96.9
N21	0.082	7.3	88.9				

6.5.1 Box plot

The mean, SD and CV of OS for individual babies at each PMA are given in Table 6.3. The empirical distributions of these three statistics are illustrated using box plots in Figure 6.2. Each box plot represents the empirical distribution of a statistic (i.e. mean, SD or CV) across individual babies at a given PMA. Once again, a trend of increasing level and decreasing variability over time can be seen (medians of the means are 96.6%, 96.8% and 97.2%, medians of the SDs are 3.0%, 2.4% and 2.0%, medians of the CVs are 0.032, 0.025 and 0.023 at the three PMAs). Due to the small sample sizes of 10 and 9, respectively, for 34 and 35 weeks PMA, neither the medians (tested using the Wilcoxon rank-sum test) nor the distributions (tested using the two-sample Kolmogorov-Smirnov test) of any of the statistics were statistically different at the 10% level of significance between pairs of PMAs (i.e. 34 versus 35 weeks, 34 versus 36 weeks and 35 versus 36 weeks). Note that this does not rule out differences actually exist but it is more likely that the tests were lacked of power due to the small sample sizes.

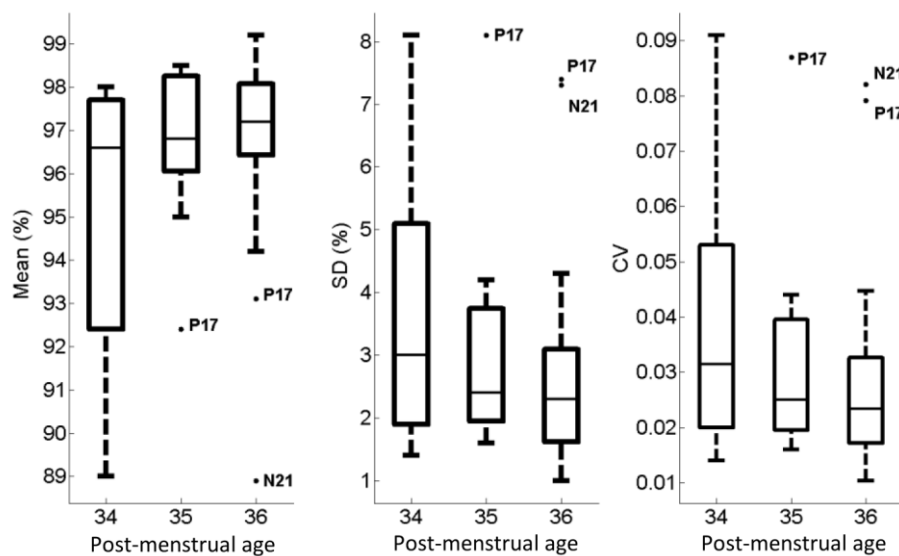


Figure 6.2: Box plots for mean (left), SD (middle) and CV (right) at 34, 35 and 36 weeks PMA

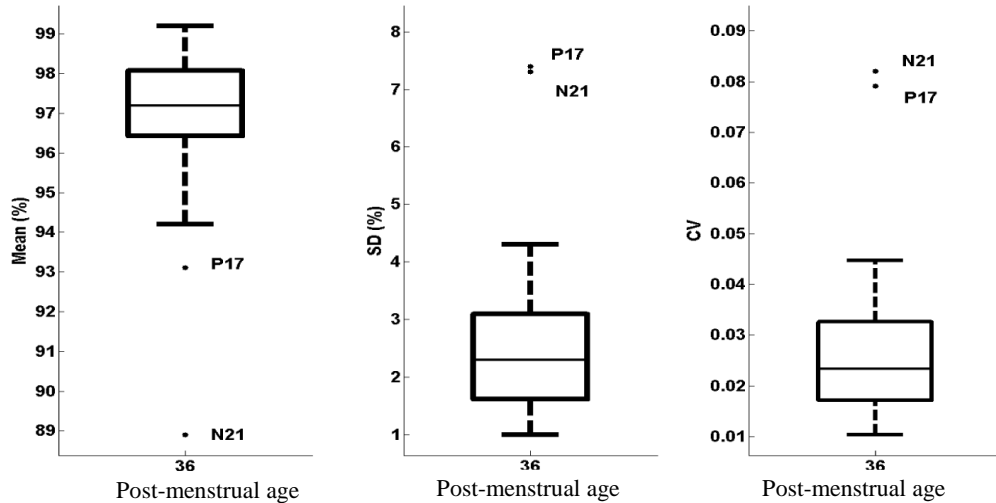


Figure 6.3: Box plots for the mean (left), SD (middle) and CV (right) at 36 weeks PMA

We now analyse the box plots for babies at 36 weeks PMA, where data from 31 babies are pooled and represented. Recall that points that are more than 1.5 times the IQR beyond the edges of the box are regarded as outliers. Thus, the box plot for each statistical measure represents the empirical distribution of that measure across the 31 babies. The line within the box plot for the mean denotes the location of the median of the means and is situated at 97.2%; likewise, the median of the SDs is at 2.0% and the median of the CVs is at 0.023. Each box plot flags babies P17 and N21 as outliers. None of the other babies is flagged as an outlier. The values of the extremeness index for babies P17 and N21 respectively are: 2.02 and 4.56 with the mean, 2.92 and 2.85 with the SD, and 2.98 and 3.18 with the CV. Recall that both babies were clinically assessed to be mildly unstable at 36 weeks PMA.

6.5.2 Clustering

Hierarchical clustering was performed by using each statistic to group babies at 36 weeks PMA into clusters. Clustering was performed on each of the three statistics (mean, SD, CV) in turn. The results from hierarchical clustering using the mean, SD and CV in turn are illustrated by dendrograms in Figure 6.4. To avoid cluttering the figure, babies P17 and N21 have been labelled explicitly but all of the other stable babies have simply been labelled as SB. Results of clustering outcomes and number of clusters depend on the decision of which horizontal lines is cut. The best choice is the one that maximises within-group similarity while minimising

between-group similarity at the same time. In the figure, the best choice in each dendrogram is denoted by a common colour for objects belonging to the same group. This is illustrated in the first dendrogram for clustering with the mean. The best outcome is shown as two groups, with N21 by itself in one group but with P17 together with all of the stable babies in the other group (coloured red). Clustering of the means was able to isolate N21 but P17 ended up in the stable group. Using the SDs or the CVs, P17 and N21 were placed in one group and the rest in another group. To determine whether clustering using CVs and SDs had any advantage over means, the CH index was computed. Successful clustering of the stable and unstable babies into two groups is achieved using the CV (CH = 72.8) and SD (CH = 63.3) but not with the mean. Even though the clustering outcomes for CV and SD are the same – P17 and N21 in one group and all other stable babies in another group – the higher CH index for CV indicates that its clusters are better separated and more homogeneous within. Thus, the combination of saturation variability and saturation level in the form of the CV appears to be more effective for detecting instability than either variability or level taken separately.

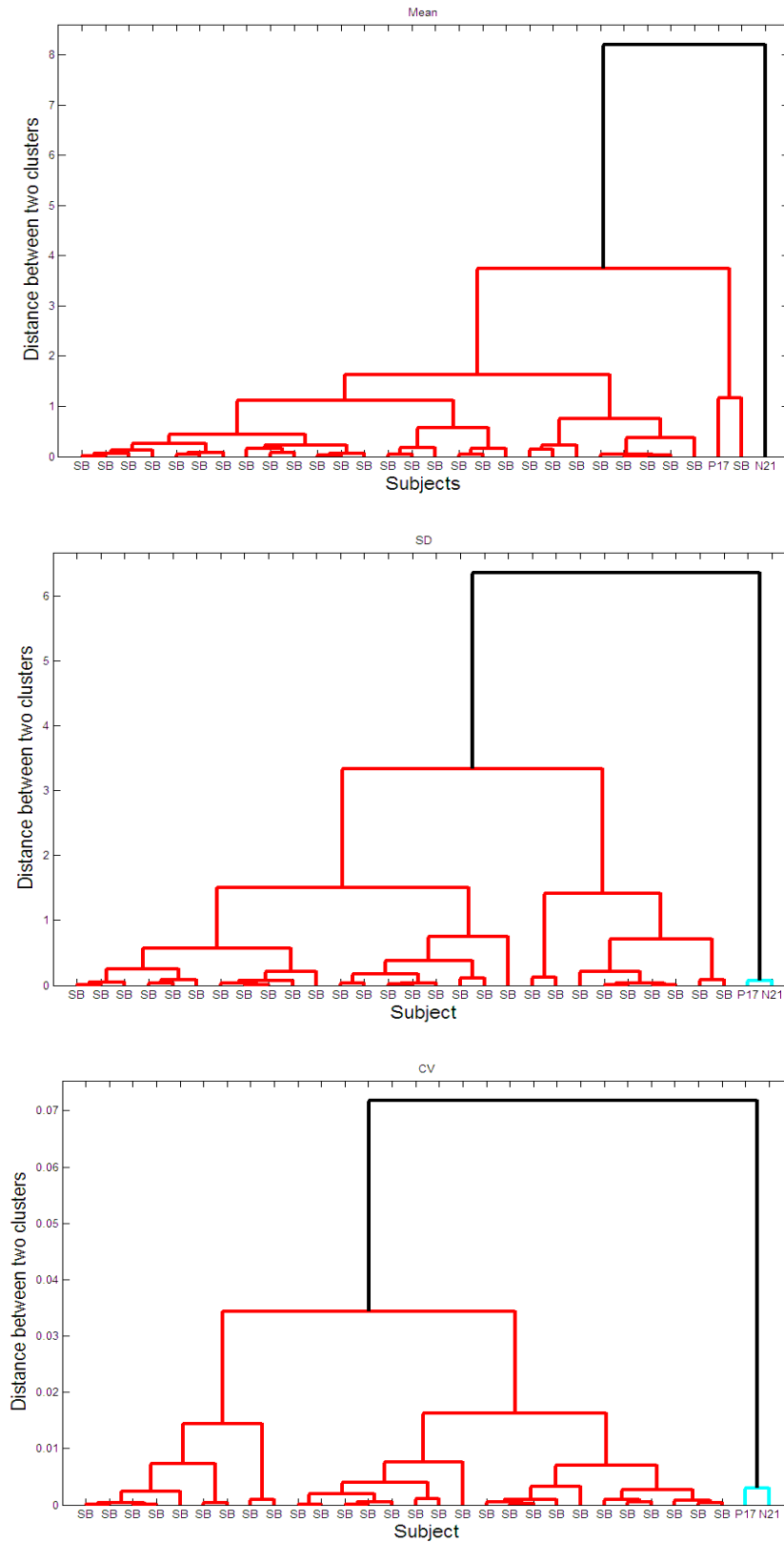


Figure 6.4: Dendrograms for hierarchical clustering with the mean (top), SD (middle) and CV (bottom)

6.6 Discussion

In this study, we observed a trend of increasing saturation level and decreasing variability with advancing PMA, which is consistent with improving maturity and stability over time. At 34, 35 and 36 weeks respectively, medians of the means were 96.6%, 96.8% and 97.2%; medians of the SDs were 3.0%, 2.4% and 2.0%; and medians of the CVs were 0.032, 0.025 and 0.023. With each of the three statistical measures at 36 weeks PMA, the box plot flagged the two unstable babies as outliers and none of the other babies. Successful clustering of the stable and two unstable babies into two groups was achieved using the CV (CH = 72.8) and SD (CH = 63.3) but not with the mean.

