Exploring the relationship between care provided and clinical outcomes for preterm babies born between 27-31 weeks of gestation in England

Thesis for the degree of Doctor of Philosophy

from the Department of Health Sciences,

University of Leicester

Submitted October 2022

Dr Abdul-Qader Tahir Ismail

PhD Supervisors:

Professor Elaine M Boyle

Professor Thillagavathie Pillay

Exploring the relationship between care provided and clinical outcomes for preterm babies born between 27-31 weeks of gestation in England

Dr Abdul-Qader Tahir Ismail

Thesis abstract

My PhD investigated relationships between care provided for babies born between 27-31 weeks gestation in England and their outcomes.

It comprised: a systematic review examining whether place of birth/care affects outcomes for babies born between 27-31 weeks; literature reviews exploring (a) heterogeneity in outcomes for this cohort, (b) heterogeneity of structure and process within neonatal units and (c) concepts around measurement of quality of care; a study exploring the relationship between quality of care and outcomes for these babies.

Systematic review revealed a lack of evidence to guide optimisation of place of birth. Literature reviews demonstrated heterogeneity of outcomes related to degrees of fetal maturity between 27-31 weeks, and that variation in structure and process within neonatal units can impact outcomes.

In exploring relationships between quality of care and outcomes, plans for data collection using questionnaires proved unsuitable during piloting. Data from the National Neonatal Research Database (NNRD) were therefore utilised. To analyse care provided I measured unit compliance with pre-determined, evidence-based standards (administration of antenatal steroids, normothermia on admission, early use of non-invasive ventilation, and appropriate nurse staffing ratio), and then assessed data completion and compliance with National Neonatal Audit Programme (NNAP) measures. I categorised neonatal units (n=113, 4986 babies) into two groups, and compared the top quartile with the rest, using multivariate analyses to look for associations with length of stay (LOS) and pre-discharge mortality.

I found no difference in mortality, but demonstrated a mean reduction in LOS by one day for babies born in neonatal units within the top quartile for compliance with evidence-based and NNAP measures. This supports a relationship between quality of care and outcomes, and the hypothesis that units striving to comply with national guidance and provide evidence-based care have better outcomes. This has potential implications for patient-flow and cost-effectiveness in neonatal care.

Executive summary

Background and preparatory work:

This PhD was undertaken as part of the OptiPrem Study; an NIHR funded national study comparing outcomes for babies born between 27-31 weeks gestation in NICU versus LNU in England, by each gestational week of birth. My work formed Workstream 2 of OptiPrem and comprises a systematic review to explore the current literature base, a review of the literature on outcomes in this population, and a study exploring the relationship between quality of care and outcomes.

Regionalisation of perinatal services in the UK means this cohort of babies are cared for in either neonatal intensive care units (NICU), providing tertiary level care, or local neonatal units (LNU), providing secondary level care. In other countries (e.g., US, Canada, Australia, parts of Europe), neonatal care is more centralised, with all babies born <32 weeks gestation being cared for in NICU. While there is convincing evidence this is beneficial for babies born <27 weeks of gestation, the systematic review I conducted demonstrated a lack of similar evidence for babies born between 27-31 weeks gestation. Furthermore, my literature review, investigating outcomes for this cohort by each gestational week of birth in the context of foetal biology, revealed a gradient of risk, with rates of mortality and morbidity increasing from birth at 31 to 27 weeks.

Workstream 1 of the OptiPrem Study sought to investigate optimal place of birth and care for this, generally understudied, group of babies (despite accounting for ~12% of all preterm babies born in England). However, neonatal units, even of the same designation and within the same healthcare system, are known to vary. This is clearly documented regarding nurse staffing, organisation culture, volume of patients, care practices, and outcomes. My study was designed to ignore designation of unit and instead look for associations between care provided and outcomes for babies born between 27-31 weeks gestation.

Aim:

To develop a method of identifying and measuring differences in care provided to babies born in 2018, between 27-31 weeks of gestation, by neonatal units, to categorise them into two groups based on this, and to investigate associations with outcomes.

Method:

I identified two broad areas through which to explore quality of neonatal care. The first was data completion and compliance with National Neonatal Audit Programme (NNAP) measures. The NNAP produces annually reviewed standards/audit measures based on published national standards or developed by a consensus method. A publicly available annual report provides comparison charts for each neonatal unit's adherence with each measure. Part of the purpose of the NNAP is in identifying outliers, which can result in regulatory action from the Care Quality Commission (CQC) if a locally produced action plan is not adhered to. Therefore, NNAP audit measures have become de facto standards defining good quality of neonatal healthcare.

The second area was compliance with pre-determined evidence-based measures (in absence of being able to use a Delphi approach to obtain consensus of expert opinion) relating to care that occurs in the peripartum period:

- Receipt of any antenatal steroids
- Normothermia within an hour of neonatal admission
- Percentage receiving non-invasive ventilation, of all babies receiving respiratory support on day 1 of life
- Percentage requiring intensive care support on day 1 of life, receiving 1:1 nursing

The NNAP audit measures and non-NNAP measures formed my 'measures of quality of care' (MQC).

For NNAP measures I downloaded publicly available data from the RCPCH website. OptiPrem dataset measures and outcomes were extracted from the data provided by the Neonatal Research Database (NNRD).

After categorising units by my MQC, and defining quartiles, I compared the demographic neonatal and unit profiles of my comparator groups, using Chi-squared, Fishers exact tests and weighted two sample t-tests, as appropriate. Univariate and multivariate analyses (linear and logistic regression) were conducted to explore associations between adherence with MQC and outcomes of mortality and length of hospital stay (LOS) whilst

in neonatal care. Variables included in the multivariate analysis were gestational age, birth weight, gender, multiplicity of pregnancy, IMD score, resuscitation at birth requiring adrenaline or cardiac massage, and place of birth.

Results:

My cohort of babies were born in 119 neonatal units, of which 44 were NICU (37%) and 75 were LNU (63%). Per unit, 64 babies born between 27-31 weeks were born in NICU compared to 30 in LNU, this difference being largely due to babies born at 27 and 28 weeks of gestation (p=<0.01). Significantly more babies in the most deprived IMD_Q quintile (1), and two least deprived quintiles (4, 5) were born in NICU compared to LNU (p=<0.01).

Univariate analysis comparing pre-discharge mortality of higher versus lower performing units by adherence to the NNAP audit measures, detected a difference for the whole cohort of babies (2.2% vs 3.6%; p=0.04), but not for any specific gestational week. Using logistic regression, this difference lost significance (aOR 1.22, 95% CI 0.43 – 1.35). Univariate analysis for LOS also detected a significant difference between groups; it was significantly less in higher performing neonatal units for the total cohort of babies (difference in weighted mean LOS 3.7 days, 95% CI 0.6 to 6.8, p=0.02). Using linear regression, a reduction in LOS by one day was still found in the higher performing group of units (95% CI 1.029-1.081, p<0.001). The variables entered into the model explained 46.2% of the variation in LOS.

Univariate analysis comparing pre-discharge mortality of higher versus lower performing units by adherence to my non-NNAP MQC did not detect any significant difference (aOR 1.22, 95% CI 0.78 – 1.93). However, a similar analysis for LOS did find a significant reduction in higher performing units (difference in weighted mean LOS 3.1 days, 95% CI 0.4 – 5.8, p=0.02). On multivariate analysis, the LOS for babies in neonatal units in the higher performing group was one day less (95% CI 1.008-1.053, p=0.007). The variables entered into the model explained 46.7% of the variation in LOS.

When analysing the cohorts based on higher vs lower performing units, I found a significantly higher proportion of LNU in the higher performing units, and they were more likely to have a less deprived (i.e., affluent) population.

Discussion:

Using multivariate analyses, I found a positive association between units that comply with my evidence based non-NNAP MQC and/or NNAP audit measures, and a reduction in length of stay for babies born between 27-31 weeks of gestation in England by one day. I did not find any association with pre-discharge mortality.

These results must be interpreted with caution, given less than half of the variance in outcomes was explained by the statistical models. This is due to exclusion of important confounders with poor data completion (e.g., condition of baby at birth, mother's health status pre- and during pregnancy, ethnicity, etc.), and other, unknown confounding factors. Therefore, it is still possible that this result does not reflect a true association.

However, these results do support my hypotheses. Units that are striving to comply with national guidance in the form of NNAP audit measures and practice more evidence-based care would be expected to have better outcomes for their babies, and this could result in the small but significant difference in length of stay. This could also be because of differences in structure or provision of other processes of care that we have not measured, which might have an indirect (e.g., early implementation of breastmilk feeds, more opportunities for parents to provide skin-to-skin care) or direct impact on length of stay (e.g., discharging on nasogastric tube feeding, and/or availability of community neonatal nurse follow-up).

If this association is true, the financial implications are significant, given this single day is likely to be a day of special care (with carer present), which neonatal units are reimbursed £535 by NHS England to provide. Furthermore, being able to discharge these babies one day earlier will have an effect on cot capacity, and therefore, movement within and between neonatal units. It is also very likely that there would be a reduction in LOS for babies of other gestational age ranges which may even be greater than for our cohort. However, my finding that higher performing units were more likely to have higher nurse:patient ratios and a more affluent/less deprived population, indicates that inadequate nurse staffing and healthcare inequality will be roadblocks in the ability of many neonatal units improving the care they provide and realising these benefits.

Bismillah-ir-Rahman-ir-Rahim

In the name of Allah, The Most Compassionate, The Most Merciful

Allahumma salli `ala Muhammadin `abdika wa nabiyyika wa rasulika al-nabi al-ummi

O Allah send peace and blessings upon Muhammad, Your servant, Your Prophet, Your Messenger, the one You taught

We created man from an essence of clay; then We placed him as a drop of fluid in a safe place. Then We made that drop of fluid into a clinging form, and then We made that form into a lump of flesh, and We made that lump into bones, and We clothed those bones with flesh, and later We made him into other forms. Glory be to Allah the best of creators.

[Qur'an 23:12-14]

Allah creates you inside the wombs of your mothers, one stage after another, in three veils of darkness. Thus is Allah, your Sustainer.

[Qur'an 39:6]

It was at a very difficult time in my life that I applied for this PhD. I truly did not know whether it would be the right decision to take on something of this magnitude. So, perhaps for the first in my life, I prayed in a complete state of equipoise. I would do what I could to prepare for the application, but I asked Him to only grant me success if this endeavour would be beneficial for me. Looking back now over the last five years, it has indeed been a blessed time. I have accomplished things I could never have dreamed of, by His grace. So, first and foremost – Alhamdolillah (all praise and thanks belong to Allah).

I would also like to thank Tilly and Elaine, my wonderful supervisors, who have patiently guided me along this path and taught me so much. I am thankful to the OptiPrem study team and the study steering committee members, all of whom have provided invaluable feedback and helped shape this PhD. A special mention must go to Dr Sarah Seaton, who was instrumental in providing me with the data my work depended on, and Dr Zia ur Rehman Tanoli, who taught me the more complex statistical methods I required.

I am grateful to the NIHR for funding this PhD as part of the OptiPrem study, and the University of Leicester (in particular, my fellow students of Room 4.11). I am also grateful to the Royal Wolverhampton NHS Trust and New Cross Hospital neonatal unit for providing my OOPR clinical contract and having me back during the pandemic.

Last but not least, this PhD would not have been possible without the prayers and emotional support of my parents and teachers. My sister and brother, having both gone through this process, gave me an idea of what to expect. And of course, my beloved wife, Taiba, my rock, without whom I am nothing, and my beautiful children, Abu-Bakr and Hannah, for being such perfect distractions.

Contents

Contents: Figures

Contents: Tables

Abbreviations

AAP	American Academy of Paediatrics
ACT	Australian Capital Territory
ADEPT	Abnormal Doppler Enteral Prescription Trial
APH	Antepartum haemorrhage
AS	Stenosis of aorta
AVSD	Atrioventricular septal defect
BAPM	British Association of Perinatal Medicine
BBW	Birthweight of baby
BC	British Columbia
BE	Base Excess
BIPAP	Bilevel positive airway pressure
BMI	Body mass index
BMJ	British Medical Journal
BP	Blood pressure
BPD	Bronchopulmonary dysplasia
BW	Birthweight
CCT	Certificate of completion of training
CD14	Cluster of differentiation 14
CH50	50% Haemolytic complement activity
CI	Chief investigator
CINAHL	Cumulated Index to Nursing and Allied Health Literature
CLABSI	Central line associated bloodstream infection
CLD	Chronic lung disease
CO2	Carbon dioxide
COVID	Coronavirus disease
COXI	Cyclooxygenase inhibitor
CP	Cerebral palsy
CPAP	Continuous positive airway pressure
CQC	Care Quality Commission
CRF	Clinical Record Form
CRIB	Clinical Risk Index for Babies
CRN	Clinical Research Network
CSF	Cerebrospinal fluid
CTG	Cardiotocography
DEBM	Donor expressed breastmilk
DOH	Department Of Health
EBM	Expressed breastmilk
EHCI	European Community Health Indicators
ELBW	Extremely low birthweight
EMNODN	East Midlands Neonatal Operational Delivery Network
EOS	Early onset sepsis
EPIPAGE	Epidemiologie des Petits Ages Gestationnels
EPIQ	Evidence-based Practice for Improving Quality
ET	Endotracheal
ETT	Endotracheal tube
EUROPAIN	European Pain Audit In Neonates
EXPRESS	Extremely Preterm Infants in Sweden Study
GA	Gestational age
GDM	Gestational diabetes mellitus
GIRFT	Getting it right first time

GP	General practitioner
GW	Gestational week
HFNPO	High flow nasal prong oxygen
HQIP	Healthcare Quality Improvement Partnership
IC	Intensive care
ICD-9	International Classification of Diseases, Ninth Revision
IGF-1	Insulin-like growth factor 1
IMD_Q	Index of multiple deprivation
IQ	Intelligence quotient
IUGR	Intrauterine growth retardation
IV	Intravenous
IVH	Intraventricular haemorrhage
LBW	Low birthweight
LGA	Large for gestational age
LNU	Local neonatal unit
LOS	Length of stay
LSOA	Lower Layer Super Output Area
MBRRACE-UK	Confidential Enquiry into Maternal Deaths in the UK
MNI-CORP	Maternal, Newborn and Infant Clinical Outcome Review Programme
MQC	Measure(s) of quality of care
MRI	Magnetic resonance imaging
NCAREIA	The National Clinical Audit for Rheumatoid and Early Inflammatory Arthritis
NDAU	Neonatal Data Analysis Unit
NDS	Neonatal dataset
NEC	Necrotising enterocolitis
NHMRC	Australian National Health and Medical Research Council
NHS	National Health Service
NICE	National Institute of Clinical Excellence
NICHD	U.S. National Institute of Child Health and Development
NICU	Neonatal intensive care unit
NIHR	National Institute of Health Research
NIV	Noninvasive ventilation
NLS	Neonatal life support
NNAP	National neonatal audit programme
NNRD	National neonatal research database
NNU	Neonatal unit
NO	Nitric oxide
NQF	U.S. National Quality Forum
NS	Not significant
NSW	New South Wales
OECD	Organisation of Economic Co-operation and Development
ONS	Office of National Statistics
OR	Odds ratio
OSHPD	California Office of State-wide Health Planning and Development
PDA	Patent ductus arteriosus
PDF	Portable document format
PEEP	Positive end expiratory pressure
PN	Parenteral nutrition
PPHN	Persistent pulmonary hypertension of the newborn
Pre-OL	Oligodendrocyte progenitors
PROM	Preterm rupture of membrane

PS	Stenosis of pulmonary artery
PVL	Periventricular leukomalacia
QIIS	Quality Improvement Implementation Survey
QIS	Wualified in specialty
QUALIFY	Instrument for the Assessment of Quality Indicators
QUIPS	Quality In Prognostic Studies
RCPCH	The Royal College of Paediatric and Child Health
RDS	Rspiratory distress syndrome
RNS	Reactive nitrogen species
ROM	Rupture of membranes
ROP	Retinopathy of prematurity
ROS	Reactive oxgen species
RR	Relative risk
SC	Special care
SCU	Special care unit
SD	Standard deviation
SGA	Small for gestational age
SIPAP	Synchronized inspiratory positive airway pressure
SMR	Standardised mortality ratio
SNAP-II	Score for Neonatal Acute Physiology
SSC	Study steering committee
TGA	Transposition great arteries
TIOP	Toward Improving the Outcome of Pregnancy
TPA	Tissue plasminogen activator
UK	United Kingdom
UKNC-NEC	UK Neonatal Collaborative Necrotising Entercolitis
UKNSS	UK Neonatal Staffing Study
US	United States
USA	United States of America
VEGF	Vascular endothelial growth factor
VIF	Variance Inflation Factor
VLBW	Very low birthweight
VON	Vermont Oxford Network
WHO	World Health Organisation
WS	Workstream

1 Introduction

1.1 This PhD in the context of the OptiPrem Study

This PhD formed workstream 2 of OptiPrem; an NIHR funded national study, addressing optimal place of birth/care for preterm babies born between 27-31 weeks of gestation (Figure 1) (2). The basis of this study was that within the UK, regionalisation of neonatal services is such that these babies can be born and cared for in both neonatal intensive care units (NICU – providing tertiary level care, formerly known as level 3 units) and local neonatal units (LNU – providing secondary level care, formerly known as level 2 units). There is currently a lack of evidence on whether this is associated with a difference in outcomes. Babies born between 27-31 weeks of gestation are a relatively understudied cohort despite accounting for ~12% of all preterm babies born in England. There is significantly greater interest in babies born extremely prematurely (<27 weeks of gestation), yet this cohort born between 27-31 weeks gestation contributes around four-fold more throughput and utilises twice as many neonatal bed-days per year (3, 4).

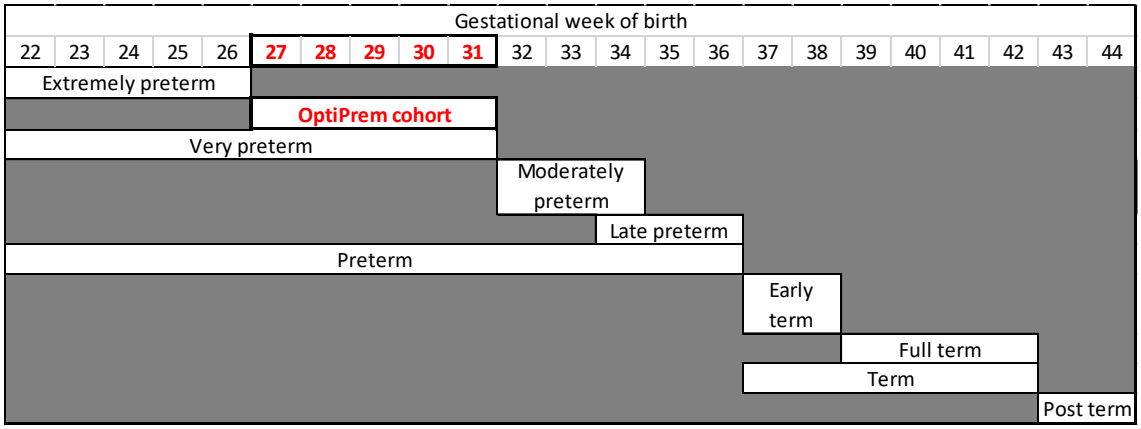


Figure 1 Schematic showing where the OptiPrem cohort sits in relation to established classifications of prematurity

OptiPrem was a three-year retrospective, one year prospective, population based, observational study using national data for England, and had five workstreams (WS). It comprised analysis of routinely collected data, health economic analysis and qualitative research, as illustrated below (Figure 2) (5). The OptiPrem team, led by TP as the chief investigator, set up the study (ISRCTN registry ID - ISRCTN74230187, registered on the Clinicaltrials.gov database - NCT02994849), obtained NIHR funding (National Institute for Health Research, Health Services and Delivery Research Stream, Project number

15/70/104), wrote the initial protocol, obtained approval from the ethics committee (Integrated Research Approvals System reference 212304; Research Ethics Committee reference 17/NE/0800), and advertised for the post, shortlisted, and chose the PhD student to tackle Workstream 2. As the PhD student, I was not involved in any of the above, except for subsequent further refinement of the protocol with regards to Workstream 2.