Limited information is available on OS values in preterm babies being monitored longitudinally, with some reports that mean of OS tends to become more stable with increasing PMA (Hofstetter et al., 2007; Beresford et al., 2005; Mok et al., 1988) Our study observed that for healthy late preterm babies, the increase in OS level and decrease in variability are discernible at weekly intervals, which accord to those prior studies. The observed changes in OS are consistent with physiological maturity over time and may also be associated with postnatal chemoreceptor activity (Rigatto et al., 1975).

The group of babies in previous studies differed from our own cohort, in that, they had examined babies at different age ranges and pooled data from different ages. Our pooled data from 31 babies for 36 weeks PMA, however, are still broadly similar in terms of OS level. For example, for babies pooled from 29 to 36 weeks GA (Harigopal et al., 2011; Beresford et al., 2005) and from 30 to 34 weeks GA (Ng et al., 1998), and studied between 2 - 14 days after birth, the medians of the means for OS were between 95% and 97%. From the box plot for the mean in Figure 6.3, the median of the means of the 31 babies at 36 weeks PMA is 97.2%. Other studies had included preterm babies with respiratory disorders (Hofstetter et al., 2007; DiPietro et al., 1994), babies monitored within a laboratory setting (Hoppenbrouwers et al., 1992), babies studied at 37 and 43 weeks PMA (Poets et al., 1992) and babies followed-up at three-month intervals over the first 12 months post-delivery period (Beresford et al., 2005). Median or median baseline OS levels ranging from 96.25% to 100% were obtained in these studies. The mean of our pooled data is 96.9% and the

median is 98%, these summary statistics are consistent with the literature. Studies involving full-term newborns found the median or median baseline of OS to be 98.3% (Samuel et al., 2013) in a 24-hour to 72-hour duration from admission to discharge, about 98% in the first month of age (Poets, 1998) and between 98% and 100% in the 2 to 6 months after birth (Terrill et al., 2014; Hunt et al., 1999; Poets, 1998).

Most previous studies on OS in preterm babies concentrated on saturation level (mean, median, median of the means or median baseline (DiPietro et al., 1994; Levesque et al., 2000; Poets, 1998; Hunt et al., 1999; Hoppenbrouwers et al., 1992; Poets et al., 1992; Beresford et al., 2005; Ng et al., 1998) for monitoring of babies in an NICU. Alternative measures such as SD and CV, however, were under-utilised to serve as measures of variability and stability. In our study, we investigated the usefulness of mean, SD, and CV statistics for distinguishing unstable preterm babies. Our findings suggest that variability in OS gives better result than saturation level, in that, the two clinically unstable babies were successfully identified by SD and CV whereas mean was able to detect only one baby. The relatively low extremeness index of 2.02 for P17 exacerbated by the large value of 4.56 for N21 helps explain the unsuccessful clustering with the mean; P17 looks more like it belongs in a group with the stable babies than with N21 in an unstable group. With the SD or CV, the values of the extremeness index for P17 and N21 are close together and relatively large (around 3); thus, aiding their clustering into a group apart from the other babies. The larger value of the CH index for the CV suggests that the resulting clusters are more pronounced than for the SD. This appears to be consistent with the larger values of the extremeness index for both P17 and N21 obtained from the CV box plot as compared to those from the SD. It therefore seems plausible that variability might be more effective than mean level for detecting instability in OS in preterm babies, and that the combination of variability and level through the CV might be even better.

Our results here add to the few known studies (DiPietro et al., 1994; DiFiore et al., 2012), suggesting that variability may reflect perinatal health conditions even though a baby may appear to be well. The box plots in Figure 6.3 show that the use of mean level or variability (SD) separately can lead to conflicting views about which of the two unstable babies is at greater risk. The mean appears to suggest that N21 is much more unstable than P17 while the SD seems to show that P17 is slightly more

unstable than N21. One way to resolve this is to use an appropriate composite measure combining mean level and variability, such as the CV.

The use of the CV for quantifying OS in preterm babies is uncommon but not new. It has been used by studies that assessed the effect of oxygen fluctuation on ROP (Saito et al., 1993; York et al., 2004). The investigators found that the CV of OS was higher for babies with threshold ROP (York et al., 2004) or with stage 3 or higher ROP (Saito et al., 1993) within the first two weeks after birth. They concluded that increased fluctuation of arterial oxygen tension might increase the risk of developing threshold ROP in very-low-birth-weight premature babies. In another study examining the impact of echocardiography on the cardio-respiratory stability of preterm babies, Groves and colleagues (Groves et al., 2005) reported that the CV of blood pressure, heart rate and OS, albeit not clinically significant, were statistically higher during the echocardiography period than during the rest period.

A limitation of this study that hampered stronger results was the small sample size of the data sets that included only two unstable babies. Nevertheless, the current results do help to draw attention to the use of both mean level and variability in the continuing quest to better understand OS in preterm babies. Further investigation is needed to develop tools that utilise these summary statistics for constructing the reference range of OS in preterm infants; this can be done in the context of studies with large number of patients.

6.7 Summary

The NZ-LP study was conducted to explore whether some statistical measures of OS recordings are useful for saturation monitoring of late preterm babies. This investigation involved ten babies at 34 weeks, nine babies at 35 weeks and 31 babies at 36 weeks PMA. All babies were healthy but two (identified as babies P17 and N21) were considered to be clinically unstable and required greater attention. OS readings were recorded continuously for at least 6 hours for each baby using a pulse oximeter. Each baby's recordings were summarised using three statistical measures, the mean for measuring saturation level, SD for measuring saturation variability and CV as a composite measure of variability relative to the mean level.

At 36 weeks PMA, in which the data set is larger, an extremeness index of box plots and a clustering performance index (CH index) of the hierarchical clustering were computed to assess the potential usefulness of each measure, i.e. by its ability to distinguish the two unstable babies from the others. The extremeness index measures the extreme level of an outlier that fell beyond the IQR while the CH index evaluates the quality of the clusters produced by the clustering method. A higher value of CH index indicates that there is a distinct separation between groups of babies, hence a better result overall.

With each of the statistics, both babies were flagged as outliers and none of the other babies. Nonetheless, the value of extremeness index with the mean was relatively low for baby P17, in comparison to the large value for baby N21. In contrast, the values of the extremeness index for the two babies, respectively, are relatively large and close together with the SD or CV. Thus, these helped in explaining the successful clustering of the babies into two groups (stable and unstable) using the CV and SD, with larger value of CH index was attained for the CV. This indicates that the grouping of the babies was markedly pronounced for the CV than that for the SD. These results were also supported by the larger values of extremeness index for CV. Taking the results together, this study show that the measures of SD and CV are believed to be more effective in detecting instability in OS in preterm babies than by using mean, but CV as the ratio of SD and mean could be even better than either SD or mean alone.

CHAPTER 7

CONCLUDING REMARKS

7.1 Overview and summary

In this thesis, the potential value of some summary statistics for OS features in pulse oximeter, particularly mean level and variation in OS readings, for saturation monitoring in preterm babies was investigated through two studies, NZ-ROP and NZ-LP. In the first study, four statistics were considered: mean (meanOS) for measuring saturation level, standard deviation (sdOS) for measuring saturation variability, coefficient of variation (cvOS) as a composite measure of variability relative to the mean level, and desaturation (dstOS) as another variability measure. With the exception of dstOS, similar OS statistics were used in the second study. A measure based on desaturation events was proposed for NZ-ROP study because generally, very young preterm babies experience a relatively higher percentage of OS readings below 80% compared to healthy term and late preterm babies (Rhein et al., 2012; McEvoy et al., 1993). The usefulness of each statistic was assessed by its ability to identify babies with unstable health conditions for NZ-LP while for NZ-ROP, statistics that could best describe the relationship between OS and severe ROP.

The first objective of NZ-ROP is to determine any plausible relationships between saturation level, three measures of variability of OS, and severe ROP. This was achieved through the fitting of logistic regression models, with the OS statistics as the predictors and the presence or absence of ROP as the response variable. Prior to the models fitting, some exploratory analysis was carried out and it was found that the four statistics were highly correlated. Since many clinicians believe that both variability and level of OS are important in describing the development of ROP, these predictors were separated into correlated and uncorrelated groups; thereby, each of the OS statistics could be assessed separately. Through single-, multi-predictor, and model averaging analyses, it was found that dstOS was strongly related to severe ROP. Importantly, after making adjustment for other predictors, dstOS was identified as having the greatest association with the risk of severe ROP among all OS statistics. These results shed additional information to the current knowledge that desaturation

episodes in extremely preterm babies could provide better insight on the adverse outcome of severe ROP than other OS statistics.

ROP is a potentially blinding disease that is related to multiple risk factors, including fluctuations in oxygenation, immaturity, growth and exposure to oxygen therapy (Cole, 2010; McColm & Fleck, 2001). Guided by the current evidence in the literature, this research also explored the respective degree of relationships of eleven predictors that are related to the four risk factors, with severe ROP. From the analysis, dstOS, GA, birth weight, and daysVentilate emerged as four of the most important predictors on the list. Subsequently, a separate analysis was conducted to identify predictor that has the strongest relationship with severe ROP for each group of the risk factors, and the following results were obtained. First, dstOS was found as the most significant OS predictor among the OS statistics, as noted previously. Second, the extent of immaturity and growth's influence on severe ROP was investigated by considering GA, birth weight and a composite variable birthweight-for-GA. Notably, weight and GA have stronger relationships with severity of ROP than when combined as birthweight-for-GA. Third, among three predictors of oxygen therapy, duration of ventilation support is the strongest predictor for severe ROP rather than duration of supplemental oxygen and use of supplemental oxygen until 36 weeks PMA. Taking these results together, the combination of desaturation of OS levels, baby's age and weight at birth, and duration of ventilation support was identified as a set of most dominant predictors for severe ROP. A multi-variable model consisted of these predictors yielded the lowest AIC among all models, and an AUC of 0.853 (95% CI=(0.792, 0.914)). These suggest that factors related to OS fluctuations, degree of immaturity, growth, and exposure to oxygen therapy are of great important in discriminating babies potentially at risk of developing severe ROP.

In NZ-LP, three OS statistics (mean, SD and CV) were investigated for their benefit in the clinical care of late preterm babies through the use of graphical and clustering approaches. OS readings of each baby were summarised using each statistic. Based on these statistics for 31 babies at 36 weeks PMA, an extremeness index that measures the extreme level of a particular outlying point, and a hierarchical clustering technique with a clustering performance index were employed to assess the ability of each measure to distinguish two clinically mildly unstable babies from a

group of healthy babies. The results from the two methods consistently showed that the two groups of babies were prominently separated with the SDs or CVs but not with the means. With SD and CV, both unstable babies were so very well distinct from the other babies that they were clustered in a same group and all stable babies in another group; the separation, however, was more prominent with the use of CV. Measures of variability such as SD and CV are believed to be more effective than mean for detecting late preterm babies with unstable health conditions, but the combination of variability and level through the CV could be even better.

7.2 Data management and adjustments

The contributions mentioned above would not be possible unless a sound data management and processing have been carried out through the implementation of several novel algorithms and procedures on OS data. Many recent studies on OS-related information had used the new generation oximeters, including the Masimo Radical that is noted for its superior performance. However, the maximum sensitivity of OS values for the Masimo oximeter is highly sensitive with OS values between 70-100% (Giuliano & Higgins, 2005). To account for acquisition of calibration errors, many studies on oxygen-related information studies had excluded artefactual readings, which were marked as 0% by the signal quality indicator, and erroneous data. For NZ-LP data, it was found that all readings between 0-69% that were suspected to be erroneous constituted only 1.5% from the total data. These less reliable readings were down-weighted linearly through the use of a weighting function to ensure that they will be given a lesser weight. The weighted readings were then incorporated into the NZ-LP analysis.

In managing and processing the NZ-ROP data, substantial efforts have been carried out to compile the NZ-ROP data sets that were obtained from the BOOST-NZ trial. The data consisted of the OS data (recorded from 24-hour until 36 weeks PMA or early discharge), clinical data and information on eye examinations for screening of ROP. Reports from 932 eyes examinations for 257 baby were compiled and the disease status was diagnosed by an ophthalmologist as severe ROP (stage three and above or requiring laser surgery) or no severe ROP. An algorithm was developed to remove erroneous observations and to assemble all data. Complete information was finally collected from 257 babies for the study.

The OS readings of Masimo oximeters used in the double-blind BOOST-NZ trial were influenced by trial-imposed offsets and oximeter artefact. The offsets occurred due to the oximeters that were specially programmed to study two groups of babies, of whom each was randomly assigned to one of two types of masked oximeter. There was a concern for the offsets and artefact as their presence would affect the choice of methods in the analyses. A procedure was outlined to identify the oximeter's type, and two algorithms involving linear and quadratic interpolations were developed and implemented on the New Zealand data to remove the effects of the offsets and artefact features. The algorithms were then validated using offsets and non-offset data sets of a UK preterm baby; the latter were considered as the reference data set for the analysis of the algorithms. The result was satisfactory: the oximeters for all babies were correctly identified, and the adjusted OS data were found to be very close to the non-offset data; hence suggesting that the features had been successfully adjusted to reflect the actual OS readings.

7.3 Further work

Managing neonatal oxygenation poses a great challenge to the healthcare providers since there are many issues on this complex topic that still require further understanding and questions that still need answers. Several critical areas have been identified for further research and these include the development of tools for measuring oxygen-related complications for newborns at various GAs, and investigating relations between varying oxygen levels over time, ROP, body systems and organs (Higgins et al., 2007).

7.3.1 *Curve-fitting methods*

In NZ-ROP study, the linear and quadratic interpolations were implemented for adjusting the offsets and oximeter artefact (described in Chapter 3), and loess and cubic spline smoothers (Hastie and Tibshirani, 1990) were applied for determining the preliminary functional form of continuous predictors in the logistic regression models (as outlined in Chapter 4). Other variants of spline functions (De Boor, 2001; Wahba, 1990) and including those within the Bayesian framework (Yue et al., 2014) could also be explored as regression splines and smoothing splines for the application of OS data.

Another curve-fitting method to be considered for the application of OS data is barycentric rational interpolation (Floater and Hormann, 2007). It is a variant of polynomial interpolation, with a set of weights and local polynomials specified for the barycentric function. The ‘roughness’ of the fitted curve depends on the value chosen for the polynomial order. Recently, Baker and Jackson (Baker and Jackson, 2014) demonstrated the application of this method to golf players data. More complicated approaches are available based on the Fourier series regression fitting for interpolating time series data (Brooks et al., 2012).

7.3.2 Measures of OS dynamics

This research had focused on time-domain measures for identifying useful measures of OS features for neonatal oxygenation monitoring. It was found that variability of OS was strongly related to severe ROP diagnosis. The variation in OS, however, may be contributed by linear and nonlinear signals, of which the latter may signify an impending or presence of a particular disease. To quantify this dynamical aspect of OS variability, sophisticated measures of nonlinear dynamical analyses such as power spectral (Scargle, 1982) and entropy measures (Richman and Moorman, 2000) will be required. These measures have been applied to quantify fetal heart rate features for predicting fetal distress (Cao et al., 2004), and to characterise patterns of hypoxemia events in preterm babies with and without severe ROP (Di Fiore et al., 2012).

7.3.3 Investigation on the behaviour of OS

The use of simple longitudinal analysis on eight-week OS data in this study (Chapter 5, Section 5.2) provided some early indication on the oxygenation patterns among babies with and without severe ROP. This approach, however, could be refined using nonlinear analyses, for instance fractal analysis, across the two groups of babies to uncover any patterns of subtle or temporal changes in oxygenation behavior of OS with advancing PMAs; thereby could confer some information on the long-range behavior between healthy and disease states (Shang et al., 2007). Analyses on the nonlinear behaviors in the OS time series data may provide a further direction on the studies of preterm babies.

7.3.4 Range of measure values

Since measures of variability of OS have been found useful, it would be helpful to clinicians if reference ranges of values are defined for each measure to monitor potential unstable babies and those at higher risk of severe ROP. Ranges (or control charts) developed for each advancing PMA would facilitate diagnosis and treatment decisions, thereby establishing the link with the state of health.

7.3.5 Time-dependent predictors

A statistical measure for desaturation episodes, which was included in the list of predictors for the logistic regression models developed in Chapter 5, was found to be strongly related to the development of ROP at 29 weeks PMA. However, given that the OS data were recorded continuously from 24 hours after birth until 36 weeks PMA, it may be possible to further exploit models with correlation structure involving longitudinal designs and time-varying OS measures. Development of such models could shed further light on the disease status over time in relation to OS fluctuations and other determinants. Using these models, it is also likely to investigate other neonatal complications associated with the use of oxygen on preterm babies, including chronic lung disease and potential damage to the postnatal growth of babies (Askie et al., 2003; Tin et al., 2001).

7.4 Concluding remarks

NZ-ROP is among the early studies that have investigated summary statistics of oximeter OS data, including desaturation episodes (Di Fiore et al., 2010; Di Fiore et al., 2012), as predictors for severe ROP. An investigation on the possible association between OS fluctuations at 29 weeks PMA and severe ROP revealed *dstOS* as the most significant OS statistic compared to *cvOS*, *sdOS*, and *meanOS* on severe ROP risk. The result therefore contributes to the ROP studies that desaturation episodes of OS along with variability of transcutaneous oxygen (Cunningham et al., 1995; York et al., 2004), are related to the risk of developing severe ROP. The present study demonstrated that greater occurrences of desaturation episodes exposed preterm babies to an increased risk of severe ROP. Particularly, babies with higher desaturation episodes were nearly three times more likely to develop severe ROP than babies with fewer episodes. Hence, these results advocate for care practices that could reduce fluctuations in oxygenation such as utilising automatic oxygen controllers

(Claire et al., 2009) or minimising frequent manual adjustments of supplemental oxygen.

It is noteworthy that the analyses of the NZ-ROP data conferred more information than other studies on the risk factors of ROP, in that, those studies have not investigated the relative contribution of each factor to the development of severe ROP. In this study for a cohort of babies of <29 weeks GA, the most significant predictor for each group of risk factors: exposure therapy, oxygenation fluctuations, immaturity and growth, was investigated. Longer periods of ventilation support, higher desaturation episodes of OS, and low GA and birth weight were found as the most important factors related to the development of severe ROP. These findings extend the results of Durand and colleagues (1992) who reported that very young babies with low birth weight, greater number of desaturation events and assisted ventilation, exhibit the early signs of chronic lung disease. Outcomes from NZ-ROP may have direct clinical implications in predicting potential babies at high risk for severe ROP.

Multiple clinical and oxygen-related predictors were considered in NZ-ROP to establish important predictors of severe ROP. The circumstances by which each one contributes to the findings of NZ-ROP are particularly relevant towards identifying and managing factors that exacerbate the development of severe ROP. Therefore, all available clinical predictors were considered in the model fitting stage. Because each model contained information of certain predictors, it was vital that all models are included in the analyses of predictors for ROP. The model averaging approach, which takes into account all models, i.e. all predictors, was able to fulfill this objective; thus, successfully enumerating all information needed.

The analysis of determining time interval for summarising OS data of NZ-ROP contributed in some way to the understanding of the links between onset of ROP and first examination after birth. The present study highlighted the pattern of OS during the early perinatal life of preterm babies, where it was found that a distinct separation on the variability and level of OS between babies with or without severe ROP could be observed as early as 28 weeks PMA. Whereas onset of ROP has been reported (Hellstrom et al., 2003; Smith, 2005) to occur from 30 weeks PMA; our

results here suggested the possibility of examining and hence detecting the retinal disorder as early as 28 weeks PMA in a premature child. Early identification of preterm babies at future risk of developing advanced stages of ROP would enable them to receive individual perinatal care; thus reducing the chance of disease progression and childhood blindness.

An increased rate of severe ROP to 13.6% from 6.9% in premature New Zealand babies was seen over the past 10 years (1998-2009). The observed increase, which is similar to those found in several developed countries, could be partly accounted for by the increasing survival rate of babies born at lower GA in recent years. This adds weight to the current trends of ROP that there is an increasing population of premature babies at greatest risk of visual impairment outcome; such an increment is likely to result in more babies being screened or treated with surgery. It is therefore important that clinical information on caring for these vulnerable patients is managed efficiently through the process of analysing, linking and delivering the information to either physical or “virtual” medical decision site. The analysis on the relationships between saturation level and variability of OS, other clinical and postnatal factors, and the risk of severe ROP exemplifies the start of this process.

Previous studies on OS in convalescing preterm babies have mostly reported the information in the form of saturation level such as mean, median and media baseline of OS. The observational NZ-LP study, however, found that the saturation variability, as measured by SD, could be more effective than saturation level for detecting instability in OS in late preterm babies; in that, variability may reflect perinatal unstable health conditions even though a baby may appear to be well; and that the combination of variability and level through the CV could be even better.

In brief, the findings from the analyses demonstrated that identification and summarisation of useful OS features quantitatively hold great promise for improving clinical monitoring and diagnosis of high-risk preterm babies. Although different cohorts of babies were examined in the two studies, i.e. extremely preterm babies in the NZ-ROP and late-preterm babies in the NZ-LP, the results found based on the two studies can be combined and become more informative for the clinicians in providing preterm babies with appropriate health assessments and timely interventions.

Appendix A

MATLAB programs

A.1 Algorithm for cleaning and assembling NZ-ROP data

This appendix gives the implementations of seven steps for preparing NZ-ROP data, which consist of OS, ROP status, clinical and postnatal data of 257 babies from 5 centres in New Zealand. These include files conversion (Excel, text and MATLAB), cleaning and aggregation of data.