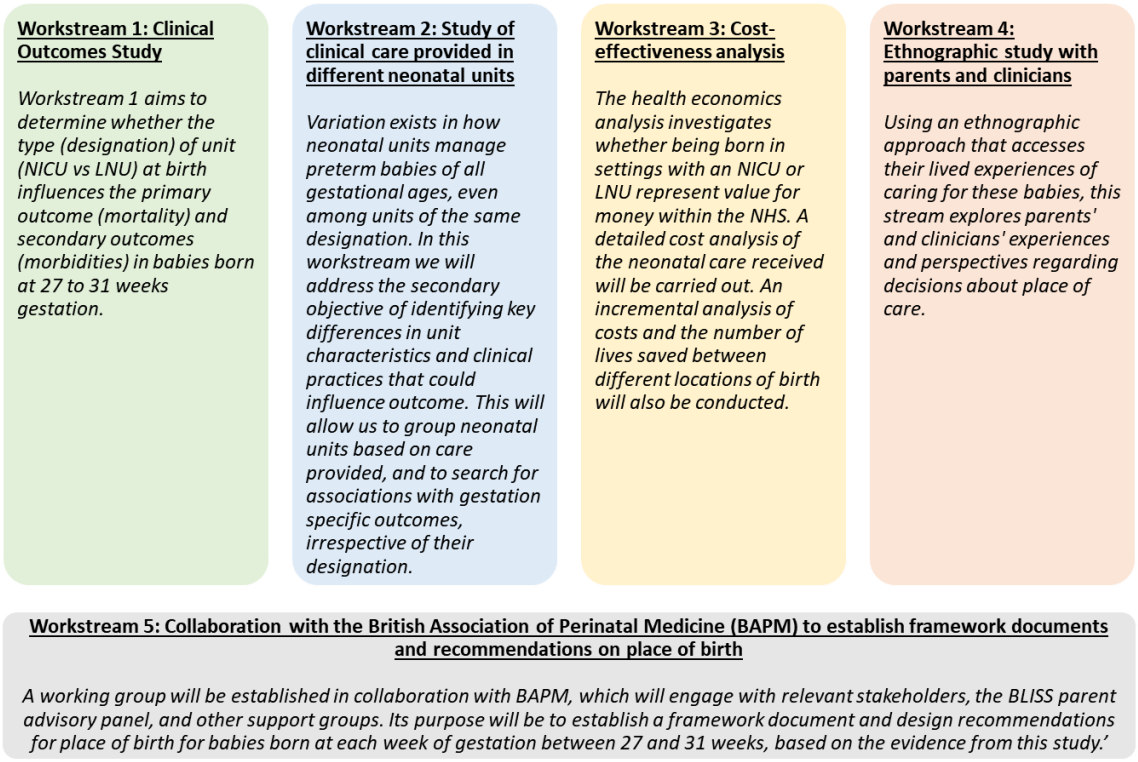


Figure 2 Description of five workstreams included in OptiPrem Study (5)

1.2 Workstream 2 of the OptiPrem study

Workstream 2 (WS2) was based on the premise that differences in the way neonatal units are set up and the care they provide affects outcomes for babies born between 27-31 weeks of gestation, and perhaps this is more important than unit designation. WS2 aimed to find a way to identify these differences and measure them using project-defined measures of quality of care (MQC) (Figure 3). Units would be categorised based on similarities in the quality of care they provide (i.e., independent of their designation). Outcomes would then be analysed by these categorisations to identify associations. Ultimately this work, along with that of WS1, 3, and 4, would feed into WS5 which would involve recommendations on how to improve service provision for this population of preterm babies.

1.3 Overarching research question of this PhD

- How can the quality of care delivered to babies born between 27-31 weeks of gestation be measured, and how can this be used to categorise neonatal units, irrespective of designation?
    - Are there associations between care provided and outcomes?

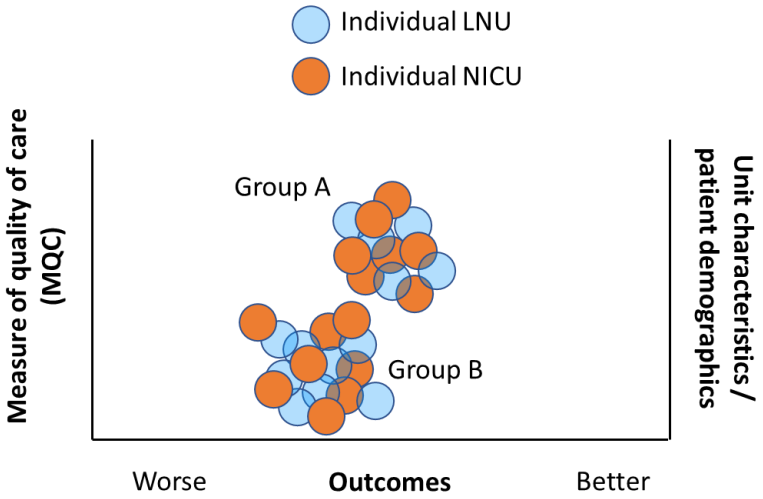


Figure 3 Schematic showing aim of OptiPrem Workstream 2 to categorise units according to measures of quality of care, analyse unit characteristics and patient demographics of resulting groups, and look for associations with outcomes, independent of designation of neonatal unit.

1.4 Covid impact statement

I returned to clinical practice between April – December 2020, to support the NHS during the COVID pandemic. During this time my PhD was on hold, and the university was closed. This created a discrepancy in the timeline of the study for the different workstreams and meant that on return to my PhD I did not have the same level of statistical support as was envisaged at the outset. This inadvertently changed my PhD to a focus on mortality and length of stay, as opposed to mortality and major morbidities. This was accepted by the university reviewers of my PhD progress. A decision to focus on univariate analyses only (as a consequence of lack of statistical support), was originally taken, but revised during my third year, based on the potential value it would add to my PhD results. As a result, multivariate analyses, limited to the most relevant groupings and outcomes was included, with external statistical support.

2 PhD aims and objectives

The overarching aim of this PhD was to explore the relationship between care provided and clinical outcomes for preterm babies born between 27-31 weeks of gestation in England, independent of unit designation.

2.1 Objectives

1. To undertake a systematic review exploring evidence on best place of care for this cohort
2. To conduct a literature review on the heterogeneity in physiology, clinical care required, and outcomes for these babies
3. To examine the literature for existing methods of assessing healthcare quality, in order to identify potential measures of neonatal quality of care
4. To use these measures to categorise neonatal units and perform statistical analyses looking for associations between care provided and clinical outcomes

2.2 Thesis roadmap

- In chapter 3, the rationale for OptiPrem, and workstream 2 is discussed.
    - 3.1 describes regionalisation of neonatal care in the UK within the worldwide context and supporting evidence.
    - This leads to my systematic review in 3.2, which explores whether there is adequate evidence available to answer the question regarding optimal place of birth/care for babies born between 27-31 weeks of gestation.
    - In 3.3, the heterogeneity in physiology, clinical care required, and outcomes for these babies is explored.
    - This heterogeneity is not only present in the patient population, but also the units that take care of them, and this is discussed in 3.4.
- In chapter 4, quality of care is discussed.
    - 4.1 begins with a discussion of the different ways this term has been defined.
    - This is followed, in 4.2, by delving deeper into how I can define and categorise healthcare itself.

    - This leads to a description of the methods by which measures of quality of care can be identified and chosen in 4.3.
    - In 4.4, different methods of evaluating and validating quality of care measures are discussed.
- In chapter 5, the questionnaire I created to collect information on quality of care is described.
    - In 5.1, the process by which the questionnaire was designed and finalised is described.
    - The results of piloting this questionnaire are laid out in 5.2, following by an analysis of these results in 5.3.
    - In 5.4 there is a discussion of what was learnt from this process, how it could be improved upon, but ultimately why it was an unsuitable method for collecting data on quality of care for the purposes of this PhD.
- In chapter 6, a suitable alternative source of data to measure quality of care, the NNRD (National Neonatal Research Database), is discussed.
    - The methods by which neonatal data is collected by the NNRD, managed by the Neonatal Data Analysis Unit (NDAU), and what relation this has with the National Neonatal Audit Programme (NNAP), are summarised in 6.1.
    - In 6.2, the process by which measures of quality of care (MQC) were chosen from the NNRD data available to OptiPrem (non-NNAP MQC) and publicly available NNAP data (NNAP audit measures), is discussed.
    - 6.3 identifies the specific research question being asked in this PhD.
    - In 6.4 and 6.5, the patient demographics and unit characteristics, and outcomes, that will be analysed are specified.
- In chapter 7, the methods are described.
    - In 7.1, a description is provided of the process of sorting the data supplied to OptiPrem by NDAU that was relevant to this work to arrive at the required patient cohort. The process by which it was determined how many babies in each unit received the care specified by my non-NNAP MQC, is also described.
    - In 7.2, the process to determine which of the NNAP audit measures were appropriate to interrogate individual unit compliance and data completion, is described.

    - The specific variables used to describe the demographic profile and unit characteristics of the comparison groups is laid out in 7.3. The statistical tests used are also described.
    - In 7.4, the univariate and multivariate statistical tests used to look for associations with the pre-specified outcomes, is described.
- In chapter 8, the results are presented.
    - The demographics of the patient cohort is described in 8.1, broken down by gestational age and designation of unit. The trends seen regarding adherence with my non-NNAP MQC is also described. Following statistical testing, any significant differences found between the comparison groups regarding demographic profile and unit characteristics is also presented here. Finally, the results of the univariate analysis looking for associations between adherence with my non-NNAP MQC and clinical outcomes, is described.
    - In 8.2, the results of the univariate analysis looking for associations between adherence / data completion and outcomes for the NNAP audit measures, is described.
    - Data for my non-NNAP MQC and NNAP audit measures, is compared in 8.3.
    - The results of the multivariate analysis are presented in 8.4, alongside any transformation of data required to meet the assumptions of the tests.
- In chapter 9, the results are interpreted and discussed.
    - The discussion regarding demographic profile and unit characteristics for the whole cohort, and when split by gestational age and my comparator groups is presented in 9.1. The findings of the univariate analysis looking for associations between adherence to my non-NNAP MQC and outcomes, is discussed. A similar discussion regarding the results for the NNAP audit measures is presented in 9.2.
    - In 9.3, the significance of the multivariate analyses results is discussed.
    - In 9.4, the significance of the findings when comparing my different comparison groups for the NNAP audit measures and my non-NNAP MQC, is discussed.
- In chapter 10, an overarching discussion regarding my PhD is presented.

- The PhD is put in the context of several other studies investigating quality of care in neonatal medicine in 9.5.
- In 9.6, the novel aspects of the PhD work are highlighted.
- The strengths and weaknesses of the PhD are discussed in 9.7.
- A summary of the most pertinent findings from the PhD are presented in the conclusion, in 9.8.
- In 9.9, future work (planned and possible), is discussed.

3 Background: rationale for OptiPrem and WS2

In this chapter I present the rationale for OptiPrem and the PhD, which forms WS2 of the study. I begin with a description of regionalisation of neonatal care in the UK within the worldwide context and supporting evidence. This leads to my systematic review, which explores whether there is adequate evidence available to answer the question regarding optimal place of birth/care for babies born between 27-31 weeks of gestation. I also explore the heterogeneity in physiology, clinical care required, and outcomes for these babies, and describe how this heterogeneity is not only present in the patient population, but also the units that take care of them.

3.1 Regionalisation of perinatal care

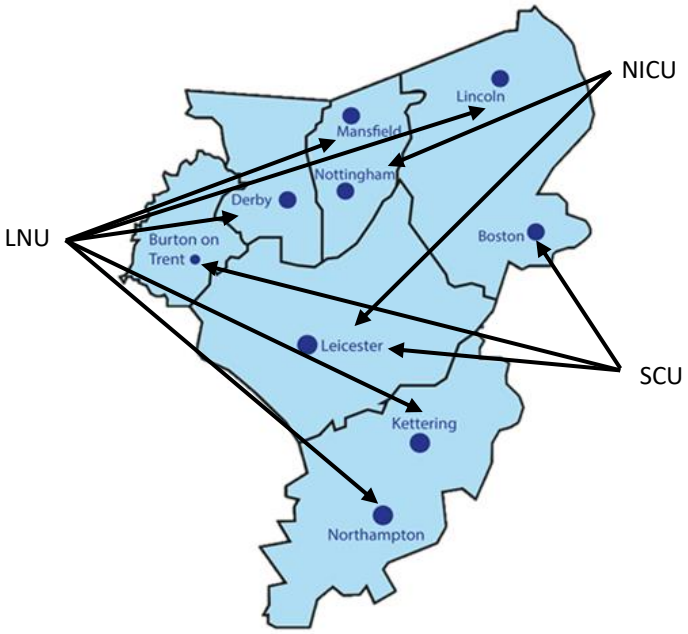


Figure 4 East Midlands Neonatal Operational Delivery Network (EMNODN) showing units of different 'levels' or 'designations'

Regionalisation of perinatal care aims to provide the appropriate level of care for each newborn baby while improving cost-efficiency. Evidence from centralisation of surgical procedures and paediatric intensive care (6-9) had shown improved outcomes, due to increased exposure to complex patients allowing staff to develop and retain expertise. Neonatal intensive care is expensive, due to specialised training of healthcare staff, advanced technology (e.g., ventilators, incubators) and costly consumables (e.g., surfactant, central lines, parenteral nutrition). Applying the industrial concept of

‘economy of scale’, grouping babies requiring intensive care into fewer units lowers overall cost to the healthcare system.

Regionalised perinatal care takes the form of managed clinical networks in which neonatal units (NNU) with different designations are joined by a transport service (Figure 4). Women with complex pregnancies (e.g., impending premature birth or detection of congenital anomalies) are transferred to hospitals with the required level of expertise. Therefore, lower level, local units need to identify cases for early referral and be able to stabilise critically ill and extremely premature neonates prior to transfer, since in-utero transfer is not always possible (e.g., due to advanced labour, maternal sepsis, antepartum haemorrhage, foetal distress). The regional specialist unit provides education and training to its referring hospitals.

Negative aspects of regionalisation include its effect on family-centred care. Separation of mother and infant causes anxiety affecting bonding, especially when mothers are hospitalised locally, but their babies require transfer (10). Travel time and costs are not insignificant (11). To alleviate pressure on intensive care spaces, recovering and growing babies requiring lower-level care are transferred out causing parental anxiety through disruption of continuity of care and relationship with staff (12). When NNU were organised into networks, many units had to stop providing services and treating preterm babies they had previously managed. This led to doctors and nurses feeling devalued, and risks leading to loss of skills and confidence in resuscitation and stabilisation of such patients (13).

3.1.1 Regionalisation of perinatal health care in the developed world

3.1.1.1 Perinatal regionalisation in the US

The first mention of regionalisation in a perinatal setting was in the U.S., in the 1950s. The Hospital Council of Philadelphia proposed a minimum of 2,000 deliveries per year per maternity unit (14). By the 1970s, within several U.S. states sick neonates were referred to hospitals providing neonatal intensive care from local community hospitals (15). In 1976 the March of Dimes foundation published the report ‘Toward Improving the Outcome of Pregnancy’ (TIOP) (16) in which level 1 hospitals would deal with uncomplicated maternity/neonatal care, stabilizing and transferring complicated cases to level 2 units (17). In 2004 the American Academy of Paediatrics (AAP) defined four

levels of care (18) in which all neonates <32 weeks gestation at birth (very preterm) or <1,500g birthweight (very low birthweight – VLBW) should be cared for in level 3 units (Table 1).

3.1.1.2 Perinatal regionalisation in Canada

The Nova Scotia Reproductive Care Program was created in 1971 as a 'voluntary system of perinatal regionalisation' (19). In Ontario, the Ministry of Health recommended a system of perinatal regionalisation in 1972 (20). In British Columbia (BC) the Reproductive Care Program was created in 1988 to support already ongoing regionalisation of perinatal care. In 2005 they defined level of perinatal care, which were renamed and refined as 'perinatal tiers of service' in 2015 (21) in which all neonates <30 weeks of gestation at birth, or 1,200g birthweight should be cared for in tier 3 units (Table 1). Data from 1996-1997 from the Canadian Neonatal Network indicates that 81% of all very preterm neonates were born in tier 3 units (22).

3.1.1.3 Perinatal regionalisation in Europe

Regionalisation of perinatal care within the countries of Europe, similar to the states of the U.S., progressed at different rates using different models (Table 1). For example, in Sweden, initially only the smallest obstetric units were closed, the rest were funded to provide neonatal intensive care to reduce need for transfers (23). In Germany there was significant variation within different regions, e.g., in Baden-Wiirtemberg in 1991 only roughly a third of very low birthweight (VLBW - <1500g) infants were born in regional NICU, while in Hesse the proportion of VLBW infants requiring transfer fell to 39% by 1988 (23). The Netherlands initially employed a voluntary system of perinatal regionalisation from the late 1970s and in France the ministry of Health released a National Policy of Perinatal Care in 1971, advocating perinatal referral centres. Based on an assessment of their perinatal healthcare systems in 1987, the Portuguese Ministry of Health recommended major changes, which included closure of smaller maternity units (<1500 deliveries per year), and stratification of neonatal units into three levels, organised in networks of referring hospitals. In Finland, a highly regionalised system operated, indicated by data from 1987 – 1988 when 52.7% of all LBW babies (<2500g birthweight), and 47.6% of all preterm deliveries (<37 weeks gestation) occurred in university teaching

hospitals (24). Overall, in Europe by 2003, 63% - 93% of very preterm babies were being delivered in level 3 units (25).

3.1.1.4 Perinatal regionalisation in Australia and New Zealand

In 1978, the Australian National Health and Medical Research Council (NHMRC) and Australian College of Paediatrics proposed a plan for regionalisation of perinatal services (26). At the time, there existed 16 NICU but no clear referral or transport system from lower level units (27). A system was implemented relatively rapidly (Table 1); in Victoria from 1985 – 1987, only 23% of extremely LBW (ELBW - <1,000g birthweight) infants were born outside of level 3 units (28, 29). In New Zealand, regionalisation of perinatal care commenced in the 1970s (30) (Table 1). By 1999, the proportion of very preterm deliveries in tertiary level units was ~84% (31).

	Level 1	**Level 2**	**Level 3**
U.S. (32)	• Care for babies born ≥35 weeks	• Care for babies born ≥32 weeks and weight ≥1500g • Stabilise babies born <32 weeks or <1500g, and brief periods of mechanical ventilation, before transfer to a NICU	• Level 3 NICU care for babies of all gestational ages and birthweight • Level 4 regional NICU have level 3 capabilities and are located within an institution with surgical and paediatric medical capabilities
Canada (21)	• Tier 1a care for babies ≥37 weeks and ≥2,500g • Tier 1b care for babies ≥35 weeks and ≥1,800g	• Tier 2a care for babies ≥32 weeks and ≥1,500g • Tier 2b care for babies ≥30 weeks and ≥1,200g	• Tier 3 care for babies of all gestational ages and birthweight with non-life-threatening conditions • Tier 4 provide tier 3 services to babies of all gestational ages and birthweight, including those with life-threatening conditions and requiring paediatric subspecialty input
Australia (33-36)	• Previously labelled level 1 now includes level 1, 2 and 3 • Level 1 and 2 do not provide routine neonatal care • Level 3 care for babies ≥36/≥37 weeks (>2000g/≥2500g)	• Previously labelled level 2a and 2b now includes level 4 and 5 • Level 4 care for babies ≥32/≥34 weeks (>1500/>1700g) • Level 5 care for babies ≥31/≥32 weeks (>1250/>1350g)	• Previously labelled level 3 now includes level 6 • Care for babies of all gestational ages and birthweight, including surgery and congenital and metabolic diseases • May be split into 6a and 6b, with only the latter providing surgical and speciality services
New Zealand (37)	• Care for babies ≥36 weeks	• Care for babies ≥32 weeks • Some units (level 2+) care for babies ≥28 weeks	• Care for babies of all gestational ages and birthweight
Finland (38)	• Smaller, non-university hospitals provide care to babies ≥32 weeks and >1500g		• University Hospitals care for babies of all gestational ages and birthweight
Sweden (39)	• Smaller, non-regional centres provide care to babies ≥28 weeks		• Regional centres care for babies of all gestational ages and birthweight
France (40)	• No neonatal ward • Not required to have a paediatrician on-site	• Care for babies ≥32 weeks • Paediatrician must be present during the day, can be on-call at nights and weekends	• Care for babies of all gestational ages and birthweight • Neonatologist must always be present

Table 1 International summary of organisation of neonatal care services. Extracted from national guidelines and relevant reviews (reproduced with permission from Ismail et al. (1))

3.1.1.5 Perinatal regionalisation in the UK

Within this international context, in 2001 the British Association of Perinatal Medicine (BAPM) advocated regionalisation of neonatal services (41). The Department of Health's (DOH) national review described a non-regionalised system of neonatal intensive care, with a lack of national standards (42). They considered two options; major centralisation of care (like the model employed in the US and many European countries – Table 1, Figure 5), however this was rejected, since *"…neonatal intensive care is often needed for some weeks. Major centralisation would impose considerable travel and other burdens on families. Also capacity and staffing factors argued against major centralisation. It is important that babies are cared for as close to home as possible and that only the sickest babies would require care in the more specialist centres."* The option they chose was of an intermediate degree of centralisation (Figure 5), in which level 2 units (in conjunction with level 3 units) can look after babies born >28 weeks of gestation and 1000g birth weight, and who generally only require intensive care for a short period of time. Other examples of similar 'intermediate' level units are found in places of low population density and large areas, e.g., Iowa, Ontario; where a centralised system is not feasible, and so they serve as regional centres for several community hospitals (43, 44). The aim of this system was to provide safe care to all babies as close to home as possible, minimising unplanned and inappropriate transfers out of region. In 2009 the DOH produced the *'Toolkit for High-Quality Neonatal Services'* (45). Levels of unit were redefined from 1, 2 and 3 to special care baby units (SCU – which would take babies born

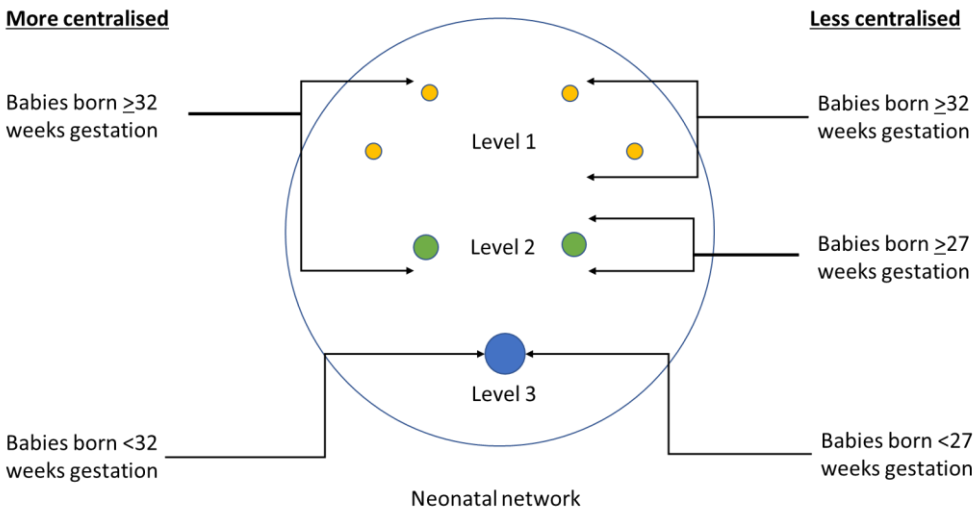


Figure 5 Schematic showing differences between more and less centralised neonatal networks regarding location of birth for babies of different gestational ages

at 34 weeks and above), local neonatal units (LNU – which would now take babies born at 27 weeks and above, and no weight limit) and neonatal intensive care units (NICU).

All of the above indicates that at least in all the countries mentioned, it was believed that regionalisation of perinatal care would improve outcomes. However, the method and speed of implementation varied. Regionalisation did not occur in a vacuum; prior to networks being established there were still neonatal units of different sizes, which were able to provide various levels of care (although many more of them provided intensive care). Therefore, there were still 'smaller' and 'larger' units, and transfers occurred regularly when units were at capacity, or when a baby required a level of care the birth unit could not provide (out-of-region and long-distance transfers were more common). Table 1 summarises the similarities and differences in the approach different countries have taken, compared to the approach taken by the UK shown in Table 2.

Level 1 (Special Care Unit - SCU)	**Level 2 (Local Neonatal Unit - LNU)**	**Level 3 (Neonatal Intensive Care Unit - NICU)**
• Care for babies born ≥34 weeks (or ≥32 weeks depending on local network policy). • Provide special care and may provide some high dependency care. • Stabilise babies who need to be transferred to an LNU or NICU. • Receive transfers from units within their network for continuing special care. • Doctors and nursing staff are on a shared rota with paediatric services. • Consultants are general paediatricians.	• Care for babies born ≥27 weeks of gestation (or ≥28 weeks depending on local network policy) • Provide all categories of care for their local population (including short periods of intensive care), but transfer babies requiring complex or longer-term intensive care to a NICU. • Depending on size and level of activity, doctors and nursing staff may be on a shared or separate rota with paediatric services. • Some consultants have neonatal expertise, others are general paediatricians	• Care for babies of all gestational ages (>22/23 weeks) • Sited alongside specialist obstetric and feto-maternal services. • Provide all categories of neonatal care (including non-conventional modes of ventilation, inhaled nitric oxide, therapeutic hypothermia, etc.). • May be co-located with surgery, and other specialised services. • Consulted for advice and receive transfers from other units within their network. • Doctors and nursing staff are not on a shared rota with paediatric services • All consultants have neonatal expertise.

Table 2 Summary of differences between three levels of neonatal care within the UK
Adapted from British Association of perinatal Medicine (BAPM) (45, 46) (reproduced with permission from Ismail et al. (1))

3.1.3 Evidence for regionalisation of preterm neonates

3.1.3.1 Babies born <27 weeks of gestation

There is compelling evidence that regionalisation of care for babies born <27 weeks of gestation improves outcomes. The EPICure 2 study looked at mortality and short-term morbidity for all babies born between 22-26 completed weeks of gestation in 2006 (47). To ensure correct classification of level 3 units they excluded 'low activity' units (provided <500 days of respiratory support per year and did not have a dedicated neonatal consultant). Their results showed that overall mortality, adjusted for gestational age and birthweight, was significantly lower in level 3 units compared to level 2 (aOR 0.73, 95% CI 0.59-0.90). Furthermore, babies born in level 3 units had significantly higher odds of survival without morbidity compared to ex-utero transfers (OR 1.92, 95% CI 1.02-3.60). A limitation of this study was the use of only two, albeit important confounders of gestational age and birthweight in most analyses.

The EXPRESS (Extremely Preterm Infants in Sweden Study) group conducted a national study using data from 2004-2007 (48). Babies born between 22-27 weeks in level 3 units had significantly reduced odds of infant mortality when adjusted for gestational age (aOR 0.49, 95% CI 0.32-0.75), however when also adjusted for perinatal interventions (e.g., tocolysis, antenatal steroids, delivery by caesarean section and surfactant administration), this difference became insignificant. This would indicate the nearly two-fold reduction in infant mortality was related to care provided in higher level units. However, generalising these findings to other healthcare settings, even in the developed world, must be done with caution. In their population setting there was a high degree of perinatal regionalisation (70% of babies born at <27 weeks were born in level 3 units, of which 75% were transferred in-utero), good general health, and near universal utilisation of pregnancy care (97% of women have routine 17-18 week scans). This is exemplified by survival rates of 26%, 65%, 73% and 84% for neonatal intensive care for births at 22, 23, 24, and 25 weeks of gestation, respectively; compared to figures of 16%, 29%, 46% and 69% from the UK EPICure 2 study (47).

The California Office of State-wide Health Planning and Development (OSHPD) examined outcomes for babies born at 22 – 28 weeks of gestation between 2007 – 2011 (49). NICU were defined as 'intermediate' (short-term ventilation), 'community' (long-term ventilation, limited surgery), and 'regional' (full range of surgery and neonatal

intensive care services). When comparing infants requiring resuscitation at birth, more survivors than non-survivors were born in regional units (21% vs. 17%, p=<0.01), compared to more non-survivors than survivors being born in intermediate units (5% vs. 3%, p=0.03). Due to limitations in their source data, they had to exclude nearly half of their eligible population; however, their mortality figures by gestational week (6%, 27%, 60%, 78% at 22, 23, 24, and 25 weeks, respectively) were similar to findings from other parts of the developed world.

In Victoria, Australia, data on all infants born between 23-27 weeks was available from 2006-2009 and classified into outborn (birth anywhere other than in a level 3 unit), and inborn (50). Outborn infants had significantly higher infant mortality (46%) than inborn infants (21%) even when adjusted for gestational age, birthweight and sex (aOR 3.16, 95% CI 2.52-3.96). However, this might have been lower had outborn births not included hospitals with no neonatal or obstetric services, and births out of hospital. This is partially balanced by inclusion of babies who were purposefully not resuscitated, which would have a greater impact on inborn mortality rates.

Similar evidence from older studies was used by BAPM (41, 51) to make recommendations, later endorsed by the DOH (42), to create networks in which babies born <27 weeks should be looked after in NICU (52-54). Less premature babies (born ≥27 weeks), who do not require long term ventilation or surgery could also be cared for in LNU. However, in much of the rest of the developed world, all babies born <32 weeks of gestation are cared for in NICU, i.e., neonatal healthcare is more centralised.

3.1.3.2 Babies born <32 weeks of gestation

Lasswell et al. (55) conducted a meta-analysis investigating outcomes for babies born with VLBW and <32 weeks of gestation by level of care at birth. They searched for studies published from 1976-2010 in which neonatal and/or pre-discharge mortality data was provided for births in level 3 units compared to lower-level units, regardless of ex-utero transfer. 41 studies (from the U.S., Canada, Europe, Australia, Israel and Ghana) met their inclusion criteria. For quality assessment, they looked at degree of adjustment for confounding factors, and adequate descriptions of levels of care. Their results showed that from the 37 studies (n=104,944), which investigated VLBW babies there was a 62% increase in odds of mortality for birth in non–level 3 units compared to birth in level 3

units (38% vs 23%; aOR 1.62, 95% CI 1.44-1.83). Four studies (n=9,300) investigated outcomes of babies born <32 weeks, which showed a 55% increase in odds of mortality (17% vs 15%; aOR 1.55, 95% CI 1.21- 1.98).

A national Finnish study using data from 2000-2003 compared infant mortality outcomes for 2021 babies born <32 weeks or <1500g birthweight, by level of neonatal unit at the hospital of birth (56). They excluded infants with lethal congenital anomalies, births at level 1 hospitals, or at hospitals with <5 births per annum at <32 weeks of gestation. The adjusted OR (by pregnancy complications, maternal health, birth during non-office hours, birthweight, gestational age, and gender) for birth in a level 2 unit (compared to level 3), was 2.1 (95% CI 1.3-3.3). This OR may be higher than similar studies from other countries due to Finnish classification of NNU, whereby hospitals with <30 annual births of babies <32 weeks of gestation are classified as level 2, and >44 annual births as level 3. Therefore, level 3 units are necessarily higher volume than level 2 units, which is generally associated with improved outcomes.

Similar findings were found in a population-based study from New South Wales and the Australian Capital Territory, looking at pre-discharge neonatal mortality data for 4454 babies born <32 weeks, between 2007-2011 (57). Using a logistic regression model, being outborn was one of the few independent risk factors for mortality (OR 1.43, 95% CI 1.05-1.95). Data from the EPIPAGE (Epidemiologie des Petits Ages Gestationnels) study looked at outcomes of babies born <32 weeks (n=585) born in nine regions of France in 1997, and found that after adjusting for confounding variables, the risk of mortality was significantly higher in level 1 and 2 units (OR 7.94, 95% CI 2.16–29.09) than level 3 units (58). Phibbs et al. (59) and Warner et al. (60) also both found that adjusted mortality odds were increased for VLBW infants by birth in lower level units.

3.1.3.3 Babies born between 27-31 weeks of gestation

Therefore, there is also strong evidence supporting the provision of care for all babies born <32 weeks of gestation in NICU. However, the patient population in these studies include babies born <27 weeks of gestation, so it is possible the significantly improved outcomes seen with care in NICU is not true for the entire gestational age range up to 32 weeks.

There are no studies directly reporting outcomes for babies specifically born between 27-31 weeks gestation. I wanted to explore outcomes specifically for this cohort of preterm babies in a regionalized healthcare system. This is of particular importance in the UK since neonatal healthcare is not fully centralized and these babies are cared for in both NICU and LNU. In the following section I discuss the systematic review of the literature conducted to address this question.

3.2 Systematic review: The impact of level of neonatal care provision on outcomes for preterm babies born between 27-31 weeks of gestation

The PICO for my systematic review was as follows:

Patients: Preterm babies born between 27-31 weeks of gestation (further extended to include babies born with birthweight between 1000-1500g)
Intervention: Birth and/or care provided in tertiary level neonatal unit (NICU)
Comparator: Birth and/or care provided in non-tertiary level neonatal unit (non-NICU)
Outcome(s): All major neonatal clinical outcomes (mortality and morbidity)

3.2.1 Methods

3.2.1.1 Selection of studies

My background reading on variation in outcomes for preterm babies by level of neonatal unit of birth/care, revealed several differences in study design that needed to be accounted for in my search strategy. Studies categorised preterm babies either by gestational age and/or birthweight. Some studies compared babies transferred in-utero versus ex-utero, others excluded ex-utero transfers. Studies categorised units by level, designation or volume of patients.

The initial search strategy was developed for use in the Medline database. To identify key search terms, the titles of journal articles read for my background reading were amalgamated into groups and free online software (https://www.online-utility.org/text/analyzer.jsp) was used to identify recurring words and phrases (Appendix I). The initial search strategy consisted of three sections (combined using 'AND', using 'OR' for terms within each section):

- Population – using the Cochrane Neonatal Review Group search filter (61) with added terms of relevance (e.g. 'preterm')
- Intervention – relating to unit level, regionalisation, volume of patients, antenatal and postnatal transfer
- Outcome – mortality and morbidity, e.g. chronic lung disease, necrotising enterocolitis, intraventricular haemorrhage, retinopathy of prematurity, etc.

I limited my search to studies since 1977. In 1976 the ‘Towards Improving Outcomes of Pregnancy’ (TIOP) paper was released in the US by the March of Dimes foundation (62), which signalled a more concerted and widespread international effort to regionalise perinatal care. I also limited my search to articles published in English due to time and monetary constraints regarding obtaining translations. I do not believe this excluded any relevant studies since the majority of countries with a regionalised perinatal care system publish their scientific articles in English. The search was also restricted to human patients and publication in journals.

The initial search strategy revealed >30,000 results the majority of which were irrelevant. This was because the intervention search terms were too broad (e.g., critical care, intensive care, level 3, level 2, level 1), and because search terms were being located within articles, even when not in relation to each other (i.e. mentioned in different sentences in different sections of the abstracts, title, or keywords). Therefore, adjacency searching was used to search for words that are near to each other – increasing the likelihood of use in the required context e.g. 'neonatal ADJ2 unit' - will find ‘level’ within two words of ‘neonatal’. This decreased the results to <4,000 but raised concerns regarding exclusion of relevant results.

While conducting my background reading, I had identified two groups of studies, those which were for possible inclusion in the systematic review and those which were not for inclusion but were quite similar to the first group. The search strategy should identify studies within both groups and so I checked whether this was true. I could find nearly all the studies in the first group, but a significant proportion of studies in the second group had not been identified.

Therefore, a new ‘intervention’ section was devised in which I produced an exhaustive list of all the phrases I would want to search for. This was used to produce a list of word combination (using ADJ) which would successfully search for the entire list. I also simplified the outcome section by removing terminology for specific morbidities but increasing the number of general terms (i.e. ‘death’, ‘survival’, ‘outcome’). This was because a significant proportion of irrelevant results stemmed from searches for specific morbidities, and because studies that might investigate these outcomes in relation to level of neonatal unit nearly always also investigate mortality. The outcome section was also

combined with the population section (using ADJ), since I was only interested in these outcomes within this population. Finally, I analysed the titles, key words and abstracts of studies within the second group that were missing from the results, to find out how the search strategy needed to be altered or expanded to identify them. The final search strategy identified 3621 results and included all articles in the first group and most articles in the second group.

This search strategy (Figure 6) was modified for use in Embase and CINAHL, utilising the same limitations, returning 3662 and 1339 results, respectively. Results were imported to Endnote and duplicates removed, resulting in 1337, 3448, and 3309 results from CINAHL, Embase and Medline, respectively. The three sets were combined, duplicates removed, yielding a final list of 5048 articles.

These 5048 articles were scanned by their titles to exclude irrelevant ones, leaving 217. The remaining abstracts were analysed by TP and I, with EB arbitrating any differences of opinion, using the following inclusion criteria:

- Availability of abstract
- Data from post-1977, regardless of year of publication
- Comparative studies (by level or volume of patients)
- Grouping of patients by gestational age
- Outcomes compared by hospital of birth or care

This left 52 articles, of which 51 full texts were available. Finally, I excluded articles if they did not compare mortality and/or morbidity outcome measures for preterm babies born between 27-31 weeks of gestation by level of hospital of birth or care. Authors were contacted for further information for studies in which the gestational age range contained or overlapped with, but was not exactly 27-31 weeks, or outcome data was in a non-numerical format. This resulted in nine articles for inclusion in the systematic review (50, 53, 63-69) (Figure 8).

The reference lists of these nine articles and two recent systematic reviews on this topic were analysed (55, 70). No further relevant articles were identified. A search for relevant

grey literature was conducted in the following databases: Opengrey, Scopus, Embase, and Web of Science, revealing no relevant additions.

1	((newborn or neonate or neonatal or premature or very low birth weight or low birth weight or VLBW or LBW or infan* or neonat* or Preterm) adj4 (Mortality or morbidity or death or survival or outcome))
2	(Regionalis* or regionaliz* or centralis* or centraliz* or care level or level of care)
3	((tertiary or regional or level 3 or level III or level 2 or level II or level 1 or level I) adj2 (perinat* or obstet* or neonat* or NICU* or hospital* or centre* or center* or unit or units))
4	((size or level or volume) adj3 (perinat* or obstet* or neonat* or VLBW or LBW or low birth weight or very low birth weight or preterm or premature))
5	((birth or delivery) adj3 (hospital* or unit or units or centre* or center*))
6	((Maternal or neonatal or perinatal or antenatal or postnatal or utero) adj2 (transport* or transfer*))
7	(Inborn or outborn)
8	2 or 3 or 4 or 5 or 6 or 7
9	1 and 8
10	limit 9 to (english language and humans and yr="1977 -Current" and journal article)

Figure 6 Search strategy (Medline)
Adapted for use in Embase and CINAHL (reproduced with permission from Ismail et al. (1))

3.2.1.2 Expanding the systematic review to include studies categorising by birthweight instead of gestation age

Studies published before the 1980s (24, 71, 72) and from the U.S. (59, 73, 74) stratify neonates by birthweight as opposed to gestational age. Historically, this provided greater accuracy, but its use continues due to a strong association with infant mortality (75). This is despite recognition that at any given gestational age there is wide variation in birthweights and that neonates born earlier have worse outcomes than their more mature counterparts of similar birthweight (76). To ensure the paucity of data revealed by my systematic review was not due to excluding studies that categorised neonates by birthweight rather than gestational age, I analysed the 94 articles excluded for this reason. Of the commonly used birthweight stratifications, the closest approximation to 27-31 weeks of gestation were babies born between 1000-1500g (Figure 7). I was aware this

would mean the potential inclusion of babies born <27 weeks of gestation that were large for gestational age (LGA) and ≥32 weeks that were small for gestational age (SGA). It could also mean exclusion of babies born at the lower end of my target gestational age range that were SGA and babies born at the higher end that were LGA. Overall, if I ended up doing a meta-analyses combining data from studies stratifying neonates by gestational age and birthweight, this had the potential to underestimate poor outcomes for babies born at 27-28 weeks, while overestimating them for babies born at 30-31 weeks. 11 articles were identified which compared outcomes for babies born between 1000-1500g birthweight by level of unit of birth or care.

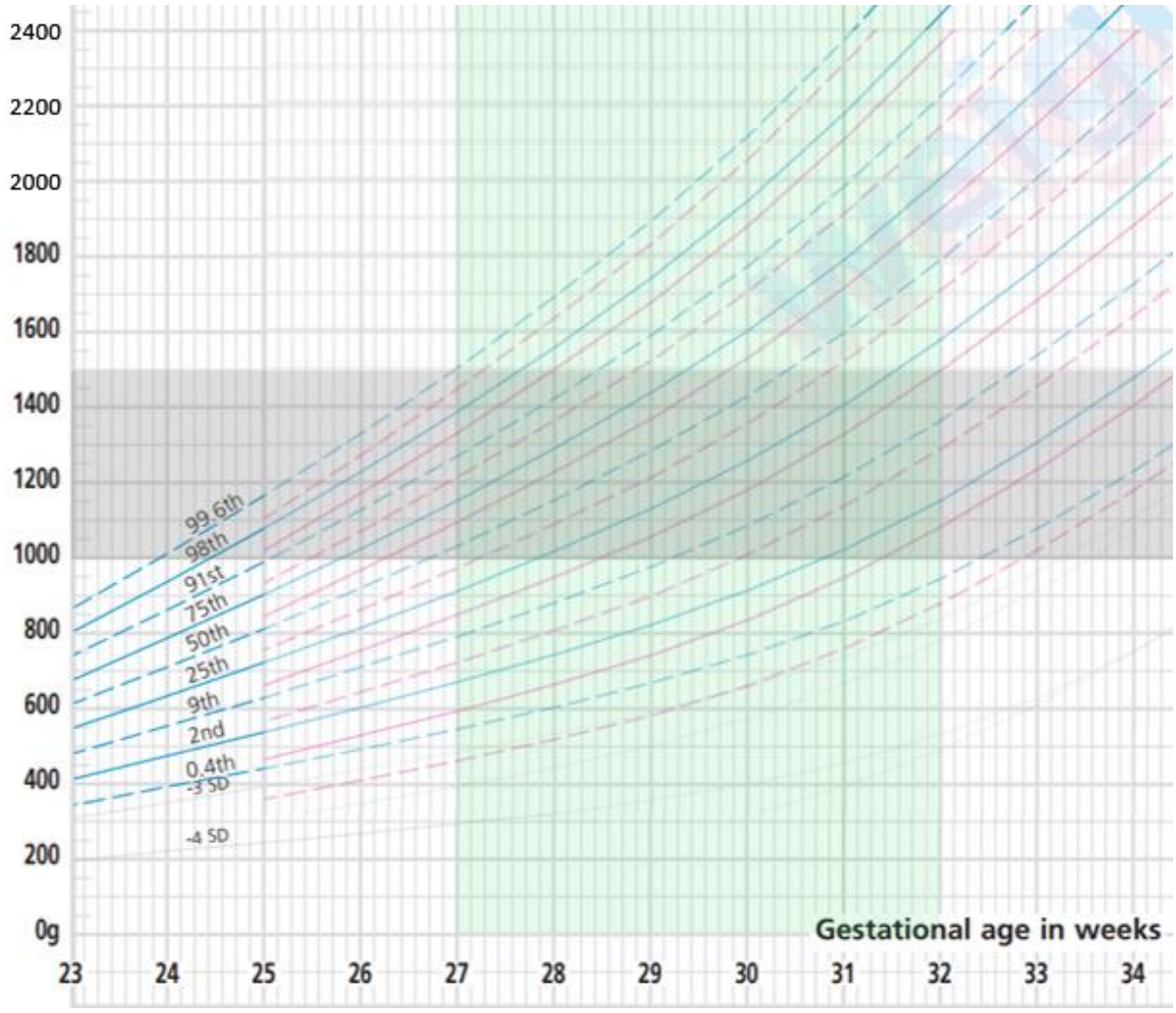


Figure 7 World Health Organisation (WHO) growth chart for preterm babies
Male and female data shown overlapping. To demonstrate overlap between 27-31 weeks of gestation and 1000-1500g birthweight (reproduced with permission from Ismail et al. (1))

3.2.1.3 Study quality

To assess the quality of included studies I used a modified version of the QUIPS (Quality In Prognostic Studies) tool (77). No studies were excluded from the review based on their quality assessment. The QUIPS tool was also used to do a brief analysis of the 11 studies that categorised their babies by birthweight.