Step 1: Creation of a folder for each baby and re-format of text files

```
% 5***_02mthyy_02mthyy.txt -> 5***_2-mm-yyyy_2-mm-yyyy.txt
```

```
% 5***_02mthyy_12mthyy.txt -> 5***_2-mm-yyyy_12-mm-yyyy.txt
```

```
% 5***_12mthyy_02mthyy.txt -> 5***_12-mm-yyyy_2-mm-yyyy.txt
```

```
% 5***_12mthyy_12mthyy.txt -> 5***_12-mm-yyyy_12-mm-yyyy.txt
```

```
function [numfolder,numfiles] = createfolder(dirfiles,dirtarget)
```

```
files = dir(dirfiles);
```

```
numfiles = length(files);
```

```
%cd(dirsource)
```

```
cd(dirfiles)
```

```
% for i = 1:numfiles
```

```
for i = 3:numfiles
```

```
    T = files(i).name;
```

```
    M = T(1:4);
```

```
    if isdir(strcat(dirtarget,M)) == 0
```

```
        mkdir(dirtarget,M)
```

```
    end
```

```
    if strcmp(M,T(1:4)) == 1
```

```
        copyfile(files(i).name,strcat(dirtarget,M))
```

```
    end
```

```
end
```

```
Folder = dir(strcat(dirtarget,'5*'));
```

```
numfolder = length(Folder);
```

```
function [numfolder,numfiles] = renamemth(dirfolder,dirsource,dirtarget)
```

```
Folder = dir(dirfolder);
```

```
numfolder = length(Folder);
```

```

numfiles = zeros(1,numfolder);
listmth = {'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec'};
listmt = {'01','02','03','04','05','06','07','08','09','10','11','12'};

for j = 1:numfolder
    cd(strcat(dirsource,Folder(j).name))
    files = dir('*.*txt');
    numfiles(j) = length(files);
    mkdir(dirtarget,Folder(j).name);

    for i = 1:numfiles(j)
        T = files(i).name;

        B = strcmp(listmth,T(8:10)); b = strcmp(listmth,T(16:18));
        C = listmt(find(B == 1)); c = listmt(find(b == 1));
        D = strcmp('0',T(6)); d = strcmp('0',T(14));
        if (find(B==1)) ~= 0
            if D && d
                T = (strcat(T(1:5),T(7),'-',C,'-20',T(11:13),T(15),'-',c,'-20',T(19:20),'.'txt'));
            elseif D && ~d
                T = (strcat(T(1:5),T(7),'-',C,'-20',T(11:15),'-',c,'-20',T(19:20),'.'txt'));
            elseif ~D && d
                T = (strcat(T(1:7),'-',C,'-20',T(11:13),T(15),'-',c,'-20',T(19:20),'.'txt'));
            elseif ~D && ~d
                T = (strcat(T(1:7),'-',C,'-20',T(11:15),'-',c,'-20',T(19:20),'.'txt'));
            else
                T = (strcat(T(1:5),'.'unknown'));
            end
        end
        movefile(files(i).name,strcat(dirtarget,Folder(j).name,'\,'char(T)))
    end
end
end

```

Step 2: Files conversion from text to MATLAB format for OS data

```
function [numfolder,numfiles] = convertdata(dirfolder,dirsource,dirtarget,centre)
```

```

%[numfolder,numfiles] = renamefile1(dirfolder2,dirtarget1,dirsource);
Folder = dir(dirfolder);
numfolder = length(Folder);
numfiles = zeros(1,numfolder);

for j = 1:numfolder
    cd(strcat(dirsource,Folder(j).name))
    files = dir('*.*txt');
    numfiles(j) = length(files);
    coxy = zeros(1,1); cflag = cell(1); cdate = cell(1); ctime = cell(1); campm = cell(1);

    for i = 1:numfiles(j)

```

```

T = files(i).name; % str contains filename.txt
fid = fopen(T,'r+');

C = textscan(fid, '%s %n %s','delimiter',' ');
RawOxy = C{2};
Flag = C{3};
DateTimeAmpm = C{1};
[Date,TimeAmpm] = strtok(DateTimeAmpm);
[Time,Ampm] = strtok(TimeAmpm);
fclose(fid);

coxy = [coxy;RawOxy]; cflag = [cflag;Flag]; cdate = [cdate;Date]; ctime =
[ctime;Time]; campm = [campm;Ampm];
end

n = length(coxy);
rawoxy = coxy(2:n); flag = cflag(2:n); date = cdate(2:n); time = ctime(2:n); ampm
= campm(2:n);
M = strcat(centre,'B',Folder(j).name);
cd(dirtarget)
save(M,'date','time','ampm','rawoxy','flag')
clear coxy cflag cdate ctime campm C datetimeampm timeampm ans

end

```

Step 3: Re-format of filenames for each baby in MATLAB directory

```
function [numfolder,numfiles] = renamefile(dirfolder,dirsourc,dirtarget)
```

```

Folder = dir(dirfolder);
numfolder = length(Folder);
numfiles = zeros(1,numfolder);

for j = 1:numfolder
    cd(strcat(dirsourc,Folder(j).name))
    files = dir('*.*txt');
    numfiles(j) = length(files);
    mkdir(dirtarget,Folder(j).name)
    for i = 1:numfiles(j)
        T = files(i).name;
        A = strcmp('-',T(7)); B = strcmp('-',T(8)); C = strcmp('-',T(17)); D = strcmp('-',T(18)); E = strcmp('-',T(19));

        if sum([A,C]) == 2 || sum([A,D]) == 2 || sum([B,D]) == 2 || sum([B,E]) == 2
            if A && C
                %T = (strcat(T(1:15),T(21:24),'-',T(18:19),'-',T(16),'.'txt'));
                T = (strcat(T(1:15),T(21:24),'-',T(18:19),'-',T(16),'.'txt'));
            elseif A && D
                T = (strcat(T(1:15),T(22:25),'-',T(19:20),'-',T(16:17),'.'txt'));
            elseif B && D
                %T = (strcat(T(1:16),T(22:25),'-',T(19:20),'-',T(17),'.'txt'));
            end
        end
    end
end

```



```

        T = (strcat(T(1:16),T(22:25),'-',T(19:20),'-',0,T(17),'.txt'));
    elseif B && E
        T = (strcat(T(1:16),T(23:26),'-',T(20:21),'-',T(17:18),'.txt'));
    else
        T = (strcat(T(1:5),'unknowntxt'));
    end
end
end
movefile(files(i).name,strcat(dirtarget,Folder(j).name,'\char(T)))
end

end

```

Step 4: Data cleaning for NZ-ROP data

```
function Ix = cleandata(rawoxy,flag,date,GAweeks,GAdays,DOB)
```

```

Ix = find(rawoxy >= 50); % Discard data below 50
X50 = rawoxy(Ix); flag = flag(Ix); date = date(Ix);
sig = {'normal'};
goodflag = strncmp(sig,flag,6);
Igoodflag = find(goodflag == 1);
X50 = X50(Igoodflag); date = date(Igoodflag);
date = char(date);
GA = 36 - GAweeks + 1;
meanGA = zeros(1,GA);
for i = 1:GA
    d = 6 - GAdays;

    if (DOB(2) == '/')
        Ia = find(date(:,2) == '/'); a = str2num(date(Ia));
        Ib = find(date(:,3) == '/'); b = str2num(date(Ib,1:2));
        startdate = str2num(DOB(1));
        enddate = startdate + d;
        Ic = find([a;b] <= enddate);

    else % DOB(3) == '/'
        startdate = str2num(DOB(1:2));
        enddate = startdate + d;
        if enddate >= 29
            e = DOB(4:5); % Detect the month of birth
            ee = strcat(DOB(4),num2str(str2num(DOB(5)) + 1)); % Detect the month of
enddate
            Ia = find(date(:,4) == DOB(4) & date(:,5) == DOB(5));

            if enddate >= 32
                if (e == '01') | (e == '03') | (e == '05') | (e == '07') | (e == '08') | (e == '10') | (e
== '12')
                    enddate = enddate - 31;
                end
            end
        end
    end
end

```

```

elseif enddate >= 31
    if (e == '04') | (e == '06') | (e == '09') | (e == '11')
        enddate = enddate - 30;
    end

elseif ((enddate >= 30) & (e == '02')) & (DOB(7:10) == '2008')
    enddate = enddate - 29;

elseif ((enddate >= 29) & (e == '02'))
    if (DOB(7:10) == '2006') | (DOB(7:10) == '2007') | (DOB(7:10) == '2009')
        enddate = enddate - 28;
    end
    % else % Those == 28,...,31 in Feb, 30-day mth and 31-day month
    %Ib = find((date(:,4) == DOB(4)) & (date(:,5)==DOB(5))); b =
str2num(date(Ib,1:2));
    %Ic = find(b <= enddate);
    end
    Ib = find(date(:,3)== ee(1) & date(:,4) == ee(2)) & str2num(date(:,1)) <=
enddate);
    Ic = [Ia;Ib];

else
    Ib = find(date(:,3) == '/'); b = str2num(date(Ib,1:2));
    Ic = find(b <= enddate); %%% CHECK
end
end
oxy = X50(Ic);
rfadj = NZraw4(oxy); % correction for dips and offsets

end

```

Step 5: Compilation of postnatal, clinical and oxygen data

```

% Convert DOB dates-format to numbers-format
% a) import ID (4 digits babies identifier) and DOB3(dates-clock format)
% from Excel to MATLAB
% b) DOB2 = datestr(datenum(DOB3,'dd/mm/yyyy'),1); output=dd-mmm-yyyy
% c) DOB = datenum(DOB,'dd-mmm-yyyy')
% d) create .mat that consists vectors of short-term outcomes (each row
% corresponds to each baby's information

```

```

function numfiles = appendData(dirfolder,dirsources)

```

```

Folder = dir(dirfolder);
numfiles = length(Folder);

```

```

for k = 1:numfiles
    cd(dirsources)
    f = dir('C*.mat');
    T = f(k).name;
    load(T,'date')
end

```

```

date2 = date;
date = datestr(datenum(date, 'dd/mm/yyyy'), 1); date = datenum(date);
load('STO_add_babies','ID','Baby','GAdays','WeightBirth','Gender',...
'HighestROP','DaysVentilate','DaysO2','O2atW36','DOB2','DOB')
t = T(1:8);
post = strmatch(t, Baby, 'exact');
DOB2 = DOB2(post,:);
DOB = DOB(post);
ID = ID(post);
HighestROP = HighestROP(post);
GAdays = GAdays(post);
Gender = Gender(post);
WeightBirth = WeightBirth(post);
DaysVentilate = DaysVentilate(post);
DaysO2 = DaysO2(post);
O2atW36 = O2atW36(post);
save(t,'date','date2','GAdays','DOB','DOB2','ID','HighestROP','Gender',...
'WeightBirth','DaysVentilate','DaysO2','O2atW36','-append')
end

```

Step 6: Computation of summary statistics of oxygen data (CV, Mean, SD)

```

% date = datestr(datenum(date, 'dd/mm/yyyy'), 1); date = datenum(date);
% Date and DOB have been converted to numbers using appendData.m
% X50 and date50 = final data with (i) values >= 50 and (ii) good signals ('normal')
% Extract cv and mean for the combined weeks
% treatB = high (1) or low (0) readings of oxygen saturation

```

```
function [cvB,meanB,F] = boostOS(dirsource,dirfolder,nweek)
```

```
Folder = dir(dirfolder);
numfiles = length(Folder);
```

```
cvB = zeros(numfiles,nweek+2); meanB = zeros(numfiles,nweek+2); F =
zeros(numfiles,1);
for k = 1:numfiles
```

```

    cd(dirsource)
    f = dir('C*.mat');
    T = f(k).name; F(k) = str2num(T(5:8));
    %load(T,'rawoxy','flag','date','DOB','HL')
    load(T,'date50','X50','DOB','HL')
    %Ix = find(rawoxy >= 50); % Discard data below 50
    %X50 = rawoxy(Ix); flag = flag(Ix); date = date(Ix);
    %sig = {'normal'};
    %goodflag = strncmp(sig,flag,6);
    %Igoodflag = find(goodflag == 1);
    %X50 = X50(Igoodflag); date50 = date(Igoodflag); % Maintain good data
('normal' signal flag)
    %xdate = unique(date50); n = length(xdate);
    %GA = 36 - GAweeks + 1;