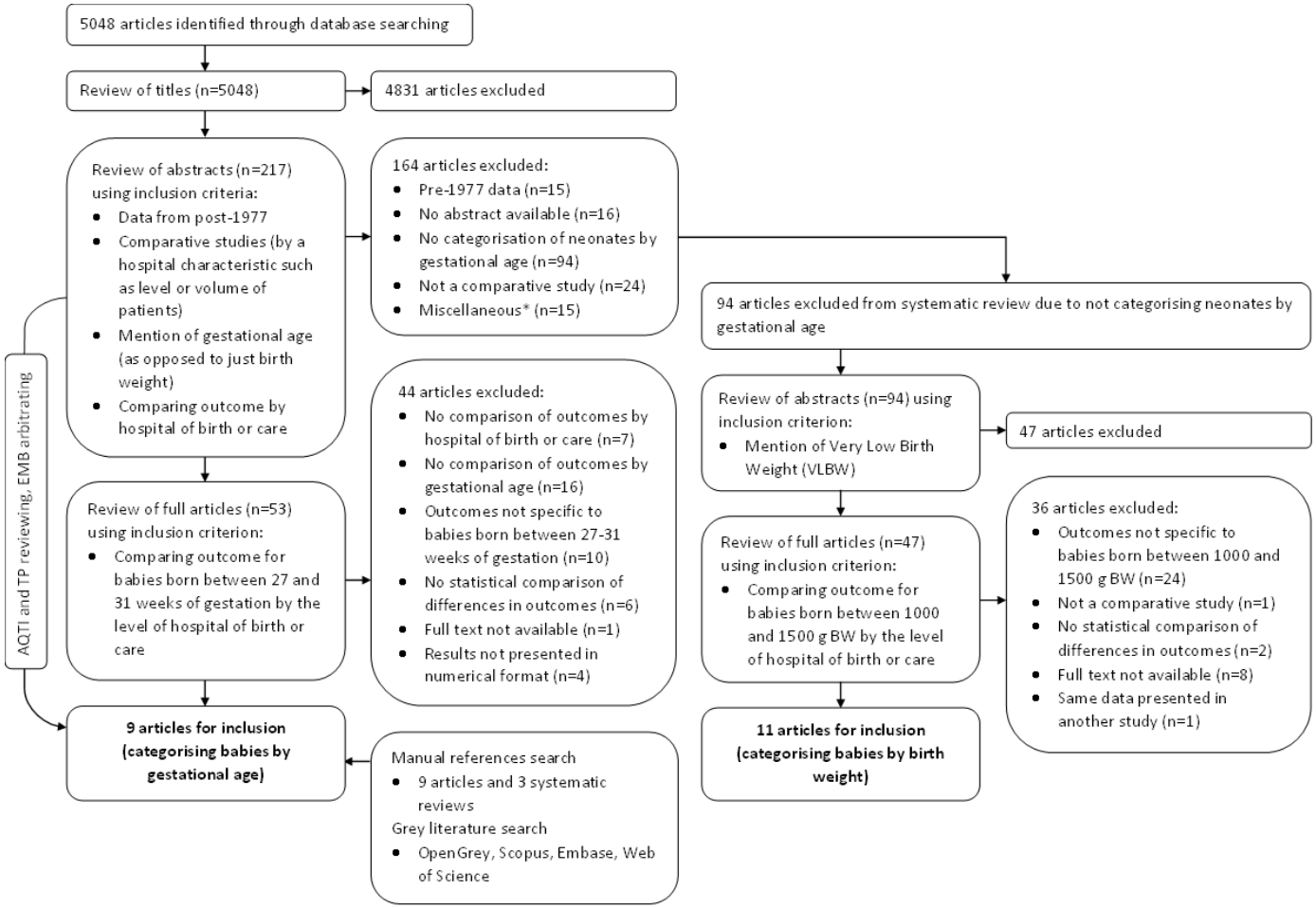


Figure 8 Flow diagram showing results from systematic review search strategy for studies categorising neonates by gestational age and birthweight
Reproduced with permission from Ismail et al. (1)
**Miscellaneous includes studies excluded due to comparing outcomes in NICU vs. NICU/a geographical area/paediatric hospitals/neonatal care in a non-regionalised healthcare system; studies investigating degree of regionalisation/incidence and avoidability of ex-utero transfers; and studies comparing birth asphyxia in term infants/success of using early nasal CPAP*

3.2.2 Results

Of the 5048 articles identified (Figure 8), nine studies were eligible for inclusion based on reporting outcomes for babies born between 27-31 weeks of gestation by designation of hospital of birth or care (Lamont 1983, Field 1991, Jonas 1997, Truffert 1998, Hauspy 2001, Holmgren 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 53, 63-66, 68, 69, 78). In all nine studies the outcomes given for babies born between 27-31 weeks was within a more complete dataset including other gestational ages. Of these, two studies included babies born at 27 weeks (Truffert 1998, Lee 2003) (65, 78), one study did not include babies born at 28 weeks (Field 1991) (53), and two studies did not include babies born at 31 weeks (Field 1991, Truffert 1998) (53, 65). The rest gave outcomes for babies born between 28-31 weeks (Lamont 1983, Jonas 1997, Johansson 2004, Hauspy 2001, Holmgren 2001, Boland 2015) (50, 63, 64, 66, 68, 69). In five studies, babies within the target gestational age range were split further, and outcomes provided separately (27/28-29 weeks and 30-31 weeks) (Lamont 1983, Jonas 1997, Truffert 1998, Hauspy 2001, Lee 2003) (63, 65, 66, 69, 78). I did not identify any gestation-specific data (i.e., by week of gestational age).

A further 11 studies were identified based on birth weight categorisation (1000-1500g) (Miller 1983, Gortmaker 1985, Watkinson 1986, Obladen 1994, Powell 1995, Powell 1997, Yeast 1998, Sanderson 2000, Gould 2002, Warner 2004, Mohamed 2010) (17, 60, 79-87). In these, it was not possible to extract information about those born at 27-31 weeks to allow comparison with the nine other studies. The aim of identifying birthweight studies was to ensure I was not missing a significant source of data from the U.S. and older studies that favoured use of birthweight over gestational age to categorise preterm babies. A summary analysis of these 11 studies revealed this was not the case, and so subsequently I focussed on the nine studies included in the systematic review.

3.2.2.1 Description of studies and assessment of quality

See Appendix I, Tables 47 and 48 for in depth summaries of the quality assessment of included gestational age and birthweight studies using the modified QUIPS tool.

3.2.2.1.1 Study participation

Study settings included single networks (Lamont 1983, Hauspy 2001) (63, 66) and populations (Field 1991, Jonas 1997, Truffert 1998 Holmgren 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 53, 64, 65, 68, 69, 78)). Population based studies are less prone to selection bias and have greater external validity (88). In the context of a regionalised perinatal service, results from a single network depend on outcomes of individual units, which can vary significantly (89-93).

All nine studies defined their exclusion criteria, which varied from including all patients (Field 1991) (53), based on gestational age (Jonas 1997, Truffert 1998, Hauspy 2001, Holmgren 2001, Johansson 2004) (64-66, 68, 69), presence of birthweight (Jonas 1997) (69), congenital anomalies (Lamont 1983, Boland 2015) (50, 63), transfer for surgical correction of congenital anomalies (Lamont 1983) (63), moribund condition at birth and transfer after four days of life (Lee 2003) (78), and birth at units without paediatric services (Johansson 2004) (64). Regarding congenital anomalies, one study did not define this further (Lamont 1983) (63); the other defined it as any baby with a congenital anomaly who died before one year of age, irrespective of a causative relationship between the two (Boland 2015) (50). The seven studies which did not exclude these infants also did not adjust for this in their statistical analysis. Babies with significant congenital anomalies are more likely to be transferred to NICU. Unless excluded or adjusted for this can introduce significant selection bias.

Seven studies compared baseline characteristics of their comparator groups by the level of unit (Lamont 1983, Field 1991, Truffert 1998, Hauspy 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 53, 63-66, 78). These included maternal characteristics, pregnancy factors, neonatal characteristics, unit, and staff characteristics (see Appendix I, Table 47). However, none of the studies specifically compared baseline characteristics for babies born between 27-31 weeks (i.e., they compared characteristics for the entire population meeting their inclusion criteria). Overall, for the category of study participation, one study was of reasonable quality (Boland 2015) (50).

3.2.2.1.2 Study attrition

Six studies were retrospective (Jonas 1997, Truffert 1998, Hauspy 2001, Holmgren 2001, Johansson 2004, Boland 2015) (50, 64-66, 68, 69), two were prospective (Field 1991, Lee

2003) (53, 78), and one could not be classified due to a lack of information (Lamont 1983) (63). The completeness of data on demographic/confounding factors was >95% in all studies (Lamont 1983, Field 1991, Jonas 1997, Truffert 1998, Hauspy 2001, Holmgren 2001, Lee 2003, Boland 2015) (50, 63-66, 68, 69, 78) apart from one (Field 1991) (53), in which it was not possible to determine. In seven studies outcome analysis was carried out on >99% of all babies born within my target gestational age range that met their inclusion criteria (Lamont 1983, Jonas 1997, Truffert 1998, Hauspy 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 63-66, 69, 78). Field et al. (53) did not provide enough information to determine whether all babies born between 28-31 weeks were included in their outcome analysis. In the study by Holmgren et al. (68) up to 13.4% of babies born between 28-31 weeks were not included in the outcome analysis. Overall, for the category of study attrition, seven studies were of reasonable quality (Lamont 1983, Jonas 1997, Truffert 1998, Hauspy 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 63-66, 69, 78).

3.2.2.1.3 Prognostic factor measurement

In all five studies that compared in-utero vs. ex-utero transfer of babies to a NICU, ex-utero transfers were from a range of different level units (Lamont 1983, Truffert, 1998, Hauspy 2001, Lee 2003, Boland 2015) (50, 63, 65, 66, 78), including out-of-hospital births and hospitals without obstetric or paediatric units (Boland 2015) (50). Jonas et al. (69) compared outcomes for babies cared for in level 3 vs. non-level 3 units with no explanation given of the facilities available in each. In the study by Field et al. (53) large/intensive care units were compared with small/special care units, which varied considerably. Some provided intensive care while others transferred all such babies out and as a group, they provided 5-420 ventilation days annually. The two studies that compared outcomes by level of unit of birth defined both the level of units from which babies were transferred and facilities available (Holmgren 2001, Johansson 2004) (64, 68).

Overall, in seven studies the lower-level local units being compared to NICU were either not defined or included birth settings in which preterm babies could not receive an adequate level of care (Lamont 1983, Field 1991, Jonas 1997, Truffert 1998, Hauspy 2001, Lee 2003, Boland 2015) (50, 53, 63, 65, 66, 69, 78). Grouping these babies with those born in level 2 units, which would be able to stabilise very preterm infants and

provide intensive care until transfer, introduces a significant source of bias favouring birth in NICU. Overall, for the category of prognostic factor measurements, two studies were of reasonable quality (Holmgren 2001, Johansson 2004) (64, 68).

3.2.2.1.4 Outcome measurement

The number of babies between 27-31 weeks in each study varied from 157-3331, determined by how many were born and met the inclusion criteria within a specified timeframe. No power calculations were used, so width of confidence intervals (CI) was relied upon to judge whether the patient population was sufficiently large for a lack of a positive outcome to be assumed true. The seven studies which provided 95% CIs showed them to be narrow (Field 1991, Jonas 1997, Hauspy 2001, Holmgren 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 53, 64, 66, 68, 69, 78), except for the low mortality rate in babies born at 30-31 weeks by Hauspy et al. which was 0.23-17.8 (66).

All nine studies examined mortality/survival, but within this there was variation regarding the timeframe (perinatal (Holmgren 2001) (68), pre-discharge (Lamont 1983, Lee 2003) (63, 78), neonatal (Jonas 1998, Hauspy 2001, Holmgren 2001) (66, 68, 69), infant (Holmgren 2001, Johansson 2004, Boland 2015) (50, 64, 68), by 2 years of age (Truffert 1998) (65), and unspecified (Field 1991) (53)). Lee et al. also looked at a combined outcome measure (survival to discharge without major morbidity, including IVH, CLD, NEC, ROP) (78), as did Truffert et al. (survival without disability at two years of age, including cerebral palsy, deafness, Brunet Lezine developmental score <80) (65). Other morbidity measures included RDS (Hauspy 2001, Lee 2003) (66, 78), IVH (≥grade 3) (78), ROP (≥grade 3) (Lee 2003) (78), CLD (Lee 2003) (78), NEC (Lee 2003) (78), and severe asphyxia (as defined by authors) (Holmgren 2001) (68). These outcome measures are routinely used in neonatal research. Field et al. (53) was the only study that did not define their outcome measure with regards to timeframe of mortality. Overall, for the category of outcome measurement, five studies were of reasonable quality (Jonas 1997, Holmgren 2001, Lee 2003, Johansson 2004, Boland 2015) (50, 64, 68, 69, 78).

3.2.2.1.5 Study confounding

Confounding factors adjusted for included gestational age (Jonas 1997, Lee 2003, Johansson 2004, Boland 2015) (50, 64, 69, 78), birthweight (Jonas 1997, Boland 2015)

(50, 69), birthweight for gestational age (Johansson 2004) (64), gender (Jonas 1997, Johansson 2004, Boland 2015) (50, 64, 69), 5-minute Apgar score (Lee 2003) (78), mode of delivery (Jonas 1997, Lee 2003, Johansson 2004) (64, 69, 78), presentation (Jonas 1997, Lee 2003, Johansson 2004) (64, 69, 78), multiple gestation (Jonas 1997, Lee 2003) (69, 78), year of birth (Jonas 1997) (69), maternal age (Jonas 1997) (69), parity (Jonas 1997) (69), marital status (Jonas 1997) (69), maternal hypertension (Lee 2003, Johansson 2004) (64, 78), placental complications (Johansson 2004) (64), antenatal corticosteroids (Lee 2003) (78), antenatal care (Lee 2003) (78), SNAP-II score (Lee 2003) (78), and intubation (Jonas 1997) (69). There is no universally recognised list of confounding variables for which neonatal studies adjust, and a great deal of variation is seen within the literature. Five studies did not adjust for confounding factors (Lamont 1983, Field 1991, Truffert 1998, Hauspy 2001, Holmgren 2001) (53, 63, 65, 66, 68).

When comparing outcomes by level of care in a regionalised healthcare system, adjusting for confounding variables is necessary because higher risk and more unwell patients are disproportionately represented in higher level units. Raw outcome data can indicate similar or worse outcomes compared to lower-level units, but this trend can reverse following appropriate statistical adjustment (58, 64, 94-96). Overall, for the category of study confounding, three studies were of reasonable quality (Jonas 1997, Lee 2003, Johansson 2004) (64, 69, 78).

3.2.2.1.6 Summary of study quality analysis

None of the studies were of reasonable quality across all five domains in my modified QUIPS tool (Table 3). One study was of reasonable quality across three domains (Johansson 2004) (64), four studies across two domains (Jonas 1997, Holmgren 2001, Lee 2003, Boland 2015) (50, 68, 69, 78), and four studies across zero domains (Lamont 1983, Field 1991, Truffert 1998, Hauspy 2001) (53, 63, 65, 66). Therefore, overall, the quality of the studies indicated potential for a high level of bias.

Type of study (comparator groups)	Study	Criteria of modified QUIPS (Quality In Prognostic Studies) tool				
		Study participation (population, exclusion criteria, comparison of baseline characteristics between comparator groups)	Study attrition (prospective or retrospective, data source, completeness of data on demographic/confounding factors, proportion of babies outcome analysis carried out on)	Prognostic factor measurement (definition of birth location, explanation of facilities available at different level units)	Outcome measurement (definition)	Study confounding (adjustment for confounding factors, which variables used)
In-utero vs. ex-utero transfer to NICU	Lamont 1983	x	✓	x	x	x
	Truffert 1998	x	✓	x	x	x
	Hauspy 2001	x	✓	x	x	x
	Lee 2003	x	✓	x	✓	✓
	Boland 2015	✓	✓	x	✓	x
Level of unit of birth (NICU vs. non-NICU)	Holmgren 2001	x	✓	✓	✓	x
	Johansson 2004	x	✓	✓	✓	✓
Level of unit of care (NICU vs. non-NICU)	Field 1991	x	x	x	x	x
	Jonas 1997	x	✓	x	✓	✓

Table 3 Summary of assessment of study quality (categorising babies by gestational age) using modified QUIPS tool
*✓denotes adequate quality, **x** indicates inadequate quality*

3.2.2.2 Outcomes by study design

The studies were all of cohort design but could be divided into three groups based on the following comparators (Table 4):

- Group 1: in-utero versus ex-utero transfer to a NICU for continued care
- Group 2: birth at a maternity service linked to a NICU versus non-NICU irrespective of subsequent main place of care
- Group 3: main place of care in a NICU versus non-NICU, irrespective of the place of birth. Here, place of care referred to either the entirety of care (peripartum and postnatal), or the level of unit of care after the baby was transferred ex-utero.

3.2.2.2.1 Mortality, based on location of birth/care:

Group 1:

I identified five studies that categorised babies by gestational age. Two found significant differences in survival to discharge (Lamont 1983) (63) and infant mortality (Boland 2015) (50) respectively, although Lamont et al. found this only for babies born between 28-29 weeks of gestation. The other three studies did not find a significant difference (Truffert 1998, Hauspy 2001, Lee 2003) (65, 66, 78). Of the four birthweight studies investigating this outcome, three found a significant difference (in neonatal mortality (Watkinson 1986) (87), pre-discharge mortality (Miller 1983) (81), and survival up to 2 years of age (Powell 1987) (85)).

Group 2:

Of the two gestational age studies, neither found a significant difference in mortality (Holmgren 2001, Johansson 2004) (64, 68). Of six studies categorising babies by birthweight, three studies (Gortmaker 1985, Yeast 1998, Sanderson 2000) (17, 79, 86) found a significant difference in neonatal and infant mortality and three did not (Powell 1995, Gould 2002, Warner 2004) (60, 80, 84).

Group 3:

Of the two gestational age studies in the third group, Jonas et al. found a significant reduction in neonatal mortality (69), but Field et al. did not (undefined timeframe) (53).

3.2.2.2.2 Morbidity, based on location of birth/care:

Group 1:

Of the five studies that categorised babies by gestational age, there were conflicting results for incidence of intraventricular haemorrhage (IVH) (Lamont 1983, Lee 2003) (63, 78) and respiratory distress syndrome (RDS) (Hauspy 2001, Lee 2003) (66, 78). A significant reduction was found in the incidence of chronic lung disease (CLD) in babies born at 27-29 weeks (but not 30-31 weeks) (Lee 2003) (78), and no significant difference found for necrotising enterocolitis (NEC) and retinopathy of prematurity (ROP) (Lee 2003) (78). Two birthweight studies also provided conflicting results for incidence of IVH (Obladen 1994, Mohamed 2010) (82, 83).

Group 2:

Of the two studies that looked at morbidity outcomes, the gestational age study found an insignificant difference in the incidence of asphyxia (Holmgren 2001) (68), the birthweight study found significant reduction in composite outcomes of BPD or death, IVH (grade III or IV) or death, ROP or death, but not NEC (Bell stage II or III) or death (Warner 2004) (60).

Categorisation method	Type of study (comparator groups)	Study	Country of study	Total number of babies	Population (gestation (weeks $^{+days}$) / birthweight (g))	Outcomes reported by included studies								
						Mortality time-frame						Survival time-frame		Morbidity
						Un-defined	Peri-natal	Neonatal	Dis-charge	Infant	2 years	Discharge	2 years	
Gestational age	In-utero vs. ex-utero transfer to NICU	Lamont 1983	UK	206	28^{+0} - 29^{+6}							**↑ (71% vs. 49%, p<0.05)**		Non-significant difference in incidence of IVH (46% vs. 57%)
					30^{+0} - 31^{+6}							**NS (94% vs. 92%)**		Significant reduction in incidence of IVH (17% vs. 37%, p<0.05)
		Truffert 1998	France	157	27^{+0} - 28^{+6}						**NS (52.5% vs. 52.5%)**			Non-significant difference in incidence of survival (up to 2 years of age) without disability* (35.3% vs. 40.0%)
					29^{+0} - 30^{+6}						**NS (40.7% vs. 24.3%)**			Non-significant difference in incidence of survival (up to 2 years of age) without disability* (50.0% vs. 58.1%)
		Hauspy 2001	Belgium	315	28^{+0} - 29^{+6}			**NS (11% vs. 12%, OR 0.88, 95% CI 0.25-3.11, p=1.00)**						Non-significant difference in incidence of RDS (69% vs. 79%, OR 0.58, 95% CI 0.23-1.52, p=0.27)
					30^{+0} - 31^{+6}			**NS (4% vs. 2%, OR 1.04, 95% CI 0.23-17.8, p=0.67)**						Non-significant difference in incidence of RDS (58% vs. 55%, OR 1.11, 95% CI 0.59-2.06, p=0.74)
		Lee 2003	Canada	2148	27^{+0} - 29^{+6}				**NS (OR 1.0, 95% CI 0.4-2.5)**					Non-significant difference in incidence of IVH (OR 1.9, 95% CI 0.9-4.0), NEC (OR 0.9, 95% CI 0.4-2.1)) and ROP (OR 1.1, 95% CI 0.4-3.3). Significant reduction in incidence of RDS (OR 2.5, 95% CI 1.6-4.0) and CLD (OR 1.7, 95% CI 1.1-2.8).

					30^{+0} - 31^{+6}				**NS (OR 1.0, CI 0.2-4.5)**					Non-significant difference in incidence of IVH (OR 1.1, 95% CI 0.2-4.8), NEC (OR 0.6, 95% CI 0.1-3.2), and CLD (OR 0.7, 95% CI 0.3-1.6). Significant reduction in incidence of RDS (OR 3.1, 95% CI 2.1-4.8).
		Boland 2015	Australia	250	28^{+0} - 31^{+6}					**↓ (OR 1.66, 95% CI 1.19-2.31, p=0.003)#**				No other outcomes measured
	Level of unit of birth (NICU vs. non-NICU)	Holmgren 2001	Sweden	394	28^{+0} - 31^{+6}		**NS (RR 1.79, 95% CI 0.99-3.25)**	**NS (RR 0.36, 95% CI 0.10-1.32)**		**NS (RR 0.36, 95% CI 0.10-1.32)**				Non-significant difference in incidence of asphyxia¥ (RR 0.95, 95% CI 0.37-2.49)
		Johansson 2004	Sweden	1636	28^{+0} - 31^{+6}					**NS (OR 0.83, 95% CI 0.51-1.33)**				No other outcomes measured
	Level of unit of care (NICU vs. non-NICU)	Field 1991	UK	171	29^{+0} - 30^{+6}	**NS (OR 0.34, 95% CI 0.07-1.61)**								No other outcomes measured
		Jonas 1997	Australia	3331	28^{+0} - 29^{+6}			**↓ (OR 0.11-0.12, 95% CI 0.04-0.05 to 0.30-0.33)˜**						No other outcomes measured
					30^{+0} - 31^{+6}			**↓ (OR 0.29-0.50, 95% CI 0.14-0.21 to 0.20-0.57)˜**						

Categorisation method	Type of study (comparator groups)	Study	Country of study	Total number of babies	Population (gestation (weeks +days) / birthweight (g))	Outcomes reported by included studies								
						Mortality time-frame						Survival time-frame		Morbidity
						Un-defined	Peri-natal	Neonatal	Dis-charge	Infant	2 years	Dis-charge	2 years	
Birthweight	In-utero vs. ex-utero transfer to NICU	Miller 1983	US	94	1000-1500							**↑ (91% vs. 67%, p<0.05)**		No other outcomes measured
		Watkinson 1986	UK	154	1001-1500			**↓ (96.3% vs. 79.2%, p<0.02)**						No other outcomes measured
		Powell 1987	UK	390	1000-1500								**↑ (75% vs. 54%, p=0.02)**	No other outcomes measured
		Obladen 1994	Germany	220	1000-1249							**NS (20.0% vs. 25.0%)**		Non-significant difference in incidence of IVH
					1250-1499							**NS (11.8% vs. 7.0%)**		No other outcomes measured
		Mohamed 2010	US	36493	1000-1500									Significant reduction in incidence of all IVH (17.1% vs. 10.6, aOR 1.73, 95% CI 1.56-1.91, p<0.001), and grade III-IV IVH (25.1% vs. 16.2%, aOR 1.6, 95% CI 1.18-2.18, p=0.003)+
	Level of unit of birth (NICU vs. non-NICU)	Gortmaker 1985	US	4874	1000-1500			**↓ (RR 1.43, p<0.001)=**		**↓ (RR 1.37, p<0.001)=**				No other outcomes measured
		Powell 1995	US	947	1000-1500					**NS (6.8% vs. 5.7%, RR 1.2, 95% CI 0.6-2.4)**				No other outcomes measured

		Yeast 1998	US	2852	1000-1500			**↓ (aOR 2.28, 95% CI 1.33-3.89)[@]**						No other outcomes measured
		Sanderson 2000	US	1345	1000-1249			**↓ (RR 2.98, 95% CI 2.09-3.75)[&]**						No other outcomes measured
					1250-1499			**NS (RR 0.91, 95% CI 0.27-2.94)**						No other outcomes measured
		Gould 2002	US	Un-defined (<4405)	1000-1500			**NS (OR 0.67, 95% CI 0.33-1.35)**						No other outcomes measured
		Warner 2004	US	474	1000-1500				**NS (OR 2.9, 95% CI 0.84-10.20)**					Non-significant difference in incidence of NEC (Bell stage II or III) or death (OR 1.14, 95% CI 0.43-2.70). Significant reduction in incidence of BPD or death (OR 3.62, 95% CI 1.83-7.30), IVH (grade III or IV) or death (OR 4.57, 95% CI 2.11-10.1), ROP or death (OR 5.08, 95% CI 1.8-15.1)[$]

Table 4 Study characteristics and outcomes for studies characterising neonates by gestational age and birthweight

*↑ denotes direction of significant difference found between comparator groups, **NS** denotes lack of significant difference between comparator groups (reproduced with permission from Ismail et al. (1)), original OR, CI and p values are provided where possible*

**Disability is a composite outcome measure consisting of cerebral palsy, deafness, Brunet Lezine developmental score <80 (97)*

[¥]Asphyxia refers to Apgar score <5 at 10 minutes of age

IVH (intraventricular haemorrhage), RDS (respiratory distress syndrome), CLD (chronic lung disease) BPD (bronchopulmonary dysplasia), NEC (necrotising enterocolitis), ROP (retinopathy of prematurity), OR (odds ratio), aOR (adjusted odds ratio), RR (relative risk), CI (confidence interval)

[#]Boland et al, reported mortality for ex-utero vs. in-utero transfers to NICU

[~]Jonas et al. reported mortality for birth in level 3 hospital and transfer from non-level 3 hospital to level 3 hospital vs. birth in non-level 3 hospital with no transfer to level 3 hospital

[+]Mohamed et al. reported IVH for transport vs. inborn groups[@]Yeast et al. reported aOR for neonatal mortality for level II units, using level III units as reference

[&]Sanderson et al. reported RR for neonatal mortality for level II units, using level III units as reference

[=]Gortmaker et al. reported RR of death for 'other urban' (non-level III) units, using level III units as reference

[$]Warner et al. reported aOR for mortality and major morbidity in non-subspecialty perinatal centres (non-SPC) vs. subspeciality perinatal centres (SPC)

3.2.3 Discussion

This is the first review to investigate outcomes of preterm babies born at 27-31 weeks of gestation by the level of neonatal unit of birth and/or care. Overall, the evidence identified in this review was limited, conflicting, and prone to bias. The literature was heterogeneous with respect to gestational ages studied, study design and outcomes.

3.2.3.1 Strengths and weaknesses

Due to a lack of studies specifically investigating babies born at 27-31 weeks, I used a comprehensive search strategy including grey literature. I also reviewed literature comparing outcomes for babies of 1000-1500g birthweight so older studies in which birthweight was favoured over gestational age were not ignored. Therefore, I believe all relevant published literature was identified. Due to time and financial constraints non-English studies were excluded but based on my background reading I do not believe this resulted in exclusion of any relevant studies. I did an extensive analysis of study quality using a modified version of the QUIPS tool. A narrative review was undertaken since a meta-analysis was not appropriate, reflecting the qualities of the available literature.

3.2.3.2 Rationale for conducting a narrative review

The nine studies identified by this systematic review were not suitable for meta-analysis of their results due to differences between studies in five respects.

These nine studies contained three different types of studies. The first group compares babies transferred in-utero vs ex-utero to a NICU from a lower-level unit. I.e., in one arm all neonates are transferred, and subsequently all babies receive the same level of care. Differences in outcome reflect transport and differences in pre-transfer peripartum and postnatal care received in different level units. The second group compares babies born in NICU vs a lower-level unit, regardless of subsequent transfer. While similar to the first group (in that both compare babies receiving different care at birth), in this type of study babies born in lower-level units may or may not be transferred, therefore it is not possible to determine the level of subsequent care unless defined by the authors. The third group compares babies cared for in a NICU vs a lower-level unit. 'Cared for' may mean entirety of care for babies not transferred postnatally, or the level of unit in which the first consecutive 'χ' number of hours of care were received, which may or may not be the unit

of birth. These studies define the different level of care babies receive after birth, which is not the case in the first group of studies and is unknown in the second. We are also not told the level of unit of birth, which is important in both other groups.

Secondly, there was variation in gestational age ranges. Within my target range, neonatal mortality varies by each gestational week (see Section 3.3, Table 5). Seven of the studies provided outcomes for babies born at 28-31 weeks, but only three of these were from the same type of study (50, 63, 66), with two each from the second (64, 68) and third groups (53, 69).

The third reason was differences in outcome measures. Three studies each reported neonatal (66, 68, 69) and infant mortality (50, 64, 68), one study reported both (68). However, only two studies reported infant mortality from the same (second) group (64, 68).

The fourth reason is heterogeneity regarding which type of hospitals or birth locations constitute the 'lower level' comparator group against which NICU are compared. Except for the two studies within the second group (64, 68), 'lower level' facilities were either undefined, or included births in level 1 units, hospitals without paediatric or obstetric units, and out-of-hospital births.

Finally, there was inadequate adjustment for confounding in six studies. The three studies that adjusted for a reasonable number of variables came from each group (64, 69, 78).

Regarding the 11 studies that categorised their babies by birthweight rather than gestational age, these showed similar heterogeneity with regards to the study type, differences in reported outcome measures, definition and facilities available in non-NICU birth locations, and adjustment for confounding variables. Amalgamating the results of these studies with the nine included in the systematic review was not appropriate because of the very imprecise overlap between birthweight and gestational age for babies born between 27-31 weeks (Figure 7).

3.2.3.3 Putting this review into context

There have been two previous similar systematic reviews. In the 1980s, Ozminkowski et al. (70) carried out a meta-analysis investigating neonatal mortality for babies with birthweight <1500g by hospital of birth. They identified 19 articles (1972 – 1984), a meta-analysis of which showed a 38% reduction in odds of neonatal mortality for inborn babies compared to outborn (OR 0.62, 95% CI 0.55-0.69), but with a significant degree of heterogeneity. Subgroup analysis of the eight studies which provided data on babies with a birthweight of 1001-1500g (n=3,180) revealed consistent, statistically significant OR in favour of inborn status (0.53, 95% CI 0.36-0.79). The type of studies included (inborn versus outborn) are similar to the five I identified comparing in-utero and ex-utero transfers (81-83, 87). However, Ozminkowski et al. did not provide information on level of unit or birth location from which outborn babies were being transferred to NICU.

Considering the overall group of preterm babies born at <32 weeks, I have previously discussed the meta-analysis by Lasswell et al. (55), on studies from 1976-2010, in which neonatal or pre-discharge mortality data was provided for births in level 3 units compared to lower level units. 41 international studies met their inclusion criteria. Studies were classified as of insufficient quality if they provided *'no hospital information or lack of clear description of the distinction between hospital levels'*. Even when excluding these studies, their meta-analysis showed increased odds of mortality for birth in non-level 3 units for VLBW (36% vs 21%; aOR 1.60, 95% CI 1.33-1.92) and very preterm babies (born <32 weeks of gestation - 12% vs 7%; aOR 1.42, 95% CI 1.06-1.88). Subgroup analyses were only performed for babies with birthweight <1000g.

Watson et al. (98) advanced this analysis, by identifying that within this cohort of babies, it was predominantly those born at ≤27 weeks of gestation for whom place of birth had a major impact. They showed that care in a high volume (within the top quartile) or NICU was associated with significantly lower mortality to discharge for babies born at ≤27 weeks, but not for those born between 27-32 weeks of gestation.

However, this analysis could be taken a step further, by exploring outcomes by week of gestation for babies born between 27-31 weeks. As discussed in the following section, this population represents a heterogeneous group, with significant differences in the clinical care they require and outcomes when comparing babies born at the lower end of

this gestational age range to those born at the higher end. If the more immature babies within this population have similar outcomes as those born at <27 weeks (regarding place of birth/care), then caring for them in LNU may be associated with worse outcomes and long-term costs. Conversely, perhaps more mature babies would do better in LNU, through the avoidance of over-medicalisation. Even if outcomes are comparable, keeping mothers and their babies in local units could avoid unnecessary transfers and improve family-centred care. The cost to the UK NHS (National Health Service) of providing the same level of care in NICU versus LNU has not been quantified but may also be different. Therefore, grouping babies of 27-31 weeks together might obscure benefits of birth/care in one type of unit over the other.

3.2.4 Conclusion

There is currently a paucity of evidence and data to guide the management of preterm babies born at 27-31 weeks of gestation with respect to place of birth or care. The OptiPrem study is designed to answer this question.

3.3 Understanding heterogeneity in postnatal outcomes for babies born between 27-31 weeks of gestation in the context of fetal biology

The paucity of evidence demonstrated during the systematic review of optimal location of birth and care for babies born between 27-31 weeks reflects a more general lack of research focused on babies born between 27-31 weeks of gestation. During this systematic review, most of the data available for this population was generated from subgroup analyses in studies of larger gestational age ranges. Of these, most reported outcomes for this group as a whole rather than by gestational week. Neonatal research is logistically difficult, especially in relation to very preterm babies, as the population size decreases with each extra gestational week of prematurity. Therefore, it is common practice to cohort babies. While not ideal, this makes more sense for certain gestational age ranges than others.

Babies born between 27 and 31 weeks do not form a 'natural' cohort as do those born extremely preterm. There is a significant degree of heterogeneity in the clinical presentation between babies born at either end of this spectrum. Over this five-week period the foetus is undergoing significant growth and developmental changes in-utero. This chapter describes the limited available literature on the variation in clinical presentation and outcomes for babies born between 27-31 weeks of gestation in the context of fetal developmental biology and preterm birth. In doing so, it highlights the importance of reporting gestation specific outcomes for this cohort in the OptiPrem study.

3.3.1 Survival and key morbidities for babies born between 27-31 weeks

Table 5 summarises outcomes for major neonatal morbidities by each gestational week between 27-31 weeks and includes international mortality data from national statistical bodies. An identical trend is evident for all, demonstrating increasing incidence with decreasing gestational age and substantially different outcomes for the most preterm babies within this gestational age range compared to the most mature. There is, on average, a greater than 4-fold difference in mortality between babies born at 27 weeks of gestation compared to 31 weeks, and a 4-fold increase in rates of survival to discharge without morbidity for babies born at 30 weeks compared to 27 weeks.

Reported outcomes			Outcome (%) by gestational week				
			27	28	29	30	31
Mortality	U.K. – England and Wales (Office of National Statistics) (2013) (99)		7.7	6.5	3.6	2.2	2.2
	U.K. – Scotland (NHS National Services Scotland) (2007-2012)[a]		11.6	7.1	5.5	4.2	2.0
	U.S. (Centre for Disease Control and Prevention) (2015) (100)		6.7	5.6	3.7	2.7	2.3
	Canada (Canadian Perinatal Surveillance System) (2013-2016)[a]		6.5	4.4	2.8	2.4	1.8
	Australia (Australian Institute of Health and Welfare) (2016)[a]		5.1	3.5	2.2	1.6	0.8
	Austria (Statistics Austria) (2015-2017)[a]		7.5	6.1	4.1	3.0	1.6
	Finland (Finnish medical birth register) (2013-2016)[a]		7.9	5.3	4.0	1.4	2.5
	Portugal (Statistics Portugal) (2010-2013)[a]		14.3	7.9	4.8	3.5	2.3
	Netherlands (Infoservice Statistics Netherlands) (2014-2015)[a]		12.0	9.9	3.6	4.0	2.7
	Belgium (Statistics Belgium) (2010-2015)[a]		8.0	5.3	3.5	2.2	1.6
Survival without morbidity (101)			10	15	26	39	-
Survival with severe morbidity[b] (101)			16	12	10	8	-
IVH (102, 103)		Any	33.0		23.0		17.0
		Severe (≥grade III)	42.0	38.0			14.0
PVL (104)			9.9			4.2	
Cerebral palsy (105)			12.3	11.0	8.2	8.3	6.8
NEC[c] (106)			4.2	3.9	1.0	1.0	0.5
CLD (57)			28.1	21.4	11.1	5.9	3.0
Renal failure (107)			9.2	6.0	4.0	3.9	-
ROP (57, 108)		All	30.5			11.0	
		Severe (≥stage III)	4.6	1.5	0.2	0.1	-
PDA (57, 109)		Day 7 of life	68.0	33.0		2.0	
		Requiring surgery	4.5	3.0	1.5	0.8	0.4
Sepsis (57, 110)		Early onset (EOS)	2.0		0.8		
		Late onset (LOS)	27.6	17.7	14.7	7.0	5.4

Table 5 Summary of outcomes from national statistical bodies and international studies showing heterogeneity with increasing gestational age at birth from 27 to 31 weeks
[a]Personal communication
[b]Survival to discharge with severe morbidity, which included grade III/IV IVH, PVL, at least stage III ROP in either eye or requiring surgery, stage III NEC, BPD necessitating oxygen and positive pressure (via non-invasive or invasive ventilation), or >1 episode of infection
[c]Severe disease, defined as disease confirmed by laparotomy, histology or autopsy, or documented as primary cause of death.

3.3.2 Fetal development and neonatal clinical presentation

3.3.2.1 Gut development, feeding, growth, and necrotising enterocolitis (NEC)

3.3.2.1.3 In-utero growth

Birth weights at 27 and 31 weeks reflect a period of rapid in-utero growth, with approximately a 725g averaged difference for babies born on the 50th centile (representing an increase of ~72.5%) (Figure 9). Delay in feeding results in poorer weight gain and longer duration of parenteral nutrition with its associated risks (111). Two main factors contribute to this: poor feed tolerance and episodes of confirmed or suspected NEC.

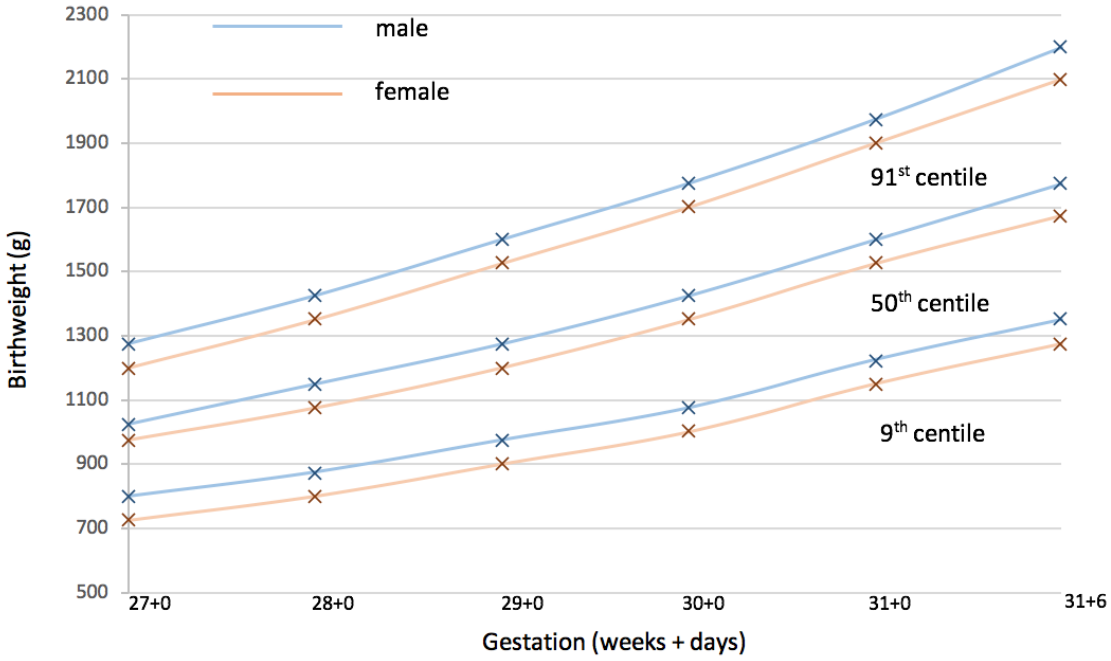


Figure 9 Graph showing increase in birthweight by gestational age from 27+0 to 31+6 weeks
Data shown for male and female babies on the 9th, 50th and 91st centile (as per WHO growth charts) (112, 113)

3.3.2.1.4 Gut development and feeding

A preterm baby's ability to tolerate enteral feeds depends on development of anatomical, motor, digestive and absorptive functions of the immature gut. By 20 weeks, the fetal gut looks anatomically identical to that of a term baby (114), but within the third trimester intestinal length doubles with an even greater increase in surface area (115). Before 31 weeks there is delayed gastric emptying and intestinal smooth muscle contractions are disorganised (116, 117). From 31-34 weeks, contractions begin to cluster, propagation increases from 50% to 90% and gastric emptying time shortens to term equivalent.

From 26-34 weeks, secretion of digestive enzymes including maltase, glucoamylase, sucrase and isomaltase, reaches 70% of adult levels, and lactase is at 30% of newborn levels (118). Pancreatic and lingual lipase is detected from 30 weeks onward (114, 119), while bile acid concentration is 33%-50% that of term babies and is below the critical micellar concentrations required for lipid solubilisation. Monosaccharide absorption by day 14 of life is higher in babies born between 28-32 weeks compared to <28 weeks (120). These anatomical, motor, digestive and absorptive functions manifest clinically, with more immature preterm babies (born at <29 weeks gestation) tending to take a longer time to establish full enteral feeds than their older preterm counterparts (121).

3.3.2.1.5 Necrotising enterocolitis (NEC)

In preterm babies, poor absorption and gut stasis lead to bacterial overgrowth in the context of a pathogenic microbiome and impaired gut wall integrity, contributing to the development of NEC (122). The incidence of developing severe NEC for babies born at 27 weeks gestation is 8 times that at 31 weeks of gestation (Table 5) (106, 123).

3.3.2.2 Maturation of central respiratory drive and lung development, respiratory distress syndrome (RDS) and chronic lung disease (CLD)

3.3.2.2.6 Maturation of central respiratory drive, and apnoea of prematurity

Between 27-31 weeks, the central respiratory drive centres, which control the diaphragm, intercostal and laryngeal muscles, are maturing. In-utero breathing stimulates lung growth (124) and by 24-28 weeks, fetal breathing movements occur for 10-20% of the time, increasing to 30-40% by 30 weeks (125). Due to this immaturity of central respiratory drive, preterm babies experience periods of hypoventilation and apnoea. The incidence falls from 54% at 30-31 weeks to 7% at 34-35 weeks (126). In those born at 24-27 weeks, apnoeic episodes are more likely to continue for longer compared to those born $\geq$28 weeks (127).