```

```

meanGA = zeros(1,nweek); sdGA = zeros(1,nweek); cvGA = zeros(1,nweek);
startGA = zeros(1,nweek); endGA = zeros(1,nweek);
startGA(1) = DOB + 1; endGA(1) = DOB + 7;
startW12 = startGA(1); endW12 = DOB + 14; % W12 = combined data of the first
and second week after birth
startM = startGA(1); endM = DOB + 28; % M = combined data of the first until the
fourth week after birth

for i = 2:nweek
    startGA(i) = startGA(i-1) + 7; endGA(i) = endGA(i-1) + 7;
end
xIpol = 97;
for i = 1:nweek
    Ioxy = find(date50 >= startGA(i) & date50 <= endGA(i));
    oxy = X50(Ioxy);
    Mem = sum(ismember(xIpol,oxy));
    E = isempty(oxy);
    if ((E == 1) || (Mem == 0)) % empty vector for that week
        meanGA(i) = NaN; sdGA(i) = NaN; cvGA(i) = NaN; % missing oxy data for
that week
    else
        [sumf,sumfadjx,sumfadjx2] = NZraw4(oxy,HL); % correction for dips and
offsets
        meanGA(i) = sumfadjx/sumf; sdGA(i) = sqrt((sumfadjx2 -
meanGA(i)*sumfadjx)/(sumf - 1));
        cvGA(i) = sdGA(i)/meanGA(i);
    end
end
IW12 = find(date50 >= startW12 & date50 <= endW12); IM = find(date50 >=
startM & date50 <= endM);
W12 = X50(IW12); M = X50(IM);
Mem = sum(ismember(xIpol,W12));
E = isempty(W12);
if ((E == 1) || (Mem == 0))
    meanW12 = NaN; cvW12 = NaN;
else
    [sumf,sumfadjx,sumfadjx2] = NZraw4(W12,HL); % correction for dips and
offsets
    meanW12 = sumfadjx/sumf; sdW12 = sqrt((sumfadjx2 -
meanW12*sumfadjx)/(sumf - 1));
    cvW12 = sdW12/meanW12;
end
Mem = sum(ismember(xIpol,M));
E = isempty(M);
if ((E == 1) || (Mem == 0))
    meanM = NaN; cvM = NaN;
else
    [sumf,sumfadjx,sumfadjx2] = NZraw4(M,HL); % correction for dips and offsets
    meanM = sumfadjx/sumf; sdM = sqrt((sumfadjx2 - meanM*sumfadjx)/(sumf -
1));
1));

```

```

    cvM = sdM/meanM;
end

cvB(k,:) = [cvGA,cvW12,cvM]; CVx = cvB(k,:);
meanB(k,:) = [meanGA,meanW12,meanM]; MEANx = meanB(k,:);
save(T,'CVx','MEANx','-append')
end

```

Step 7: Computation of desaturation episodes (OS<=80% per week)

```
function [dstB,nB,F] = BOOSTdesat(dirsource,dirfolder,nweek)
```

```
Folder = dir(dirfolder);
numfiles = length(Folder);
```

```
dstB = zeros(numfiles,nweek+1); nB = dstB; F = zeros(numfiles,1);
for k = 1:numfiles
```

```

    cd(dirsource)
    f = dir('C*.mat');
    T = f(k).name;
    F(k) = str2num(T(5:8));
    load(T,'date50','X50','DOB') % Use good data ('normal' signal flag)

```

```

    dstGA = zeros(1,nweek); ndst = dstGA;
    startGA = zeros(1,nweek); endGA = zeros(1,nweek);
    startGA(1) = DOB + 1; endGA(1) = DOB + 7;
    startW14 = startGA(1); endW14 = DOB + 28; % M = combined data of the first
until the fourth week after birth

```

```

    for i = 2:nweek
        startGA(i) = startGA(i-1) + 7; endGA(i) = endGA(i-1) + 7;
    end

```

```

    for i = 1:nweek
        Ioxy = find(date50 >= startGA(i) & date50 <= endGA(i));
        oxy = X50(Ioxy); nGA = length(oxy);
        Idst = find(oxy <= 80);
        dst = oxy(Idst);
        E = isempty(oxy);
        if (E == 1)
            dstGA(i) = NaN; % oxy data are missing for that week
        else
            ndst(i) = length(dst);
            dstGA(i) = ndst(i)/nGA;
        end
    end

```

```

end
IW14 = find(date50 >= startW14 & date50 <= endW14);
W14 = X50(IW14); nW14 = length(W14);
Idst = find(W14 <= 80);

```

```
dst = W14(Idst); nd = length(dst);
E = isempty(W14);
if (E == 1)
    dstW14 = NaN;
else
    dstW14 = nd/nW14;
end
nB(k,:) = [ndst,nd];
dstB(k,:) = [dstGA,dstW14]; DST = dstB(k,:);

save(T,'DST','-append')

end
```

A.2 Algorithms for adjusting the offsets and artefact of

Masimo oximeters

This appendix gives the implementations of procedure for identifying types of oximeters (high or low saturation), and algorithms to remove the effects of the offsets and artefact from the OS data.

```
function [sumfreq,sumfadjx,sumfadjx2,freqadj,rfadj,rf] = NZraw4(rawoxy,highlow)
```

```
tabfreq = tabulate(rawoxy); ox = tabfreq(:,1);
sumfreq = sum(tabfreq(:,2));
rf = tabfreq(:,3); % All freqs (in %)
%maxrf = find(tabfreq(:,3)==max(tabfreq(:,3)));
%H = [rf(85) rf(86) rf(87)];
%L = [rf(93) rf(94) rf(95)];

%if std(H) < std(L)
% highlow = 1;
%else
% highlow = 0;
%end

v2 = ones(2,1); v4 = ones(4,1);
rfadj = rf;
idip = (87:90)';

if highlow == 1 % High reading @ low SAT
% Interpolation for r(84:85)
x = (84:85)'; x2 = x.^2;
B = [rfadj(83);rfadj(89);sum(rfadj(84:88))];
A = [83^2 83 1;86^2 86 1;sum(x2) sum(x) 2];
c = A \ B;
rfadj(x) = [x2 x v2] * c;

rfadj(86:92) = rfadj(89:95);
rftemp = rfadj(idip);

% Interpolation (linear) for r(87:90)
x = (87:90)';
A = [ 86 1; 91 1];
B = [ rfadj(86); rfadj(91)];
c = A \ B;
rfadj(x) = [x v4] * c;

rfadj(idip) = max([rfadj(idip) rftemp])';

% Interpolation for r(93:96)
x = (93:96)'; x2 = x.^2;
```

```

B = [rfadj(92);rfadj(97);rfadj(96)];
A = [92^2 92 1;97^2 97 1;sum(x2) sum(x) 4];
c = A \ B;
rfadj(x) = [x2 x v4] * c;

Ineg = length(find(rfadj(x) <= 0));
if Ineg > 0
    rfadj(x) = rfadj(96)/4;
end

% Low reading @ high SAT
elseif (highlow == 0)
    % Interpolation for r(95:96)
    x = (95:96)'; x2 = x.^2;
    B = [rfadj(91);rfadj(97);sum(rfadj(92:96))];
    A = [94^2 94 1;97^2 97 1;sum(x2) sum(x) 2];
    c = A \ B;
    rfadj(x) = [x2 x v2] * c;

    rfadj(88:94) = rfadj(85:91); % Shift to get r(88:94)

    % Interpolation for r(84:87)
    x = (84:87)'; x2 = x.^2;
    B = [rfadj(83);rfadj(88);rfadj(84)];
    A = [83^2 83 1;88^2 88 1;sum(x2) sum(x) 4];
    c = A \ B;
    rfadj(x) = [x2 x v4] * c;

    rftem = rfadj(idip);

    % Interpolation (linear) for r(87:90)
    x = (87:90)';
    A = [ 86 1;91 1];
    B = [ rfadj(86);rfadj(91)];
    c = A \ B;
    rfadj(x) = [x v4] * c;

    rfadj(idip) = max([rfadj(idip) rftem]');

else
    error('highlow is 1 or 0')

end

adj = (84:96)';
rfadj(adj) = rfadj(adj) * (sum(rf(adj))/sum(rfadj(adj))); % Normalisation of rel. freq
freqadj = rfadj * sumfreq/100;
sumfadjx = freqadj'* ox;
sumfadjx2 = freqadj'*(ox.^2);
%ymax = max(max(rf),max(rfadj)) + 0.5;

```



```

%plot((1:1:100),rf,-b',(1:1:100),rfadj,-
r.','LineWidth',2,'LineWidth',2,'MarkerSize',25);
%switch colr
% case '-r.'
% plot((1:1:100),rf,-b',(1:1:100),rfadj,-
r.','LineWidth',2,'LineWidth',2,'MarkerSize',25);
% case '-g.'
% plot((1:1:100),rfadj,-g.','LineWidth',2,'LineWidth',2,'MarkerSize',25);
%end
% xlabel('Saturation','fontsize',16); ylabel('Relative frequency (%)','fontsize',16)
% axis([60 100 0 ymax])
% title(['Saturation readings ' ' inputname(1)],'fontsize',16);
% h = legend('raw',adjusted',2);
% set (h,'Location','NorthWest','fontsize',13);
% grid on

%x1=CHC5150(1:22000); x2=CHC5150(22001:44000);
x3=CHC5150(44001:65536);
%plot((1:1:100),rfadjSEG,'--g','LineWidth',2,'LineWidth',2,'MarkerSize',25);

```

References

- Allegaert, K., Vanhole, C., Casteels, I., Naulaers, G., Debeer, A., Cossey, V., & Devlieger, H. (2003). Perinatal growth characteristics and associated risk of developing threshold retinopathy of prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 7, 34-37.
- American Academy of Pediatrics. Committee on fetus and newborn. (2004). Hospital stay for healthy term newborns. *Pediatrics*, 113, 1434-1436.
- American Academy of Pediatrics & American College of Obstetricians and Gynecologists. (2002). *Guidelines for perinatal care*.
- Section on Ophthalmology American Academy of Pediatrics; American Academy of Ophthalmology; American Association for Pediatric Ophthalmology and Strabismus. (2006). Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*, 117, 572-576.
- Anderson, C. G., Benitz, W. E., & Madan, A. (2004). Retinopathy of prematurity and pulse oximetry: a national survey of recent practices. *Journal of Perinatology*, 24, 164-168.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46, 243-256.
- Ashton, N., & Cook, C. (1954). Direct observation of the effect of oxygen on developing vessels. *Brit. J. Ophthal*, 38, 433.
- Askie, L. M., Henderson-Smart, D. J., Irwig, L., & Simpson, J. M. (2003). Oxygen-saturation targets and outcomes in extremely preterm infants. *New England Journal of Medicine*, 349, 959-967.
- Askie, L. M., Henderson-Smart, D. J., & Ko, H. (2009). Restricted versus liberal oxygen exposure for preventing morbidity and mortality in preterm or low birth weight infants. *Cochrane Database Syst Rev*, 1.
- Askie, L. M. (2013). Optimal oxygen saturations in preterm infants: a moving target. *Current Opinion in Pediatrics*, 25, 188-192.
- Askie, L. M., Brocklehurst, P., Darlow, B. A., Finer, N., Schmidt, B., & Tarnow-Mordi, W. (2011). NeOProm: neonatal oxygenation prospective meta-analysis collaboration study protocol. *BMC Pediatrics*, 11, 6.
- Austeng, D., Blennow, M., Ewald, U., Fellman, V., Fritz, T., Hellstrom-Westas, L., & Westgren, M. (2010). Incidence of and risk factors for neonatal morbidity after active perinatal care: extremely preterm infants study in Sweden (EXPRESS). *Acta Paediatrica*, 99, 978-992.
- Baker, R. D., & Jackson, D. (2014). Statistical application of barycentric rational

interpolants: an alternative to splines. *Computational Statistics*, 1-17.

Bakr, A. F., & Habib, H. S. (2005). Normal values of pulse oximetry in newborns at high altitude. *Journal of Tropical Pediatrics*, 51, 170-173.

Balfour-Lynn, I. M., Field, D. J., Gringras, P., Hicks, B., Jardine, E., Jones, R. C. & Wallis, C. (2009). BTS guidelines for home oxygen in children. *Thorax*, 64, ii1-ii26.

Bharwani, S. K., & Dhanireddy, R. (2008). Systemic fungal infection is associated with the development of retinopathy of prematurity in very low birth weight infants: a meta-review. *Journal of Perinatology*, 28, 61-66.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons.

Beresford, M. W., Parry, H., & Shaw, N. J. (2004). Twelve-month prospective study of oxygen saturation measurements among term and preterm infants. *Journal of Perinatology*, 25, 30-32.

Bianco, A. M., & Yohai, V. J. (1996). *Robust estimation in the logistic regression model*. Springer, New York.

Blencowe, H., Lawn, J. E., Vazquez, T., Fielder, A., & Gilbert, C. (2013). Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. *Pediatric Research*, 74, 35-49.

Bohnhorst, B., Peter, C. S., & Poets, C. F. (2000). Pulse oximeters' reliability in detecting hypoxemia and bradycardia: Comparison between a conventional and two new generation oximeters. *Critical Care Medicine*, 28, 1565-1568.

Bohnhorst, B., Peter, C. S., & Poets, C. F. (2002). Detection of hyperoxaemia in neonates: data from three new pulse oximeters. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 87, F217-F219.

Bolton, D. P. G., & Cross, K. W. (1974). Further observations on cost of preventing retrolental fibroplasia. *The Lancet*, 303, 445-448.

BOOST-II, Stenson, B. J., Tarnow-Mordi, W. O., Darlow, B. A. & Armstrong, N. (2013). Oxygen saturation and outcomes in preterm infants. *New England Journal of Medicine*, 368, 2094-2104.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30, 1145-1159.

Brooks, E. B., Thomas, V. A., Wynne, R. H., & Coulston, J. W. (2012). Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis. *Geoscience and Remote Sensing*, 50, 3340-3353.

- Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, 3, 1-27.
- Cantoni, E., & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96, 1022-1030.
- Cao, H., Lake, D. E., Ferguson, J. E., Chisholm, C. A., Griffin, M. P., & Moorman, J. R. (2006). Toward quantitative fetal heart rate monitoring. *Biomedical Engineering, IEEE Transactions*, 53, 111-118.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 158, 419-466.
- Chen, J., & Smith, L. E. (2007). Retinopathy of prematurity. *Angiogenesis*, 10, 133-140.
- Chow L. C, Wright K. W., & Sola A. (2003). Can changes in clinical practice decrease the incidence of severe retinopathy of prematurity in very low birth weight infants? *Pediatrics*, 111, 339-345.
- Claure, N., D'Ugard, C., & Bancalari, E. (2009). Automated adjustment of inspired oxygen in preterm infants with frequent fluctuations in oxygenation: a pilot clinical trial. *The Journal of Pediatrics*, 155, 640-645.
- Coats, D. K., Paysse, E. A., & Steinkuller, P. G. (2000). Threshold retinopathy of prematurity in neonates less than 25 weeks' estimated gestational age. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 4, 183-185.
- Cole, C. H., Wright, K. W., Tarnow-Mordi, W., & Phelps, D. L. (2003). Resolving our uncertainty about oxygen therapy. *Pediatrics*, 112, 1415-1419.
- Cole, C. H. (2010). Making sense of clinical determinants of retinopathy of prematurity. *The Journal of Pediatrics*, 157, 5-7.
- Coleman, R. J., Beharry, K. D., Brock, R. S., Abad-Santos, P., Abad-Santos, M., & Modanlou, H. D. (2008). Effects of brief, clustered versus dispersed hypoxic episodes on systemic and ocular growth factors in a rat model of oxygen-induced retinopathy. *Pediatric Research*, 64, 50-55.
- Collins, M. P., Lorenz, J. M., Jetton, J. R., & Paneth, N. (2001). Hypocapnia and other ventilation-related risk factors for cerebral palsy in low birth weight infants. *Pediatric Research*, 50, 712-719.
- Console, V., Gagliardi, L., De Giorgi, A., & De Ponti, E. (1996). Retinopathy of prematurity and antenatal corticosteroids. The Italian ROP Study Group. *Acta bio-medica de L'Ateneo parmense: organo della Societa di medicina e scienze naturali di Parma*, 68, 75-79.

- Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. Monographs on Statistics and Applied Probability. Chapman and Hall, New York.
- Cross, K. W. (1973). Cost of preventing retrolental fibroplasia? *The Lancet*, 302, 954-956.
- Croux, C., & Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics & Data Analysis*, 44, 273-295.
- Cunningham, S., McIntosh, N., Fleck, B. W., & Elton, R. A. (1995). Transcutaneous oxygen levels in retinopathy of prematurity. *The Lancet*, 346, 1464-1465.
- Cust, A. E., Darlow, B. A., & Donoghue, D. A. (2003). Outcomes for high risk New Zealand newborn infants in 1998–1999: a population based, national study. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 88, F15-F22.
- Darlow, B. A., Clemett, R. S., Horwood, L. J., & Mogridge, N. (1997). Prospective study of New Zealand infants with birth weight less than 1500 g and screened for retinopathy of prematurity: visual outcome at age 7–8 years. *British Journal of Ophthalmology*, 81, 935-940.
- Darlow B. A., Hutchinson J. L., Simpson J. M., Donoghue D. A., Henderson-Smart D. J. & Evans N. J. On behalf of the Australian and New Zealand Neonatal Network. (2004). Size matters: the risk of chronic lung disease and significant retinopathy of prematurity. *Pediatric Research*, 56, 474.
- Darlow, B. A., Hutchinson, J. L., Henderson-Smart, D. J., Donoghue, D. A., Simpson, J. M., & Evans, N. J. (2005). Prenatal risk factors for severe retinopathy of prematurity among very preterm infants of the Australian and New Zealand Neonatal Network. *Pediatrics*, 115, 990-996.
- Darlow, B., Mogridge, N., Horwood, J., Wynn-Williams, E. and Austin, N. (2007). Follow-up at 2 years of age of all admissions to a regional NICU compared with controls. *E-PAS*, 61, 7925.7
- Darlow, B. A., Ells, A. L., Gilbert, C. E., Gole, G. A., & Quinn, G. E. (2013). Are we there yet? Bevacizumab therapy for retinopathy of prematurity. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 98, F170-F174.
- Darlow, B. A., Marschner, S. L., Donoghoe, M., Battin, M. R., Broadbent, R. S., Elder, M. J., & et al. (2014). Randomized controlled trial of oxygen saturation targets in very preterm infants: two year outcomes. *The Journal of Pediatrics*, 165, 30-35.
- De Boor, C. (2001). *A practical guide to splines*. Applied Mathematical Sciences. Springer, New York.

- Di Fiore J, Bloom JN, Orge F, Schutt A, Mark S, Cheruvu VK, Walsh M, Finer N, Martin RJ. (2010). A higher incidence of intermittent hypoxemic episodes is associated with severe retinopathy of prematurity. *J Pediatrics*, *157*, 69-73.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, *44*, 837-845.
- Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*, 265-282.
- Di Fiore, J. M., Kaffashi, F., Loparo, K., Sattar, A., Schluchter, M., Foglyano, R., & Wilson, C. G. (2012). The relationship between patterns of intermittent hypoxia and retinopathy of prematurity in preterm infants. *Pediatric Research*, *72*, 606-612.
- Dipietro, J. A., Caughy, M., Cusson, R., & Fox, N. A. (1994). Cardiorespiratory functioning of preterm infants: Stability and risk associations for measures of heart rate variability and oxygen saturation. *Developmental Psychobiology*, *27*, 137-152.
- Durand, M., McEvoy, C., & MacDonald, K. (1992). Spontaneous desaturations in intubated very low birth weight infants with acute and chronic lung disease. *Pediatric Pulmonology*, *13*, 136-142.
- Engle, W., Tomashek, K.M. and Wallman, C., Committee on Fetus and Newborn, American Academy of Pediatrics. (2007). "Late-preterm" infants: a population at risk. *Pediatrics*, *120*, 1390-1401.
- Chavez-Valdez, R., McGowan, J., Cannon, E., & Lehmann, C. U. (2011). Contribution of early glycemic status in the development of severe retinopathy of prematurity in a cohort of ELBW infants. *Journal of Perinatology*, *31*, 749-756.
- Escobar, G. J., Clark, R. H., & Greene, J. D. (2006). Short-term outcomes of infants born at 35 and 36 weeks gestation: we need to ask more questions. *Seminars in Perinatology*, *30*, 28-33.
- Faraway, J. J. (2006). *Extending the linear model with R*. Chapman and Hall, Florida.
- Fierson, W. M., Saunders, R. A., Good, W., Palmer, E. A., Phelps, D., Reynolds, J. & Siatkowski, R. M. (2013). Screening Examination of Premature Infants for Retinopathy of Prematurity. *Pediatrics*, *131*, 189-195.
- Fortes Filho, J. B., Valiatti, F. B., Eckert, G. U., Costa, M. C. D., Silveira, R. C., & Procianny, R. S. (2009). Is being small for gestational age a risk factor for retinopathy of prematurity? A study with 345 very low birth weight preterm infants. *Jornal de Pediatria*, *85*, 48-54.

- Floater M.S., Hormann K. (2007). Barycentric rational interpolation with no poles and high rates of approximation. *Numer Math*, 107, 315-331.
- Flynn, J. T., Bancalari, E., Bawol, R., Goldberg, R., Cassady, J., Schiffman, J., & Sim, E. (1987). Retinopathy of prematurity. A randomized, prospective trial of transcutaneous oxygen monitoring. *Ophthalmology*, 94, 630-638.
- Flynn, J. T., Bancalari, E., Snyder, E. S., Goldberg, R. N., Feuer, W., Cassady, J., & Gillings, D. (1992). A cohort study of transcutaneous oxygen tension and the incidence and severity of retinopathy of prematurity. *New England Journal of Medicine*, 326, 1050-1054.
- Gallo, J. E., Jacobson, L., & Broberger, U. (1993). Perinatal factors associated with retinopathy of prematurity. *Acta Paediatrica*, 82, 829-834.
- Gaultier, C. (1995). Cardiorespiratory adaptation during sleep in infants and children. *Pediatric Pulmonology*, 19, 105-117.
- Gilbert C, Fielder A, Gordillo L, et al. (2005). Characteristics of infants with severe retinopathy of prematurity in countries with low, moderate, and high levels of development: implications for screening programs. *Pediatrics*, 115, e518–525.
- Giudice, L. C., de Zegher, F., Gargosky, S. E., Dsupin, B. A., De Las Fuentes, L., Crystal, R. A., & Rosenfeld, R. G. (1995). Insulin-like growth factors and their binding proteins in the term and preterm human fetus and neonate with normal and extremes of intrauterine growth. *The Journal of Clinical Endocrinology & Metabolism*, 80, 1548-1555.
- Giuliano, K. K., & Higgins, T. L. (2005). New-generation pulse oximetry in the care of critically ill patients. *American Journal of Critical Care*, 14, 26-37.
- Good, W. V., Hardy, R. J., Dobson, V., Palmer, E. A., Phelps, D. L., Quintos, M., & Tung, B. (2005). The incidence and course of retinopathy of prematurity: findings from the Early Treatment for Retinopathy of Prematurity Study. *Pediatrics*, 116, 15-23.
- Griffin, M. P., Lake, D. E., O'Shea, T. M., & Moorman, J. R. (2007). Heart rate characteristics and clinical signs in neonatal sepsis. *Pediatric Research*, 61, 222-227.
- Groves, A. M., Kuschel, C. A., Knight, D. B., & Skinner, J. R. (2005). Cardiorespiratory stability during echocardiography in preterm infants. *Archives of Disease in Childhood*, 90, 86-87.
- Gunn, D. J., Cartwright, D. W., & Gole, G. A. (2012). Incidence of retinopathy of prematurity in extremely premature infants over an 18-year period. *Clinical & Experimental Ophthalmology*, 40, 93-99.