3.3.2.2.7 Lung development, respiratory distress syndrome (RDS)

During the saccular stage of fetal lung development (24-26 weeks to 36-38 weeks), surface area for gas exchange and vascularisation increases. Although surfactant production commences with type II pneumocyte differentiation at 23-24 weeks, levels are not sufficient to prevent alveolar collapse until near term. Before around 30 weeks of gestation, the fetal adrenal cortex cannot produce cortisol to initiate the switch from lung fluid production to resorption (128). Respiratory compromise is therefore more likely with increasing prematurity, and the incidence of RDS is 60-80% at 26-28 weeks, falling to 15-30% by 32-36 weeks (129).

3.3.2.2.8 Chronic lung disease (CLD)

The more preterm babies often require mechanical ventilation rather than non-invasive respiratory support. Ventilator-associated lung injury causes increased capillary permeability, alveolar and interstitial oedema, cytokine production and inflammation

(130, 131). The more immature the lung, the greater the potential for subsequent abnormal development and CLD (132). Its incidence is 9 times greater in babies born at 27 weeks than at 31 weeks of gestation (Table 5) (57, 133).

3.3.2.3 Nephron development and renal compromise

A similar association between decreasing gestational age and renal failure is evident (107, 134). Here, the incidence is 2-fold higher for a baby born at 27 weeks compared with 30 weeks of gestation (Table 5). The degree of prematurity correlates with numbers of functioning nephrons. Two thirds of new nephrons form between 28-36 weeks, after which no new glomeruli develop (135, 136). Following preterm birth, nephrogenesis can continue for up to 40 days (137, 138), but a significant proportion of new glomeruli have cystic dilatation of the Bowman's capsule (139).

3.3.2.4 Retinal vascularisation and retinopathy of prematurity (ROP)

In-utero retinal vascularisation continues until term (140). Following preterm birth, even for a baby breathing in air this is a relatively hyperoxic environment, causing suppression of vascular endothelial growth factor (VEGF) and erythropoietin (141). There is a concomitant fall in maternally derived insulin-like growth factor 1 (IGF-1), which normally increases with gestational age. Vascularisation stops, causing hypoxia due to increasing metabolic demands of the developing retina. This stimulates VEGF and erythropoietin production causing abnormal neovascularisation.

Gestational age is the most important risk factor for developing ROP (OR 1.788, 95% CI 1.418-2.254) (Table 1) (108). In babies born at 27 weeks who survive to discharge there is a 46-fold increase in incidence of severe ROP compared to those born at 31 weeks (Table 5) (57) .

3.3.2.5 Cardiovascular adaptation: patent ductus arteriosus (PDA)

Following birth at term, the ductus arteriosus constricts in response to reduced flow, increasing oxygen partial pressure and fall in prostaglandin levels. In preterm babies this is less likely because of reduced vessel tone and reduced pulmonary clearance of prostaglandins (109). There is a 10-fold increase in the likelihood of requiring surgery for

a PDA (i.e., a clinically significant PDA) in those born at 27 weeks gestation when compared to those at 31 weeks (Table 5).

3.3.2.6 Maturation of components of the innate immune system, and neonatal sepsis

Sepsis is a major cause of mortality in preterm babies, with incidence increasing with prematurity (57, 110). There is a 5-fold increase in the incidence of late onset sepsis (LOS) in a baby born at 27 weeks of gestation compared with 31 weeks (Table 5), that may relate to rates of development of immune system components.

Endothelial cell and neutrophil adhesion molecule expression and selectin mediated capture are decreased in babies born between 30-36 weeks compared with term babies, and nearly absent before 30 weeks (142). Plasma concentrations of the FcRIII receptor (reflecting mass of circulating neutrophils and bone marrow production (143)) are around 15% at 24-32 weeks, reaching normal adult values by term (144). Monocytes in babies born <29 weeks have reduced expression of CD14 cell surface markers (145). Such antimicrobial pattern recognition receptors (including toll like receptors) continue development until 33 weeks. However, for up to 28 days after preterm birth at <30 weeks toll like receptor responses are significantly reduced compared to term babies (146). Regarding the complement system in term and preterm babies (147), levels of C3 increase from 48% at <30 weeks to 60% between 30-37 weeks compared to adult levels. CH50 assay results (indicating overall functionality of the complement system) increase from 32%-36% at 26-27 weeks, to 52%-81% at term. During the third trimester transplacental transfer of IgG rises from 10% of maternal levels at 17-22 weeks to 50% at 28-32 weeks (148).

There is obviously a complex interplay between these individual components of the innate immune system, how they interact with the developing adaptive immune system, the microbiome, and physical and external factors such as skin barrier integrity, repeated invasive procedures and indwelling plastic catheters, all of which are also related to degree of prematurity.

3.3.2.7 Neurological development, brain injury and developmental outcome

3.3.2.7.9 Intraventricular haemorrhage (IVH)

A 2-fold increase in incidence of any intraventricular haemorrhage (IVH) in babies born at 27-28 weeks of gestation vs 31 weeks has been reported (Table 5). In one series, severe (>stage III) IVH appeared to be three times more common in those born at 27 weeks than those born at 31 weeks (102, 103).

The germinal matrix is a rich, cellular region from which all neuronal cells migrate outwards in the developing brain. It is active from 8 weeks of gestation, reaching maximum size at around 23 weeks of gestation and beginning its regress from 26-28 weeks (149). Its dense supply of fragile blood vessels are prone to rupture with fluctuations in cerebral blood flow, causing the bulk of what is described in the literature as IVH. The risk is increased due to immature cerebral autoregulation, in which hypoxaemia, hypercapnia, hypocapnia, and acidosis (common in the immediate postnatal period following very preterm birth) cause pressure passivity (150, 151). This, combined with increasing severity of respiratory illness and homeostatic disturbances in the more preterm baby, may explain the inverse association of IVH with gestational age.

3.3.2.7.10 Periventricular leukomalacia (PVL)

The trend is similar for periventricular leukomalacia (PVL) (Table 5) (152). Non-cystic PVL is characterised by hypomyelination, primarily due to loss of oligodendrocyte progenitors (pre-OL) and pathogenic activation of microglia (153). By 28-30 weeks, increasing differentiation of pre-OL to immature oligodendrocytes coincides with the start of myelination (154, 155). Microglia also increase in concentration within developing white matter during this time. They have a number of roles, including phagocytosing apoptotic neurons, secreting growth factors, and stimulating oligodendrocytes to produce myelin (156-158). Hypoxia, infection or inflammation cause microglia activation and release of reactive nitrogen and oxygen species (RNS/ROS), in turn causing cell death of pre-OL (159, 160). Falls in cerebral perfusion pressure are also most likely to affect cells within watershed areas, such as the periventricular region. This explains why babies born preterm before the third trimester are at markedly higher risk of developing PVL.

3.3.2.7.11 Cerebral palsy

Preterm babies with severe IVH (grade III/IV) and cystic PVL are at increased risk of cerebral palsy (161). There is a nearly 2-fold increase in incidence of cerebral palsy for a baby born at 27 weeks compared with 31 weeks of gestation (Table 5).

3.3.2.7.12 Brain growth

In-utero, cortical volume increases from 13% at 28 weeks to 53% at 34 weeks. Babies born preterm have reduced growth trajectories of their cerebrum, cerebellum, and brainstem compared to fetuses within the last trimester (162). Compared to children born at term, those born between 26-33 weeks of gestation have reduced volumes of their premotor and sensorimotor cortex, parieto-occipital, subgenual and midtemporal regions, and cerebellum, which correlates to their IQ (163). Each extra week of maturity at birth between 27-32 weeks is associated with an increased IQ of 2.5 points (164).

3.3.3 Comparing clinical profiles of babies born at 27 weeks and 31 weeks of gestation

This summary of fetal development during the 5-week period from 27 to 31 weeks helps explain the difference between the presentation, and the level of postnatal support required by preterm babies born at either end of this gestational age range.

Many of those born at 27 weeks require mechanical ventilation in the immediate postnatal period and surfactant therapy, to reduce risk of developing CLD, and oxygen delivery is carefully managed to reduce risk of developing ROP. Monitoring of arterial blood pressure and $PaCO_2$, together with appropriate response to deviations from normal help maintain homeostasis to reduce risk of developing IVH. When mechanical ventilation is no longer needed, non-invasive respiratory support is often used to minimise risk of respiratory failure. Parenteral nutrition is given via a central line while nasogastric tube feeds with mother's milk or donor breastmilk are slowly increased. Not infrequently, an episode of bilious aspirates or abdominal distension will cause feeds to be paused as part of conservative management of suspected or proven NEC. Some will require nephrotoxic drugs, e.g., ibuprofen for treatment of a haemodynamically significant PDA, or vancomycin for treatment of central line associated bloodstream infection (CLABSI), warranting close monitoring of renal function and urine output.

In contrast, the majority of babies born at the higher end of this gestational age range will generally require less intensive support. A brief period of continuous positive airway pressure (CPAP) or high flow nasal prong oxygen is often deemed more appropriate management than invasive mechanical ventilation. A peripheral infusion of glucose will maintain hydration while nasogastric tube feeds of either mother's milk or preterm formula are generally, quite rapidly established.

3.3.4 Conclusion

This section highlights the degree of heterogeneity in outcomes for babies born between 27-31 weeks of gestation. I have examined how these relate to key aspects of organ/system development occurring in-utero during this 5-week period. The data summarised in Table 5 consistently demonstrate a gradient of risk across multiple outcomes with rates of mortality and morbidity increasing from birth at 31 to 27 weeks. Outcomes at the two extremes of this range may differ significantly. This is the rationale for investigating outcomes by each gestational week of birth in OptiPrem. This heterogeneity is not only present in the patient population, but also the units that take care of them, and I discuss this in the next section.

3.4 Exploring heterogeneity of structure, process, and outcomes in neonatal units of the same designation

The primary analysis in OptiPrem (WS1) will look at gestation specific outcomes between 27-31 weeks by designation of unit (i.e., outcomes for babies born in LNU versus NICU). However, neonatal units, even of the same designation and within the same healthcare system, are known to vary, regarding how they are set-up, the care they provide, and outcomes.

3.4.1 Obstetric services

Snowden et al. (165) investigated obstetric volume by number of deliveries of term, >2500g birthweight babies in 268 hospitals in California, in 2006. Rural, generally low volume hospitals had births per annum from <600 to >1700, while urban, generally high-volume hospitals (including teaching hospitals) had births per annum from <1200 to >3600. Pyykonen et al. (166) found similar variation in volume of births at term, at non-university hospitals in Finland between 2006-2010 (from <500 to >5000 births per annum).

Joyce et al. (167) used birth data (n=540,834) from 65 Thames Region hospitals from 1994-1996 to ascertain the effect of obstetric factors on stillbirth rates. Their statistical analysis was adjusted for birthweight. It revealed that for obstetric interventions associated with reduced stillbirth rate, incidence varied from 53.0-81.5/100 births for spontaneous vaginal deliveries, from 8.0-33.4/100 deliveries for number of caesarean sections, from 5.0-19.1/100 births for number of instrumental vaginal deliveries, from 2.6-55.5/100 deliveries for epidurals for labour, and from 1.5-52.5/100 caesarean sections for general anaesthetic. There was also an association with the number of consultant obstetricians, which varied from 0.7-4.7/1000 births.

3.4.2 Nurse staffing

Hamilton et al. (168) used data from the UK Neonatal Staffing Study, involving 54 units, and 35651 records of nursing shifts. From this they calculated the total number of registered nurses and expected number of nurses per shift (for specialist and non-specialist nurses). 57% of shifts were understaffed, and 65% of units had an average ratio of <1.0 (indicating inadequate nurse staffing according to BAPM guidance at the time).

Each unit had understaffed shifts, ranging from 13.4% to 90% (both of these were from 'large' units in a pre-regionalised neonatal healthcare system). 23.6% of shifts were understaffed regarding specialist nurses, ranging from 0.7% to 65.1%. Rogowski et al. (169) used data from the Vermont Oxford Network to examine effects of nurse staffing levels on rates of nosocomial infection. The study included 2009 data from 67 NICU, involving 3645 nurses. They found that 55% of units understaffed ≥25% of their patients and 16% understaffed ≥50% of their patients, with only five adequately staffed units. Callaghan et al. (170) used 1996-1999 data for VLBW babies admitted to an Australian NICU. They collected information on characteristics of the babies (including required level of care), maximum number of babies per shift, and nurse staffing for the 72 hours subsequent to each admission. The infant to nurse ratio varied from 1.16 to 1.97.

3.4.3 Medical staffing

Goodman et al. (74) investigated the regional supply of neonatologists in 246 neonatal 'intensive care regions' in 1995. These regions were based on travel patterns of mothers of LBW babies from county of residence to birth, with the aim that travel or transfer outside of the regions was very infrequent and so could not bias results. The study included the equivalent of 2407 full time neonatologists who were not found to be spread throughout the regions in relation to number of intensive care beds or need for intensive care (based on maternal and pregnancy risk factors). Supply within these regions ranged from 2.7 to 11.6.

3.4.4 Organisational culture

The first use of 'culture' to describe an organisation was by Dr Jacques in 1951 (171): *"the culture of the factory is its customary and traditional way of thinking and doing of things, which is shared to a greater or lesser degree by all its members, and which new members must learn, and at least partially accept, in order to be accepted into service in the firm..."* It has no standard definition, but is a holistic term including the values, beliefs, behaviours, expectations, assumptions, and norms which help explain how and why things are done the way they are in a specific organisation. Mahl et al. (172) used the Quality Improvement Implementation Survey (QIIS) which classifies organisational culture into four types (hierarchical, developmental, rational and group) to assess 18 Canadian NICU. They found that different units identified themselves to greater and

lesser degrees with the different categories, and that after adjustment, a higher group culture was associated with lower survival, whereas the opposite was true for hierarchical culture (in babies born <29 weeks).

3.4.5 Volume of patients

Watson et al. conducted a study using data from 2009-2011, from 165 neonatal units within the UK (98). They were investigating the effect of volume of patients (defined as annual number of care days at any level of care provided to babies born <33 weeks of gestation), compared to designation on outcomes. They defined 'high' volume as a unit within the top quartile, and by doing so found that 14 of 44 NICU were not classified as high volume, and 9 of 39 high volume units were not NICU. Chung et al. (73) examined 1997-2002 data from 167, level 2 and above neonatal units in the U.S. They found that the annual volume of babies born with VLBW in these units ranged from <10 to >100. Similarly, Rogowski et al. (173) reported mortality of VLBW babies born in 1995-2000, utilising Vermont Oxford Network data from 332 U.S. hospitals. The average hospital admitted ~80 VLBW babies per annum, with 10^{th} and 90^{th} percentile values of $\leq$25 and $\geq$153 admissions, respectively.

3.4.6 Care practices

As a relatively new specialty, much of neonatal medicine is not evidence based. However, there are some practices which do have strong evidence supporting them and are promoted by national guidelines (e.g., administration of antenatal steroids to pregnant women at risk of delivering a preterm baby between 24-34 weeks of gestation, and magnesium sulphate at below 30 weeks of gestation (174)). However, even in these, there is significant variation in practice between units. The 2018 National Neonatal Audit Programme report found that in UK neonatal units in 2017, administration of magnesium sulphate varied from 0-100%, and antenatal steroids from 64.3-100% (175).

Furthermore, there are many areas of practice where evidence indicating optimal practice is lacking, and expert opinion is divided. Ojha et al. looked at use of high flow nasal prong oxygen ($HFNPO_2$) in 44 level 2 and 3 UK neonatal units (LNU and NICU) (176). Of the 34 which used $HFNPO_2$, half did not have a guideline, and use varied from alternatives for nasal CPAP (77%), for weaning off CPAP (71%), and for post-extubation respiratory

support (53%). Regarding treatment of PDA in VLBW babies, Hagadorn et al. found that median rates of treatment with cyclooxygenase inhibitor (COXI) agents ranged from 20-79% and for ligation from 0-54% from 2010-2014 in 19 children's hospitals within the U.S. (177). COXI agents were started at a median age of two days (range 0-51), and for a median duration of 3 days (range 1-14). 5.6% of babies had ligation without prior treatment with COXI agents, compared to 22.0% that had both medical and surgical management. EUROPAIN (European Pain Audit In Neonates) was a prospective study looking at use of sedation and analgesia in 243 NICU in 18 European countries between 2012-2013 (178). Of these, 6 countries had national guidelines and 182 units (75%) had local guidelines for neonatal sedation/analgesia. Of the 6,680 enrolled babies, 34% were administered sedation/analgesia at least once. Of the babies that had been intubated and ventilated, 82% had received sedation/analgesia, of the babies requiring non-invasive ventilation (NIV), it was 18%, and of the babies not requiring ventilatory support, 9%. The most used drugs included opioids (morphine, fentanyl, sufentanil) sedatives/hypnotics (midazolam, chloral hydrate, phenobarbital) and general anaesthetics (ketamine and propofol). Similarly, there is variation in use of vasoactive drugs to treat hypotension, which itself has a variable definition among neonatologists and neonatal units (179, 180), and the indications for, and specific antibiotics used to treat various infective conditions commonly encountered in neonates (181), etc.

3.4.7 Outcomes

The Maternal, Newborn and Infant Clinical Outcome Review Programme (MNI-CORP), run by the MBRRACE-UK collaboration (Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK) (89) analyses data on all UK births and perinatal/neonatal deaths, from which crude hospital-based mortality rates are calculated (number of deaths divided by the number of total live births). To minimise bias these crude rates are 'stabilised' and 'adjusted.' Stabilisation involves allowing for fluctuations in mortality rates due to chance, which are more pronounced in smaller hospitals with lower numbers of deliveries. Adjustment considers maternal and neonatal risk factors (maternal age, socio-economic deprivation based on residence, ethnicity, gender, plurality, gestational age), which can result in increased mortality rates in areas of social deprivation, and in large hospitals that serve as referral units for high-risk pregnancies. Deaths due to congenital anomalies are also excluded. The 2021 report shows that when

comparing 2019 neonatal mortality rates for NICU with surgical provision, even after adjustment and stabilisation, the crude rate varied from 1.58 to 4.49/1,000 live births.

In Switzerland, Berger et al. (90) analysed centre to centre variability in the nine NICU that care for the most preterm babies. Prospective data was used to calculate centre-specific, risk-adjusted (for birthweight, gestational age, gender, plurality) survival rates for babies born between 23-25 weeks (group A, n=976), 26-28 weeks (group B, n=1,943), and 29-31 weeks (group C, n=3,399), excluding those with severe or lethal congenital anomalies. The adjusted OR for survival in each centre was compared with the average survival of all the other centres. Their results showed that for group A infants, three NICU had significantly higher adjusted OR for survival, and three had lower (range of OR from 0.24, 95% CI 0.15-0.36 to 2.73, 95% CI 1.72-4.34); for group B infants, four NICU had higher adjusted OR, and one had lower (range of OR from 0.34, 95% CI 0.26-0.46 to 4.26, 95% CI 1.96-9.23); and for group C infants, one NICU had lower adjusted OR for survival (OR 0.61, 95% CI 0.37-0.98). Across the groups (i.e., from 23-31 weeks), it was the same NICU that performed either better or worse than the average.

In Australia, Abdel-Latif et al. (91) also examined centre to centre variability for the same population of babies admitted to eight NICU in the New South Wales (NSW) and Australian Capital Territory (ACT) neonatal network. They used multiple logistic regression analysis to estimate probability of pre-discharge mortality, adjusting for confounding factors (not specified). They added these probabilities to obtain each NICU's expected death rate, and the actual number of deaths was divided by this for the risk-adjusted standardised mortality ratio (SMR). In turn, this was used to calculate the adjusted mortality rate for each NICU. For the 7212 babies born <32 weeks of gestation, actual mortality (5.3%-10.4%), expected mortality (6.7%-8.9%), SMR (0.77, 95% CI 0.61-0.97 – 1.21, 95% CI 0.98-1.46), and risk adjusted mortality (6.1%-9.6%), all varied between NICU.

Kusuda et al. investigated centre to centre variability in VLBW babies born in 2003 in 42 Japanese NICU (182). They found incidence of CLD ranged from 0-100% and IVH (any grade) was diagnosed in 0-42%. Lee et al. did the same for all babies admitted to 17 Canadian NICU in 1996-1997 (183). Out of all babies born <28 weeks, from 28-32 weeks, 33-37 weeks and >37 weeks of gestation, the percentage who had positive CSF

cultures ranged from 0-14%, 0-5%, 1-3% and 0-1% respectively, and the percentage who had seizures ranged from 0-24%, 0-9%, 1-10% and 3-23%. Vohr et al. examined variation in outcomes for ELBW babies born at 12 NICU of the U.S. National Institute of Child Health and Development (NICHD) Neonatal Research Network (93). At 18-22 months, incidence of cerebral palsy varied from 6-30%, visual impairment from 9-20%, hearing impairment from 0-28%, and hydrocephalus with a shunt from 0-8%.

3.4.8 Conclusion

The heterogeneity I have described between neonatal units of the same designation, even within the same healthcare system, exemplifies the importance of not just categorising units by designation when comparing outcomes, but to also consider the care they provide. This is of even greater importance for OptiPrem, since the patient population I am investigating is born and cared for in both LNU and NICU. The degree of heterogeneity between units of the same designation can be significant enough to overshadow the differences between units of different designations (as summarised in Section 3.1.1, Tables 1 and 2). This is the rationale for OptiPrem to include WS2, which will categorise units by the quality of care they provide babies born between 27-31 weeks of gestation, irrespective of their designation. In the next chapter I explore how quality of care can be defined and measured.