- Hagadorn, J. I., Richardson, D. K., Schmid, C. H., & Cole, C. H. (2007). Cumulative illness severity and progression from moderate to severe retinopathy of prematurity. *Journal of Perinatology*, *27*, 502-509.
- Halliday, H. L., Ehrenkranz, R. A., & Doyle, L. W. (2003). Early postnatal (< 96 hours) corticosteroids for preventing chronic lung disease in preterm infants. *The Cochrane Library*.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36.
- Hagadorn, J. I., Richardson, D. K., Schmid, C. H., & Cole, C. H. (2007). Cumulative illness severity and progression from moderate to severe retinopathy of prematurity. *Journal of Perinatology*, *27*, 502-509.
- Harigopal, S., Satish, H. P., Taktak, A. F. G., Southern, K. W., & Shaw, N. J. (2011). Oxygen saturation profile in healthy preterm infants. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *96*, F339-F342.
- Harrell, F. E., Lee, K. L., Califf, R. M., Pryor, D. B., & Rosati, R. A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, *3*, 143-152.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. CRC Press.
- Hauspurg, A. K., Allred, E. N., Vanderveen, D. K., Chen, M., Bednarek, F. J., Cole, C., & Dammann, O. (2011). Blood gases and retinopathy of prematurity: the ELGAN Study. *Neonatology*, *99*, 104-111.
- Hay, E. (1969). The relationship between environmental temperature and oxygen consumption in the newborn. *Journal of Physiology*, *200*, 589-594.
- Hellström, A., Engström, E., Hård, A. L., Albertsson-Wikland, K., Carlsson, B., Niklasson, A. & Smith, L. E. (2003). Postnatal serum insulin-like growth factor I deficiency is associated with retinopathy of prematurity and other complications of premature birth. *Pediatrics*, *112*, 1016-1020.
- Hellström, A., Hård, A. L., Engström, E., Niklasson, A., Andersson, E., Smith, L., & Löfqvist, C. (2009). Early weight gain predicts retinopathy in preterm infants: new, simple, efficient approach to screening. *Pediatrics*, *123*, e638-e645.
- Henderson-Smart, D. J., Butcher-Puech, M. C., & Edwards, D. A. (1986). Incidence and mechanism of bradycardia during apnoea in preterm infants. *Archives of Disease in Childhood*, *61*, 227-232.
- Higgins, R. D., Bancalari, E., Willinger, M., & Raju, T. N. (2007). Executive summary of the workshop on oxygen in neonatal therapies: controversies and opportunities for research. *Pediatrics*, *119*, 790-796.
- Hintz, S. R., Benaron, D. A., & Stevenson, D. K. (2002). *Biomedical engineering*

aspects of neonatal monitoring. Neonatal-Perinatal Medicine. Mosby, St Louis.

Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, 32, 17-22.

Hofstetter, A. O., Legnevall, L., Herlenius, E., & Katz-Salamon, M. (2007). Cardiorespiratory development in extremely preterm infants: vulnerability to infection and persistence of events beyond term-equivalent age. *Acta Paediatrica*, 97, 285-292.

Hoppenbrouwers, T., Hodgman, J. E., Arakawa, K., Durand, M., & Cabal, L. A. (1992). Transcutaneous oxygen and carbon dioxide during the first half year of life in premature and normal term infants. *Pediatric Research*, 31, 73-79.

Horbar, J. D., Carpenter, J. H., Badger, G. J., Kenny, M. J., Soll, R. F., Morrow, K. A., & Buzas, J. S. (2012). Mortality and neonatal morbidity among infants 501 to 1500 grams from 2000 to 2009. *Pediatrics*, 129, 1019-1026.

Huber, P. J. (1981). *Robust statistics*. Springer, Berlin.

Hunt, C. E., Corwin, M. J., Lister, G., Weese-Mayer, D. E., Neuman, M. R., Tinsley, L., & Cabral, H. J. (1999). Longitudinal assessment of hemoglobin oxygen saturation in healthy infants during the first 6 months of age. *The Journal of Pediatrics*, 135, 580-586.

Ingemarsson, I. (2003). Gender aspects of preterm birth. *BJOG: An International Journal of Obstetrics & Gynaecology*, 110, 34-38.

Jacobson, L., Hård, A. L., Horemuzova, E., Hammarén, H., & Hellström, A. (2009). Visual impairment is common in children born before 25 gestational weeks-boys are more vulnerable than girls. *Acta Paediatrica*, 98, 261-265.

Jagielska, I., Kazdepka-Zieminska, A., Żolnierzewicz, K., Kazmierczak, J., & Szymanski, W. (2008). Antenatal corticosteroid therapy as a benefit for extremely premature infants. *Archives of Perinatal Medicine* 14, 49-53.

Jain, L. and Raju, T.N.K. (2006). Late preterm pregnancy and the newborn. *Clin Perinatol*, 33: xv-xvi.

Jobe, A. H., & Bancalari, E. (2001). Bronchopulmonary dysplasia. *American Journal of Respiratory and Critical Care Medicine*, 163, 1723-1729.

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91, 401-407.

Karna, P., Muttineni, J., Angell, L., & Karmaus, W. (2005). Retinopathy of prematurity and risk factors: a prospective cohort study. *BMC Pediatrics*, 5, 18.

- Kim, J. H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, *53*, 3735-3745.
- Kinsey, V. E., Jacobus, J. T., & Hemphill, F. M. (1956). Retrolental fibroplasia: cooperative study of retrolental fibroplasia and the use of oxygen. *AMA Archives of Ophthalmology*, *56*, 481-543.
- Kinsey, V. E., Arnold, H. J., Kalina, R. E., Stern, L., Stahlman, M., Odell, G., & Patz, A. (1977). PaO₂ levels and retrolental fibroplasia: a report of the cooperative study. *Pediatrics*, *60*, 655-668.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI*, *14*, 1137-1145.
- Kong, L., Fry, M., Al-Samarraie, M., Gilbert, C., & Steinkuller, P. G. (2012). An update on progress and the changing epidemiology of causes of childhood blindness worldwide. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *16*, 501-507.
- Lala-Gitteau, E., Majzoub, S., Saliba, E., & Pisella, P. J. (2007). Epidemiology for retinopathy of prematurity: risk factors in the Tours hospital (France). *Journal Francais D'ophtalmologie*, *30*, 366-373.
- Landwehr, J. M., Pregibon, D., & Shoemaker, A. C. (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, *79*, 61-71.
- Levesque, B. M., Pollack, P., Griffin, B. E., & Nielsen, H. C. (2000). Pulse oximetry: what's normal in the newborn nursery? *Pediatric Pulmonology*, *30*, 406-412.
- Madan, A., Brozanski, B. S., Cole, C. H., Oden, N. L., Cohen, G., & Phelps, D. L. (2005). A pulmonary score for assessing the severity of neonatal chronic lung disease. *Pediatrics*, *115*, e450-e457.
- Mäkikallio, T. H., Huikuri, H. V., Hintze, U., Videbæk, J., Mitrani, R. D., Castellanos, A., & Møller, M. (2001). Fractal analysis and time-and frequency-domain measures of heart rate variability as predictors of mortality in patients with heart failure. *The American Journal of Cardiology*, *87*, 178-182.
- Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*, 354-381.

- Mallows, C. L. (1973). *Influence Functions*. "On Some Topics in Robustness", unpublished memorandum, Bell Telephone Laboratories. Murray Hill, NJ.
- Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*, 591-612.
- Marron, J. S., & Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B: Methodological*, *653-671*.
- Mattia, F. R. (1998). Chronic physiologic instability is associated with neurodevelopmental morbidity at one and two years in extremely premature infants. *Pediatrics*, *102*, e35-e35.
- McColm, J. R., Geisen, P., & Hartnett, M. E. (2004). VEGF isoforms and their expression after a single episode of hypoxia or repeated fluctuations between hyperoxia and hypoxia: relevance to clinical ROP. *Molecular Vision*, *10*, 512.
- McColm, J. R., & Fleck, B. W. (2001). Retinopathy of prematurity: causation. *Seminars in Neonatology*, *6*, 453-460.
- McDonald, A. D. (1964). The Aetiology of Spastic Diplegia A Synthesis of Epidemiological and Pathological Evidence. *Developmental Medicine & Child Neurology*, *6*, 277-285.
- McEvoy, C., Durand, M. & Hewlett, V. (1993), Episodes of spontaneous desaturations in infants with chronic lung disease at two different levels of oxygenation. *Pediatric Pulmonology*, *15*, 140-144.
- McGregor, M. L., Bremer, D. L., Cole, C., McClead, R. E., Phelps, D. L., Fellows, R. R., & Oden, N. (2002). Retinopathy of prematurity outcome in infants with prethreshold retinopathy of prematurity and oxygen saturation > 94% in room air: the high oxygen percentage in retinopathy of prematurity study. *Pediatrics*, *110*, 540-544.
- Miller, A. J. (1990). *Subset selection in regression*. Chapman and Hall.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159-179.
- Mintz-Hittner, H. A., Kennedy, K. A., & Chuang, A. Z. (2011). Efficacy of intravitreal bevacizumab for stage 3+ retinopathy of prematurity. *New England Journal of Medicine*, *364*, 603-615.
- Mok, J. Y., Hak, H., McLaughlin, F. J., Pintar, M., Canny, G. J., & Levison, H. (1988). Effect of age and state of wakefulness on transcutaneous oxygen values in preterm infants: A longitudinal study. *The Journal of Pediatrics*, *113*, 706-709.

- Msall, M. E., Phelps, D. L., DiGaudio, K. M., Dobson, V., Tung, B., McClead, R. E., & Palmer, E. A. (2000). Severity of neonatal retinopathy of prematurity is predictive of neurodevelopmental functional outcome at age 5.5 years. *Pediatrics*, *106*, 998-1005.
- Murdoch, D. R., & Darlow, B. A. (1984). Handling during neonatal intensive care. *Archives of Disease in Childhood*, *59*, 957-961.
- Naumburg, E., Bellocco, R., Cnattingius, S., Jonzon, A., & Ekblom, A. (2002). Supplementary oxygen and risk of childhood lymphatic leukaemia. *Acta Paediatrica*, *91*, 1328-1333.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *J R Statist. Soc. A*, *135*, 370-384.
- Ng, A., Subhedar, N., Primhak, R. A., & Shaw, N. J. (1998). Arterial oxygen saturation profiles in healthy preterm infants. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *79*, F64-F66.
- Ng, E. Y., Connolly, B. P., McNamara, J., Regillo, C. D., Vander, J. F., & Tasman, W. (2002). A comparison of laser photocoagulation with cryotherapy for threshold retinopathy of prematurity at 10 years: part 1. Visual function and structural outcome1. *Ophthalmology*, *109*, 928-934.
- Nicolini, U., Nicolaidis, P., Fisk, N. M., Vaughan, J. I., Rodeck, C. H., Fusi, L., & Gleeson, R. (1990). Limited role of fetal blood sampling in prediction of outcome in intrauterine growth retardation. *The Lancet*, *336*, 768-772.
- Niijima, S., Shortland, D. B., Levene, M. I., & Evans, D. H. (1988). Transient hyperoxia and cerebral blood flow velocity in infants born prematurely and at full term. *Archives of Disease in Childhood*, *63*, 1126-1130.
- O'Brien, L. M., Stebbens, V. A., Poets, C. F., Heycock, E. G., & Southall, D. P. (2000). Oxygen saturation during the first 24 hours of life. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *83*, F35-F38.
- Palmer, E. A. (2006). The factor of time in retinopathy of prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *10*, 500-506.
- Parker, B. J., Günter, S., & Bedo, J. (2007). Stratification bias in low signal microarray studies. *BMC Bioinformatics*, *8*, 326.
- Penn, J. S., Tolman, B. L., & Henry, M. M. (1994). Oxygen-induced retinopathy in the rat: relationship of retinal nonperfusion to subsequent neovascularization. *Investigative Ophthalmology & Visual Science*, *35*, 3429-3435.
- Penn, J. S., Henry, M. M., Wall, P. T., & Tolman, B. L. (1995). The range of PaO₂ variation determines the severity of oxygen-induced retinopathy in newborn rats. *Investigative Ophthalmology & Visual Science*, *36*, 2063-2070.

- Petrini, J. R., Dias, T., McCormick, M. C., Massolo, M. L., Green, N. S., & Escobar, G. J. (2009). Increased risk of adverse neurological development for late preterm infants. *The Journal of Pediatrics*, *154*, 169-176.
- Pierce, E. A., Foley, E. D., & Smith, L. E. (1996). Regulation of vascular endothelial growth factor by oxygen in a model of retinopathy of prematurity. *Archives of Ophthalmology*, *114*, 1219-1228.
- Poets, C. F., & Southall, D. P. (1994). Noninvasive monitoring of oxygenation in infants and children: practical considerations and areas of concern. *Pediatrics*, *93*, 737-746.
- Poets, C. F., Stebbens, V. A., Alexander, J. R., Arrowsmith, W. A., Salfield, S. A., & Southall, D. P. (1991). Oxygen saturation and breathing patterns in infancy. 2: Preterm infants at discharge from special care. *Archives of Disease in Childhood*, *66*, 574-578.
- Poets, C. F., Stebbens, V. A., Alexander, J. R., Arrowsmith, W. A., Salfield, S. A., & Southall, D. P. (1992). Arterial oxygen saturation in preterm infants at discharge from the hospital and six weeks later. *The Journal of Pediatrics*, *120*, 447-454.
- Poets, C. F. (1998). When do infants need additional inspired oxygen? A review of the current literature. *Pediatric Pulmonology*, *26*, 424-428.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 705-724.
- Qiu, X., et al. (2012). Neonatal outcomes of small for gestational age preterm infants in Canada. *American Journal of Perinatology*, *29*, 87-94.
- Raju, T. N., Higgins, R. D., Stark, A. R., & Leveno, K. J. (2006). Optimizing care and outcome for late-preterm (near-term) infants: a summary of the workshop sponsored by the National Institute of Child Health and Human Development. *Pediatrics*, *118*, 1207-1214.
- Razi, N. M., DeLauter, M., & Pandit, P. B. (2002). Periodic breathing and oxygen saturation in preterm infants at discharge. *Journal of Perinatology*, *22*, 442-444.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In *Encyclopedia of database systems* (pp. 532-538). Springer US.
- Regev, R. H., Lusky, A., Dolfen, T., Litmanovitz, I., Arnon, S., & Reichman, B. (2003). Excess mortality and morbidity among small-for-gestational-age premature infants: a population-based study. *The Journal of Pediatrics*, *143*, 186-191.

- Rhein, L., Simoneau, T., Davis, J., Correia, C., Ferrari, D., Monuteaux, M., & Gregory, M. (2012). Reference values of nocturnal oxygenation for use in outpatient oxygen weaning protocols in premature infants. *Pediatric Pulmonology*, *47*, 453-459.
- Richard, D., Poets, C. F., Neale, S., Stebbens, V. A., Alexander, J. R., & Southall, D. P. (1993). Arterial oxygen saturation in preterm neonates without respiratory failure. *The Journal of Pediatrics*, *123*, 963-968.
- Richardson, D. K., Gray, J. E., McCormick, M. C., Workman, K., & Goldmann, D. A. (1993). Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. *Pediatrics*, *91*, 617-623.
- Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*, H2039-H2049.
- Rigatto, H., Brady, J. P., & de la Torre Verduzco, R. (1975). Chemoreceptor reflexes in preterm infants: II. The effect of gestational and postnatal age on the ventilatory response to inhaled carbon dioxide. *Pediatrics*, *55*, 614-620.
- Roberts, C. L., & Lancaster, P. A. (1999). Australian national birthweight percentiles by gestational age. *Medical Journal of Australia*, *170*, 114-118.
- Romagnoli, C. (2009). Risk factors and growth factors in ROP. *Early Human Development*, *85*, S79-S82.
- Rousseeuw, P. J., & Leroy, A. M. (2005). *Robust regression and outlier detection*. John Wiley & Sons.
- Ruppert, D., & Cline, D. B. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *The Annals of Statistics*, 185-210.
- Saigal, S., & Doyle, L. W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, *371*, 261-269.
- Saito, Y., Omoto, T., Cho, Y., Hatsukawa, Y., Fujimura, M., & Takeuchi, T. (1993). The progression of retinopathy of prematurity and fluctuation in blood gas tension. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *231*, 151-156.
- Samuel, T. Y., Bromiker, R., Mimouni, F. B., Picard, E., Lahav, S., Mandel, D., & Goldberg, S. (2013). Newborn oxygen saturation at mild altitude versus sea level: implications for neonatal screening for critical congenital heart disease. *Acta Paediatrica*, *102*, 379-384.
- Saugstad OD. (1998). Oxygen radical disease in neonatology. *Semin Neonatol*, *3*, 231-238.

- Scargle, J. D. (1982). Studies in astronomical time series analysis. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263, 835-853.
- Schwarz, E. C., Grauel, E. L., & Wauer, R. R. (2011). No increase of incidence of retinopathy of prematurity and improvement of its outcome in a university perinatal centre level III-a prospective observational study from 1978 to 2007. *Klinische Monatsblätter für Augenheilkunde*, 228, 208-219.
- Schmidt, B., Roberts, R. S., Whyte, R. K., Asztalos, E. V., Poets, C., Rabi, Y., & Canadian Oxygen Trial Group. (2014). Impact of study oximeter masking algorithm on titration of oxygen therapy in the Canadian Oxygen Trial. *The Journal of Pediatrics*, 165, 666-671.
- Seely, A. J., Bravi, A., Herry, C., Green, G., Longtin, A., Ramsay, T., & Marshall, J. (2014). Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients? *Critical Care*, 18, R65.
- Shang, P., Wan, M., & Kama, S. (2007). Fractal nature of highway traffic data. *Computers & Mathematics with Applications*, 54, 107-116.
- Silverman, W. A. (2004). A cautionary tale about supplemental oxygen: the albatross of neonatal medicine. *Pediatrics*, 113, 394-396.
- Silvestri, J. M., Weese-Mayer, D. E., & Hunt, C. E. (2002). Home monitoring during infancy: what is normal? *Paediatric Respiratory Reviews*, 3, 10-17.
- Slidsborg, C., Olesen, H. B., Jensen, P. K., Jensen, H., Nissen, K. R., Greisen, G., & la Cour, M. (2008). Treatment for retinopathy of prematurity in Denmark in a ten-year period (1996–2005): Is the incidence increasing? *Pediatrics*, 121, 97-105.
- Smith, L. E. (2005). IGF-1 and retinopathy of prematurity in the preterm infant. *Neonatology*, 88, 237-244.
- Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician*, 33, 57-62.
- Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy. The Principles and Practice of Numerical Classification*. Freeman and Company, St. Francisco.
- Stebbens, V. A., Poets, C. F., Alexander, J. R., Arrowsmith, W. A., & Southall, D. P. (1991). Oxygen saturation and breathing patterns in infancy. 1: Full term infants in the second month of life. *Archives of Disease in Childhood*, 66, 569-573.
- Subhedar, N. V., Tan, A. T., Sweeney, E. M., & Shaw, N. J. (2000). A comparison of indices of respiratory failure in ventilated preterm infants. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 83, F97-F100.

- Sun S. (2002). Relation of target SpO₂ levels and clinical outcome in ELBW infants on supplemental oxygen. *Pediatr Res*, *51*, 350A.
- SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. (2010). Target ranges of oxygen saturation in extremely preterm infants. *The New England Journal of Medicine*, *362*, 1959-1969.
- Terrill, P. I., Dakin, C., Hughes, I., Yuill, M., & Parsley, C. (2014). Nocturnal oxygen saturation profiles of healthy term infants. *Archives of Disease in Childhood*, *100*, 18-23.
- The Committee for the Classification of Retinopathy of Prematurity. (1984). An international classification of retinopathy of prematurity. *Arch Ophthalmol*, *102*, 1130-1134.
- Tin, W., Milligan, D. W. A., Pennefather, P., & Hey, E. (2001). Pulse oximetry, severe retinopathy, and outcome at one year in babies of less than 28 weeks gestation. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *84*, F106-F110.
- Tommiska, V., Tuominen, R., & Fellman, V. (2003). Economic costs of care in extremely low birthweight infants during the first 2 years of life. *Pediatric Critical Care Medicine*, *4*, 157-163.
- Vaucher, Y. E., Peralta-Carcelen, M., Finer, N. N., Carlo, W. A., Gantz, M. G., Walsh, M. C., & Higgins, R. D. (2012). Neurodevelopmental outcomes in the early CPAP and pulse oximetry trial. *New England Journal of Medicine*, *367*, 2495-2504.
- Vijayakumar, E., Ward, G. J., Bullock, C. E., & Patterson, M. L. (1996). Pulse oximetry in infants of < 1500 gm birth weight on supplemental oxygen: a national survey. *Journal of Perinatology*, *17*, 341-345.
- Wahba, G. (1990). Spline Models for Observational Data, CBMS-NSF Regional Conference series in applied mathematics. *SIAM*, Philadelphia.
- Wallace, D. K., Kylstra, J. A., Phillips, S. J., & Hall, J. G. (2000). Poor postnatal weight gain: a risk factor for severe retinopathy of prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *4*, 343-347.
- Wand, M. P., & Jones. (1995). *M. C. Kernel smoothing*. Chapman and Hall, London.
- Ward, J.H. (1963). Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.
- Wilkinson, A. R., Haines, L., Head, K., & Fielder, A. R. (2008). UK retinopathy of prematurity guideline. *Early Human Development*, *84*, 71-74.

- Wikstrand, M. H., Hård, A. L., Niklasson, A., Smith, L., Löfqvist, C., & Hellström, A. (2011). Maternal and neonatal factors associated with poor early weight gain and later retinopathy of prematurity. *Acta Paediatrica*, *100*, 1528-1533.
- Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, *36*, 181-191.
- York, J. R., Landers, S., Kirby, R. S., Arbogast, P. G., & Penn, J. S. (2004). Arterial oxygen fluctuation and retinopathy of prematurity in very-low-birth-weight infants. *Journal of Perinatology*, *24*, 82-87.
- Yue, Y. R., Simpson, D., Lindgren, F., & Rue, H. (2014). Bayesian Adaptive Smoothing Splines Using Stochastic Differential Equations. *Bayesian Analysis*, *9*, 397-424.
- Yum, M. K., & Kim, J. H. (2003). A very-short-term intermittency of fetal heart rates and developmental milestone. *Pediatric Research*, *53*, 915-919.
- Zin, A., & Gole, G. A. (2013). Retinopathy of prematurity-incidence today. *Clinics in Perinatology*, *40*, 185-200.