4 Quality of care

In WS2, I am interested in identifying and measuring the quality of care provided by neonatal units that care for babies born between 27-31 weeks of gestation. Preterm neonates represent a group for which quality of care can have a large impact. Their early care involves high short-term costs associated with neonatal intensive care but also potential long-term costs in terms of education, social and healthcare needs for those with disabilities (184-186). Neonatology is a relatively new specialty, in which much of the practice remains non-evidence based. Variation in the quality of care delivered is likely but measuring quality of care is challenging.

In this section I discuss the different ways healthcare and quality of care has been defined and categorised. This leads to a description of the methods by which measures of quality of care can be identified and chosen. Finally, I discuss the different methods of evaluating and validating quality of care measures.

4.1 What is 'quality of care'?

Donabedian described seven attributes which define quality of healthcare (187):

1. *Efficacy: the ability of care, at its best, to improve health*
2. *Effectiveness: the degree to which attainable health improvements are realized*
3. *Efficiency: the ability to obtain the greatest health improvement at the lowest cost*
4. *Optimality: the most advantageous balancing of costs and benefits*
5. *Acceptability: conformity to patient preferences regarding accessibility, the patient-practitioner relation, the amenities, the effects of care, and the cost of care*
6. *Legitimacy: conformity to social preferences concerning all of the above*
7. *Equity: fairness in the distribution of care and its effects on health'*

The U.S. Institute of Medicine defines quality of care as the *'degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge,'* and suggest that healthcare should be *'effective, safe, patient-centred, timely, efficient, equitable'* (188). Campbell et al. defined it by asking two questions: *'first, can an individual get the care they need when they need it? Second, when they get the care, is it effective both in terms*

of clinical effectiveness and inter-personal relationships?' (189). Brook et al. argued that all definitions contained the following components; *'providing care of high technical quality,'* treating patients in a *'humane and culturally appropriate manner,'* and to invite them *'to participate fully in deciding about their therapy'* (190). Such definitions describe the aims of healthcare institutions to provide good quality care and are used as such by modern-day institutions. For example the National Health Service (NHS) England has a constitution with seven principles (191):

1. To provide a comprehensive service available to all
2. For access to services based on clinical need not ability to pay
3. To aspire to the highest standards of excellence and professionalism
4. For the patient to be at the heart of everything the NHS does
5. To work across organisational boundaries
6. To provide best value for taxpayer's money
7. To be accountable to the public communities and patients that it serves

4.2 What is healthcare?

All these definitions share similar principles in describing the care that we would each like to deliver as healthcare professions and organisations, and receive as individual patients and the general public. Before moving onto descriptions of how to measure quality of care, it is useful to have a structured way of defining or understanding healthcare. The framework introduced by Donabedian in the 1960s, of three domains or categories of care (structure, process, and outcome) is ubiquitous (Figure 10) (192).

4.2.1 Structure

Structure refers to the setting in which healthcare takes place. It can be broadly divided into two categories, physical and staff characteristics. Physical characteristics include equipment (e.g., availability of MRI scanners, ventilators, incubators) and buildings (e.g., number of intensive care beds, density of hospitals in a geographical area, regionalisation of perinatal care). Staff characteristics include staff type and training (e.g., nurses with specialist qualifications, specialist and allied healthcare staff, such as paediatric surgeons, radiologists, physiotherapists), and numbers (e.g., nurse to patient ratios, number of surgeons and anaesthetists).

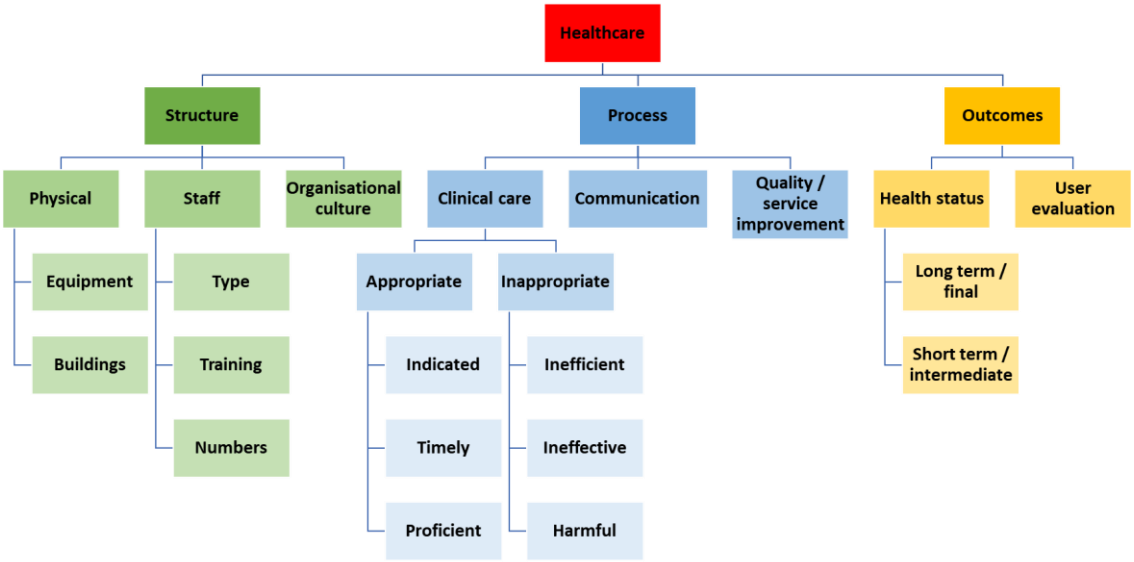


Figure 10 Aspects of healthcare described using classification into structure, process and outcomes

Organisational culture (as described in Section 3.4.4) is also part of healthcare structure and plays an important role in the ability of an organisation to deliver high quality care, thereby indirectly affecting outcomes (193). While it can be difficult to find evidence of this relationship, it is recognised that to produce long-lasting improvements in the quality of care an organisation provides requires concerted efforts to modify structure. For example, inquiries into the quality of care delivered at the Mid Staffordshire NHS Foundation Trust, Morecambe Bay, and Shrewsbury and Telford, have all highlighted the need for a change in the culture of healthcare staff (194-196).

4.2.2 Process

Process refers to the components of healthcare that are delivered to the patient and the level of proficiency or technical expertise with which they are delivered. To put it in simpler terms; what is done to the patient and how well it is done. This can also be divided into two categories, clinical care, and communication. Both may be appropriate (indicated, timely, proficient) or inappropriate (not delivered when indicated or delivered when not indicated, delayed, lacking required level of technical skill). Campbell et al. divided healthcare into prophylaxis, acute care, and chronic disorders, arguing that the two categories of process can be applied to each (189). 'Inappropriate' healthcare can also be categorised as overuse (providing healthcare to a recipient for whom the risks outweigh the benefits), underuse (failure to provide healthcare to a recipient for whom

the benefits outweigh the risks), and misuse (inappropriate provision of healthcare resulting in lack of benefit or harm) (197).

Clinical studies can reveal relationships between process and outcomes, from causal (randomised control trials and their meta-analyses) to associations (case control, cohort and observational studies). Processes supported with strong evidence are often incorporated into clinical guidelines, e.g. time to administration of first dose of antibiotics in septic patients (198), door-to-needle time for administration of tissue plasminogen activator (TPA) in acute ischaemic stroke patients (199), door-to-balloon time in patients with ST-elevation myocardial infarction (200). However, there may be processes of care which are associated with improved outcomes but for which evidence is lacking, e.g., due to rarity of the condition, or already widespread use of a treatment and so lack of equipoise to be able to carry out a randomised trial. The evidence may even seem contradictory, e.g., due to confounding by indication, where sicker patients receive better care but have poorer outcomes.

Because healthcare staff can directly impact delivery of care processes, when there is belief that certain processes will improve outcomes, there can be a greater perceived responsibility to meeting standards. However, focussing on specific care processes (which may be dictated by ease of measurement or availability of data) risks losing sight of the broader context in which healthcare is delivered, which can be to the detriment of other important processes.

Within this category, I would also include quality or service improvement. This can be described as a manifestation of organisational culture in process. It requires multidisciplinary team working, effective communication and organisation, and a desire to improve service provision and patient outcomes. For example, units seeking to improve patient outcomes may be more actively involved in auditing their care and monitoring outcomes. They may implement ‘care bundles’ or ‘packages of care’ – combinations of evidence-based measures which all impact on a specific outcome. Similarly, units may implement risk identification strategies to increase interdisciplinary communication enabling pro-action (e.g., regular discussions between obstetricians and neonatologists regarding impending delivery of preterm babies and those with congenital disabilities).

4.2.3 Outcomes

Outcome refers to the results or consequences of interaction with the structure and processes of a healthcare system. This can be divided into '*health status and user evaluation*'. The latter involves assessing how satisfied the patient is with the care they received, which can impact subsequent interaction with healthcare (189), as well as perceptions of healthcare staff regarding the quality of care they are providing (further discussed in Section 9.7.6). Health status includes morbidity and mortality. Outcomes can also be classified as short-term or intermediate (e.g., blood pressure in hypertensive patients, HbA1c in diabetic patients), and long-term or final (e.g., paralysis, death, blindness). Intermediate outcomes are useful to evaluate care quality and disease progression in the intervening timeperiod between treatment and long-term outcomes, which can be many years or even decades.

This long timeframe can make it difficult to use outcomes as care quality measures to feedback and impact the clinical care that is currently being provided. Just because a suboptimal or difference in outcomes has been identified it does not tell us how to change the structure or processes of the healthcare system or individual to ameliorate this. Furthermore, there are many factors other than care provided which will affect patient outcomes, e.g., illness severity at time of presentation, natural history of the condition, age, health status, and socioeconomic background of the patient, adherence with treatment, etc. Unless all such confounding factors are adjusted for (which arguably is impossible, hence the value of randomisation to balance unknown confounding factors), using outcomes to make comparisons between healthcare systems or individuals is not fair. The differences found could be due to factors other than the quality of care provided. However, especially for the public, outcomes are the most important indicators of care quality.

Structure	Process	Outcome
• Midwives and neonatal nurses have received training regarding benefits of breastfeeding, especially for preterm babies • Breastfeeding support workers available to help teach mothers of newborns to express colostrum • Breastpumps and privacy available for mothers to express in hospital	• If colostrum is available, it is administered as oropharyngeal or buccal colostrum and to commence enteral feeds in newborn preterm babies (unless contraindicated) • A galactogogue is prescribed for mothers of preterm babies whose milk supply is decreasing • Once their babies are sufficiently mature, mothers are encouraged to provide kangaroo care, and put the baby to breast	• Proportion of babies born at less than 33 weeks gestation, receiving any of their own mother's milk at discharge to home from a neonatal unit[§]

Table 6 Example of how structure and process relate to a specific outcome
[§]A UK NNAP (National Neonatal Audit Programme) audit measure (201)

4.3 Choosing measures of quality of care (MQC)

The broad steps involved in formulating measures of quality of care (MQC) are summarised in Figure 11. The first step is considering the purpose of measuring quality of care (202). Is it to produce a local or national benchmark against which individual practitioners, units or hospitals can compare themselves, to incentivise those consistently providing below average quality of care to improve, and reward those providing excellent care? Is it to investigate how quality of care can be improved to tackle observed poor outcomes in a specific field of medicine or surgery? Is it an academic exercise to further knowledge of the process of MQC development, which would need testing prior to implementation in the real world?

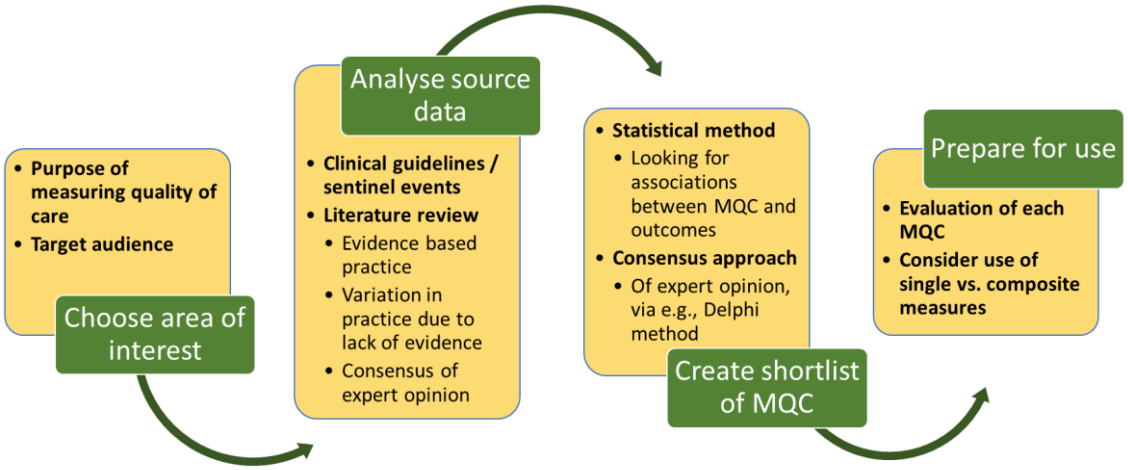


Figure 11 Schematic summarising steps in formulating measures of quality of care (MQC)

In a recent systematic review of 48 studies which extracted MQC from guidelines, criteria for selection of a topic included *'public health relevance, existence of a gap between potential and actually achieved quality, uncertainty about quality of care provided for a specific healthcare setting, economic impact of a specific healthcare problem, and individual impact on quality of life'* (203).

This is linked to the target audience for any new MQC (204). Is it for the public to be able to differentiate between and choose healthcare providers? Is it for the government or other healthcare purchasing organisations to guide investment? Is it for clinicians to evaluate their practice, or for units or hospitals to identify aspects of the structure of the healthcare they provide which could be improved? These considerations will guide which topics form the basis of the chosen MQC.

Once topics of interest are identified, specific MQC can be formulated. One method of doing this is to extract them from clinical guidelines (local, national, international), since these should be regularly updated to reflect current evidence on best practice (203). Similarly, sentinel events can be used, these are processes or outcomes which represent the worst level of care, also known as 'never events' in the NHS (205). This saves time and expense associated with development of new MQC and since they relate to guidelines that are already in place guiding clinical care patients are receiving, there is less need for pretesting prior to implementation. However, if development of the initial guidelines is not robust there is a risk of producing MQC which do not reflect actual quality of care delivered. Therefore, some form of critical appraisal is required.

Another method is to explore the scientific literature to find structures and processes with strong evidence linking them to outcomes. To find suitable outcomes (i.e., that can act as surrogate markers for quality of care received), they should depend largely on known structure/processes of care, occur relatively frequently, soon enough after the care episode and for which all relevant confounding factors can be adjusted for. Processes of care can also be used where there is known to be significant variation in practice. This is often because there is a lack of strong evidence indicating optimal practice and a lack of consensus amongst experts in the field. While the primary determinant for MQC is often the strength of supporting evidence linking to outcomes, there are many aspects of healthcare for which it is not possible to find such direct evidence; nonetheless they quite

obviously constitute good quality of care. For example, ensuring screening tests and patient follow-up are carried out within their designated timeframes. Such measures can be identified through consensus of expert opinion. Using scientific literature in conjunction with expert opinion, as opposed to guidelines, takes longer and can be more expensive but has the potential to produce more robust MQC.

Examining guidelines, scientific literature or clinical practice will identify many potential MQC which can form an initial list. This can be done either by the individual researchers or by employing a consensus method. This involves putting together a panel, which should include clinicians (considered experts in their field), academics (studying both in the area of interest but also quality of care and epidemiology), patient representatives, and health service officials (either at the national or regional level). This group should include representation from the target audience. The advantage of using a consensus method is to minimise potential selection bias, especially if pre-designed proforma with uniform rating and selection criteria are used to standardise the procedure (potential criteria discussed below) (206).

The initial list can then be analysed and narrowed down to produce a final list of MQC. This process can also be done one of two ways, using statistical methods or consensus approach. The statistical method involves searching for associations in the data between MQC in the initial list and outcomes, and only those for which a positive or negative association is found are put on the final list. This method is quick and ensures chosen MQC, by definition, impact outcomes but can lack face validity (especially among clinicians), since they may not reflect structure or process which are deemed clinically important. The consensus approach can use the same panel as for the earlier process. Depending on the amount and homogeneity of evidence, the expert clinical opinion of the panel can be used to augment the process of MQC development using a modified Delphi method (206-208). This involves multiple rounds of questionnaires, after each round the marks each item has received from the panel of clinical experts is averaged and an anonymised report is presented to the panel. The process is repeated, with the aim of narrowing down the chosen items from the list through the combined expertise of the panel. The identity of the panel members is meant to remain anonymous, and there is no face-to-face discussion during the process to avoid issues usually found in group activities where dominant or respected personalities are more likely to exert the major influence,

and panellists are unlikely to revise earlier decisions. As an extra measure of validity, a larger selection of clinicians within the relevant specialty can be asked to rate their level of agreement with the final choices of the Delphi panel (209).

4.4 Evaluating measures of quality of care

There should be predefined criteria by which the suitability of MQC are judged, and in the literature we find descriptions of various systems devised for this purpose.

QUALIFY (Instrument for the Assessment of Quality Indicators) was developed as such a method of assessing MQC (210). In 2005 researchers from the BQS Institute of Quality and Patient Safety (Germany) conducted a systematic literature search for articles in English, French and German relating to health care quality or similar terminology. From the 43 relevant articles in the 1128 identified by the search strategy, 208 criteria were extracted. These were grouped according to similar definitions or concepts, and then a process of '*pretest and refinement*' was carried out in which they were used to assess six MQC composed of varying processes and indications for processes, sentinel events and outcomes. This resulted in a final list of 20 criteria within three domains.

Category	**Criterion**
Relevance	Importance of the quality characteristic for patients and the healthcare system
	Benefit
	Consideration of potential risks/side effects
Scientific soundness	Indicator evidence
	Clarity of definitions
	Reliability
	Ability of statistical differentiation
	Risk adjustment
	Sensitivity
	Specificity
	Validity
Feasibility	Understandability and interpretability for patients and public
	Understandability for physicians and nurses
	Indicator expression can be influenced by providers
	Data availability
	Data collection effort
	Barriers for implementation considered
	Correctness of data can be verified
	Completeness of data can be verified
	Complete count of data sets can be verified

Table 7 Criteria to assess quality-of-care measures, extracted from Reiter, A., et al. "QUALIFY: Instrument for the assessment of quality indicators." Dusseldorf: Bundes Geschafts Stelle Qualitats Sicherung (2007) (210)

The U.S. National Quality Forum (NQF) does not develop MQC themselves, but other organisations can submit recommendations they have created, tested and validated which the NQF will assess according to their predefined criteria using a consensus method (211). Endorsement by the NQF is taken as a gold standard for use of the MQC by healthcare organisations and quality of care researchers, at least within the U.S. (212-214).

Conditions for consideration of MQC (must all be met)	The measure is in the public domain or a measure steward agreement is signed.
	The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.
	The intended use of the measure includes **both** accountability applications (including public reporting) **and** performance improvement to achieve high-quality, efficient healthcare.
	The measure is fully specified and tested for reliability and validity.
	The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.
	The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.
Criteria for evaluation (each measure to be assessed according to these criteria in order but measures must meet minimum requirements for first two criteria to be acceptable and assessed on other criteria)	Importance to measure and report
	Scientific acceptability of measure properties
	Feasibility
	Usability and use
	Related and competing measures

Table 8 Criteria to assess quality-of-care measures, extracted from https://www.qualityforum.org/measuring_performance/submitting_standards/measure_evaluation_criteria.aspx (211)

Hearnshaw et al. conducted a literature review (1990-1999) for articles relating to quality indicators, from which 40 characteristics were identified (215). An international panel of 38 academic and clinical experts in healthcare quality improvement was put together to engage in a two-round modified Delphi process. This resulted in a final list of 26 desirable characteristics that can be used as criteria to judge suitability of MQC (in order of importance).

Criteria are described in unambiguous terms
Criteria are based on a systematic review of research evidence
The validity of identified research is rigorously appraised
Criteria include clear definitions of the variables to be measured
Criteria explicitly state the patient populations to which they apply
Criteria are capable of differentiating between appropriate and inappropriate care
Criteria are linked to improving health outcomes for the care being reviewed
Criteria explicitly state the clinical settings to which they apply
The collection of information required for criteria-based review minimises demands on staff
The method of selecting criteria is described in enough detail to be repeated
Criteria are accompanied by clear instructions for their use in reviewing care
The systematic review used to guide the selection of criteria is up to date
Criteria are pilot tested for practical feasibility
Criteria include aspects of care that are relevant to patients
The collection of information for criteria-based review is acceptable to those patients whose care is being reviewed
The bibliographic sources used to identify research evidence are specified
In selecting criteria, decisions on trade-offs between outcomes from different treatment options are stated
The collection of information required for criteria-based review minimises demands on patients
The method of synthesising evidence and expert opinion is made explicit
Criteria are prioritised according to the quality of supporting evidence
Criteria are prioritised according to their impact on health outcomes
The criteria used to assess the validity of research are stated
Similar criteria should emerge if other groups review the same evidence
The collection of information for criteria-based review is acceptable to those staff whose care is being reviewed
Expert opinion is included in the process of developing review criteria
Criteria used in previous quality reviews of the same clinical topic are considered for exclusion

Table 9 Criteria to assess quality-of-care measures, extracted from Hearnshaw, H. M., et al. "Expert consensus on the desirable characteristics of review criteria for improvement of health care quality." BMJ Quality & Safety 10.3 (2001): 173-178 (215)

In 2006 the Organisation of Economic Co-operation and Development (OECD) examined national documents from the UK, Canada, Australia, USA, EHCI (European Community Health Indicators), Common-wealth Fund and WHO (World Health Organisation) to create a list of common criteria used to assess technical quality of MQC (216).

Dimensions	**Inclusion count**
Effectiveness or Improving health or Clinical focus	7
Accessibility	5
Patient-centredness or Patient focus or Responsiveness	5
Efficiency	4
Equity	4
Appropriateness	3
Competence or capability	3
Continuity	3
Safety	3
Acceptability	2

Timeliness	2
Capacity	1
Sustainability	1

Table 10 Criteria to assess quality-of-care measures, extracted from Kelley, Edward, and Jeremy Hurst. "Health care quality indicators project: conceptual framework paper." (2006) (216)

Examining these lists, I have identified common criteria that can be used to judge the suitability of MQC:

- Evidence of effect on outcomes
    - The MQC should relate to structure or processes of care for which there is strong evidence linking them to outcomes. If such evidence is lacking, consensus of expert opinion can be used.
    - The MQC may relate to structure or processes of care which do not directly impact on outcomes, but according to consensus of expert opinion still constitute good quality of care.
- Relevance for patients and/or healthcare system
    - MQC should target areas of healthcare of importance to individual patients. These are often areas where common or significant negative outcomes result from perceived poor quality of care.
    - MQC should target areas of healthcare of importance to the healthcare system as a whole. Along with the above, this will also include areas where there is significant expenditure, and it is believed cost-saving can occur without compromising care or outcomes (i.e., improving efficiency of care).
- Care providers can influence area of care being measured
    - It should be possible and feasible for healthcare providers to make changes to process or structure that will affect the area of care for which quality is being measured.
    - However, this depends on the purpose of measuring quality of care. If it is a descriptive exercise to document variation in care quality, then this is not a prerequisite. If the purpose is for quality improvement, then it is a necessity to assess this criterion.

- Reliability
    - This is a measure of how consistently the MQC measures quality of care, i.e., if the measurement is repeated, is the same result obtained? This includes intra- and inter-rater reliability, and internal consistency.
    - Intra-rater (also known as test-retest) reliability:
        - If the same parameter is measured on the same patient using the same test by the same investigator, are comparable results obtained? This confirms whether varying results from the MQC are due to the method of measurement itself rather than actual differences in quality of care.
    - Inter-rater reliability:
        - If the same parameter is measured on the same patient using the same test, but by different investigators, are comparable results obtained? This confirms whether varying results from the MQC are due to the way different investigators are conducting the measurements, rather than actual differences in quality of care.
    - Internal consistency:
        - We would expect multiple parameters that measure similar or interrelated aspects of care to provide similar results.
- Validity
    - Does the MQC measure what it is meant to measure? This is related to its sensitivity (what is the probability of the measure identifying a difference in care quality that exists) and specificity (what is the probability of the measure identifying a lack of difference in care quality, when one does not exist).
    - This can be answered by comparing results against a previously validated quality of care measurement tool. In the absence of this, face validity – based on consensus of expert opinion may be relied upon.
    - Careful consideration needs to be given to whether ‘gaming’ of the MQC can occur. This is where it can be made to seem as if a target is achieved, or it can actually be achieved, without the purpose of achieving that target being realised.

        - For example, a neonatal unit increasing their proportion of babies born at less than 33 weeks gestation receiving their own mother's milk at discharge by keeping some frozen, to be given on the day of discharge regardless of whether the mother has decided to stop breastfeeding by that time. This obviously defeats the purpose of measuring this parameter, which is meant to encourage units to strongly advocate the benefits of breastmilk, especially for babies born preterm, so mothers will continue breastfeeding post-discharge.
        - Gaming can be avoided by more careful construction of the MQC, or by addition of a parallel indicator, e.g., receipt of mother's breastmilk each day for last seven days prior to discharge, or continuation of breastfeeding at first clinic visit (210).
    - Validity will also depend on completeness and accuracy of data. For many processes of care this involves knowing the denominator (number of eligible patients) and numerator (number of patients receiving the process of care). When significant amounts of data are missing it is important to consider whether this is at random (rarely the case (217)) or not (e.g., units not actively involved in quality improvement projects not taking part in a survey on quality of care). In the case of the latter, this can result in introduction of bias (e.g., self-selection bias).
- Appropriate risk adjustment
    - Especially for outcome measures it is important all confounding factors are adequately adjusted for to ensure any comparison is fair, and differences identified are due to the indicator being measured rather than differences between comparator groups.
    - Care should be taken not to adjust for factors which can be influenced by care provided, since this can hide true differences between comparator groups.
- Unintended consequences / risk
    - This can occur due to false incentives being introduced by measuring certain indicators, e.g., reduction in rate of bowel wall perforation secondary to necrotising enterocolitis (NEC). In the early stages of NEC it is difficult to diagnose since the neonatal response is similar for many

disease processes. However, if the diagnosis is missed or delayed this increases risk of intestinal perforation, which significantly increases the risk of death. Such an indicator might lead surgical teams to intervene earlier, which would potentially expose more neonates who do not have NEC to major surgery. This can be avoided by having parallel indicators, e.g., confirmation of NEC on histology.

    - This criterion also includes consideration of whether pursuing improvement in the quality of care for this specific area is a better use of finite healthcare expenditure than alternatives.

- Feasibility
    - Data entry and collection
        - Are there systems already in place, or can they be introduced to allow easy, accurate and complete entry of data relating to the MQC?
        - How will this data be collated and accessed in a way compliant with individual organisations' information governance policies and national legal requirements for confidentiality regarding patient information?
    - Time and cost
        - Will the time and cost of the above be prohibitive? If the required data is part of routine care, and is recorded in a standardised format, perhaps via electronic means, this can greatly increase the speed and ease with which it can be collected.

4.5 Single vs. composite measures of quality of care

The scores of individual MQC can be combined to form an aggregate, i.e., a composite indicator. This can be used as an indication of overall quality of care compared to individual MQC which focus on one specific area and may not be representative of the whole, since performance in one process or outcome is often not correlated with performance in others (218, 219). An advantage of this is that to improve the composite score requires a healthcare provider to improve the quality of care provided in several areas. This 'system-based' improvement is desirable over improvement in a single area,

potentially at the cost of care quality in another area, which can be promoted by presenting MQC data separately (220).

The process of combining scores requires careful consideration. First, the scores of different MQC need to be on the same scale. If they are not, the method used to convert them can significantly impact results. For example, banding a continuous measure into categories creates larger differences between value at the highest and lowest borders of consecutive bands than exists in reality (221). The basis for such conversions is also often arbitrary, whereas it should have some statistical or clinical basis (e.g., based on local or national targets).

Once the scores for all MQC are on the same scale, a decision needs to be made regarding the method of aggregation. The simplest is to use addition, however this means that poor performance in one area can be hidden or compensated for by good performance in other areas. This can be partially counteracted by providing scores for each MQC alongside their aggregate score so any individual or organisation utilising the score can determine if this is the case. Another option is to differentially weight MQC depending on how 'important' they are, thereby avoiding poor performance in an important MQC from getting masked by good performance in a less important MQC. However, even if a consensus method is used to determine weighting (including which criteria to use and the value of each weight), this introduces another potential point of disagreement which can hinder engagement with the quality measurement process. However, equal weighting (due to avoiding allocating different weighting) is not a default position and must also be justified lest it face the same criticisms.

Other methods of aggregating include multiplying scores of individual MQC together to form the aggregate, and an all-or-none approach. The latter involves specifying a set standard of care across the different MQC (e.g., whether a process was carried out, or an outcome was achieved), and scoring them in a dichotomous manner (e.g., pass or fail). Only if the healthcare provider achieves the set standard in all the MQC making up the composite are they classified as providing good quality care (222). Proponents of the 'all-or-none' approach argue that this is the only system which does not allow providers to achieve good scores if care in a specific area is sub-standard, and as such is a better promoter for a systems-wide approach to improving quality of care. However, this is a

very harsh system, giving the lowest composite scores compared to other scoring systems, even for healthcare providers who would otherwise be thought of as providing good quality of care (223). If an all-or-none system is used, the less MQC used the higher the likelihood that providers of good quality of care will not 'fail', and these should relate to care practices or outcomes that are considered the most important or necessary. If the MQC relate to processes of care, they can be interrelated, i.e., steps leading to a specific outcome that requires all of them to be performed to occur. In such circumstances, anything other than an all-or-nothing approach is misleading since performance of a single process does not guarantee the outcome.

Aside from an explanation of the technical methods used to construct the composite score, the conceptual justification for combining multiple MQC should be made clear. This should not just be based on availability of measured data but reflect the thought process by which the composite score is purported to reflect overall quality of care and is superior to using the individual MQC which form it.

4.6 Conclusion

Having discussed the general concepts and principles by which quality of care can be defined and measured, in the subsequent section I discuss the first method I developed for collecting data on the quality of care provided by LNU and NICU.

5 Designing and piloting a questionnaire to collect data on quality of care

The first method I developed for gathering information on the quality of care provided by neonatal units was a questionnaire. Ethical approval had already been gained for this in the original OptiPrem protocol (Section 1.1). In this chapter I describe how the questionnaire was designed, finalised, and piloted. I analyse the results of the pilot and discuss what I learned from this process, how it could be improved upon, but ultimately why it was an unsuitable method for collecting data on quality of care for the purposes of this PhD.

5.1 Developing the questionnaire

5.1.1 Aim

Development of the questionnaire would be a two-stage process:

1. To choose questions, the answers of which would serve as measures of quality of care (to categorise units into groups).
2. To pilot the questionnaire to understand the feasibility of collecting data using this method.

If at the piloting stage significant problems became apparent, this would allow me to develop an alternate method for collecting data on quality of care. If I decided to continue with the questionnaire, the feedback received from units involved in the pilot, and my experience administering it would help develop the questionnaire further.

5.1.2 Methods

To produce the initial list of questions I used clinical guidelines produced by the 'Bedside Clinical Guidelines Partnership' in association with Staffordshire, Shropshire and Black Country Newborn and Maternity Network, and Southern West Midlands Maternity and Newborn Network (a group of 24 NHS Trusts, with over 60 contributing neonatal doctors and nurses) (224). These guidelines are updated on a three-yearly basis, and *"have been drafted with reference to published medical literature and amended after extensive consultation. Wherever possible, the recommendations made are evidence based. Where no clear evidence has been identified from published literature the advice given*

represents the consensus of the expert authors and their peers and is based on their practical experience."

Ideally a consensus method would have been used to produce the initial list of questions from the guidelines; however due to time and financial constraints, and because at this stage the questionnaire was only one of the methods of data collection being considered, this process was undertaken by me.

The main criterion was whether a structure or process of care had strong evidence of impact on outcomes, often in the form of national guidance from BAPM / NICE. I would use these to separate out units into two categories based on adherence. I would then look for association with outcomes, expecting to find units with better adherence to also have better outcomes. I also included questions interrogating practice for which there was a lack of strong evidence and national guidance and known variation between units. These would also be used to categorise units into two groups, and analysis would be undertaken to see if there was an association with a difference in outcomes between groups.

The original OptiPrem outcomes of interest were mortality, chronic lung disease, line sepsis, receipt of breast milk on discharge, and length of stay in hospital. Originally, a prospective assessment of neurodevelopment at 1 year of age was also planned, however this was not feasible due to financial constraints. Therefore, earlier versions of the questionnaire contained questions regarding neurodevelopment which were later excluded (Appendix II).

Apart from the quality improvement question, the rest were of multiple-choice format to allow me to compare answers and use them to group units. Regarding quality improvement, I have previously discussed how it can be described as a manifestation of organisational culture in process (Section 4.2.2). To use this to compare units I required a standard definition that healthcare staff would understand. I was trying to find out how proactive units are in identifying and improving patient outcomes. Therefore, I used a comparable term often used in healthcare: the 'ethos' or 'philosophy' of a unit (with an explanation of *'what their priorities are, what they pride themselves on, what they feel they do better than others'*). I also provided some examples of categories of care processes, e.g., identification of clinical practice or outcomes requiring improvement, use

of care bundles, encouragement of quality improvement projects, implementation of risk identification strategies, and regular use of simulation teaching sessions. However, given I could not identify or list all possible processes units use to improve outcomes, there needed to be an open-ended component to this question. This is a hallmark of qualitative research methodology, in which questions are more exploratory, allowing the respondent to play a leading role in identifying key aspects of the phenomenon being studied. Researchers are not looking to fit answers into predefined categories based on current understanding (*a priori*), but instead increase their understanding of the subject through experience and empirical evidence (*a posteriori*) (225). Therefore, my question interrogating quality improvement necessitated a discussion with the respondent via telephone interview. This semi-qualitative approach, combining an open-ended discussion with several predefined examples, would sufficiently capture information regarding quality improvement units looking after babies born between 27-31 weeks were involved in (Appendix II), and was devised in consultation with social scientists working on Workstream 4 of OptiPrem.

I would also need to develop a method to categorise units based on answers to this question, but in keeping with a qualitative approach this could not be predefined (since that risks applying the researcher's prejudices onto the data) and would instead be based on analysis of the answers for common themes that emerged. In general, I would be looking to categorise units into two or more groups based on how proactive they are in seeking to improve patient outcomes. I would expect to find an association between units that are more proactive and better patient outcomes.

The initial version of the questionnaire required considerable shortening to avoid discouraging units from taking part in the pilot. Each question was analysed by TP, EMB and I to determine:

- Whether the structure or process it was interrogating was suitable for inclusion, based on:
    - Convincing evidence of an 'optimal' practice that affects outcomes, or
    - Variation in practice associated with lack of evidence but which could potentially lead to differences in my outcomes of interest

- Whether how the question is posed, and its available answers:
    - Were easy to understand (i.e., not confusing, vague, or open to misinterpretation), and
    - Would capture significant difference in practice and not minor variation which would be difficult to use for categorising NNU

Table 49 (Appendix II) shows which questions were refined or removed during development of the questionnaire, and for what reason.

5.1.2.1 Piloting the questionnaire

I chose 20 NNU at random (using an online random number generator) from those participating in OptiPrem (ten LNU and NICU each). I excluded units in which I had previously worked, since I assumed it would be easier to obtain answers from them and so not representative of units in general. I also chose units which had and had not chosen to submit their data to the CRN (NIHR Clinical Research Network) to recognise their activity within the CRN research local and national research portfolio (this process involves monthly submission of data for numbers of research participants; in this case, babies born between 27-31 weeks). This was because it may have been more difficult getting engagement for the questionnaire from those units, which were not willing to participate in the CRN accrual process.

I sent an email to the clinical director of the units chosen for the pilot (Appendix II). I requested a copy of their unit guidelines (which would be used to answer as many questions as possible) and included a copy of the questionnaire. This was to be given to an allocated senior trainee or consultant, to familiarise themselves with prior to arranging a suitable date and time to have a telephone conversation (226). This would allow verification of questions answered using the guidelines, completion of any unanswered questions, and the discussion regarding 'ethos'. By agreeing to take part in the telephone conversation, consent was implied.

For the pilot, handwritten notes would be made of the telephone conversation. Following the pilot, if I decided to use the questionnaire as the main source of data collection for Workstream 2, I would record the telephone conversations and transcribe them, having asked for permission to do so in my correspondence with each unit. If permission was not

granted, I would instead make notes and send them to whomever I was having the conversation with, to verify the accuracy of their contents.

Because all, apart from the 'ethos' question were objective (designed to interrogate policies and practices), an assumption was made that inter-rater reliability would be high, allowing me to only fill in one questionnaire per unit (providing my respondent was relatively senior and knowledgeable of or had access to clinical guidelines). Although, for the purposes of the pilot I could have requested multiple healthcare personnel from each unit to answer the questionnaire to allow me to compare answers (e.g., senior and junior doctors and nurses), this would not have been feasible on a larger scale and so was not done.

5.2 Pilot results

In the following sections, I discuss the results of the questionnaire pilot, in terms of its primary purpose, which was to understand the feasibility of using this method to collect data on quality of care provided by neonatal units.

Over the four months the pilot was conducted (mid-July 2018 to mid-November 2018), I received replies from seven NNU. Four separate contacts were made with units during these four months with an aim to increase the number of respondents (at the start and end of August, mid-September, and second week of October). The first three of these were via email, the last via a phone call to the neonatal secretary who was sent the email to forward onto the clinical director to ensure they had received it. Of the seven respondents, three were LNU, four were NICU. At the time of initial contact, four of the seven units were not participating in CRN accruals. In all units, the clinical director (i.e., the point of contact) chose to engage with the pilot themselves. The number of times the respondent in the seven units was contacted varied from one to six (this included the group emails sent to the 20 units chosen for the pilot, as well as personal communication to serve as reminders if the respondent had indicated in an earlier correspondence they were willing to participate but I had not received a timely reply to my subsequent emails). The mode number was one. The time taken to complete the questionnaire (spanning from first contact to when the telephone conversation occurred or completed questionnaire was returned) varied from 5 to 85 days (mean of 46). Three of the units provided their clinical

guidelines as requested, in advance of a telephone conversation. Two units opted to forego this step and complete the entire questionnaire via the telephone conversation. The remaining two completed the questionnaire and emailed it back by themselves. For the five units that engaged in a telephone conversation, the duration ranged from 11 to 45 minutes, with an average duration of 25 minutes (of which 11 were spent on the final questions regarding unit ethos). Sample sizes were too small to investigate associations between whether clinical guidelines were provided in advance to the telephone conversation and its duration.

I received one incomplete questionnaire. This was from a NICU in which the questionnaire was allocated to a neonatal research nurse who was unable to complete all the questions. Both this nurse and I requested information from the clinical director for the unanswered questions (including the question on unit ethos), but we did not receive a reply. Of the remaining 12 NNU I did not receive any correspondence apart from two. The clinical director of an LNU agreed to take part but did not reply to any subsequent correspondence. The clinical director of another LNU replied to say they were too busy to take part.

5.3 Analysis of results

Despite repeated contact over the four-month period, my response rate was 35%, which would be insufficient to obtain a representative sample of NNU; it would need to be at least double this. Within the 20 units randomly selected for my pilot were equal numbers of NICU and LNU, and this split was roughly represented in my seven respondents. However, if the questionnaire was employed on a national level, given the low response rate it is probable the composition of respondents would not match the proportion of NICU and LNU in the country. Furthermore, there is significant risk of self-selection bias. Some, especially higher-volume and more geographically remote LNU believe they are able to provide as good, if not better care than their network NICU (personal communication), but since regionalisation, are no longer allowed to care for babies <27 weeks or those requiring long-term intensive care. Such units might be expected to engage more readily, especially with the unit ethos question in which they can describe their quality improvement projects.

The unit ethos question was answered ‘well’ by the five respondents who engaged in a telephone conversation (in that the answers contained information the question was designed to ascertain), giving multiple examples. For example, an LNU that lost a significant number of its senior nurses post-regionalisation, now uses trained healthcare assistants who are valued by the parents, nursing and medical staff. A NICU has a community of ex-parents who engage in a biweekly support group, a social media group for parents with babies currently on the unit, and arrange fundraising activities for the unit. Another unit prides itself on a being a place people like to work, with nurses feeling valued and playing a strong role in managing the unit and doctors rating it within the top three for neonatal units in England on the General Medical Council trainees survey. An LNU sees itself as a family-focussed organisation, with good feedback from parents (and no complaints in the last three years), doubling of breastfeeding rates since a large quality improvement project ten years ago focussed on championing early expression of colostrum, and creation of an information resource for fathers of preterm babies to increase their engagement. They also pride themselves on a strong team working ethic (winning an international award), despite losing their paediatric trainees post regionalisation and restructuring their unit to provide care via consultants, advanced nurse practitioners and staff grade doctors. Units also highlighted areas that need improvement, e.g., use of care bundles, family centred care, communication between neonatal and obstetric teams. The two units that did not engage in a telephone conversation filled in the ethos question very briefly, one of which wrote ‘all of the above’, the other circled three of the five examples provided. It is possible, although unlikely their answer would have been as brief in a telephone conversation. This highlights the importance of a direct conversation in completing this question.

Other questions which worked well were those interrogating practice which is clearly dichotomised and ‘all or nothing’ (this is rarely ever the case in medicine, but beyond a degree of conventionality certain practices can be considered routine even if a case-by-case decision is still being made). For example, default mode of ventilation (‘pressure limited’ or ‘volume guaranteed’), whether they use probiotics (‘yes’ or ‘no’), whether feeds are stopped during blood transfusions (‘yes’ or ‘no’). However, questions relating to practices that are new and may not be widely accepted or implemented yet, or which are significantly affected by the specifics of the clinical situation, did not prove suitable to categorise units into groups. For example, four of the respondents were not able to

answer the question on delayed cord clamping, because whether or not it occurred depended on whether the obstetrician delivering the baby practised milking of the cord, the location of birth and the temperature of the room, and how aware the midwife/obstetrician and paediatrician were of the benefits of this practice for preterm babies. During the telephone conversation I was told units had guidelines on the practice, but anecdotally they knew these were only followed '*50% of the time*'. For some practices where I predicted this was going to be an issue, appropriate answers were provided (e.g., being able to choose from 'generally only continuous positive airway pressure (CPAP)', 'both CPAP and high flow', and 'generally only high flow' for the question on post-extubation ventilatory support). These issues highlight a fundamental problem with many questionnaire studies, their validity, i.e., what is being reported may not be the reality (227). In my case, there may be a clinical guideline relating to a specific practice but this does not mean it is followed, or there may be a difference in practice based on personal preference of the clinician in charge and so the answer obtained is not reflective of the unit as a whole.

Other questions which did not serve my purpose related to practices that only affect one type of unit. For example, use of diuretics or postnatal corticosteroids for long-term ventilated patients developing or with established chronic lung disease. These babies would, by definition, have been transferred to a NICU.

5.4 Discussion

Following the pilot, if I had decided to continue using a questionnaire method for data collection, there are several changes I would have made.

The first related to increasing the response rate. For the pilot I directly emailed the clinical directors of NICU and LNU, with two repeat emails serving as reminders. A month before the pilot closed, I contacted the neonatal secretaries of units that had not replied and sent them the email to forward onto the clinical director in case my emails were not reaching their inbox or were being ignored due to being from an external email address to their organisation. A copy of the questionnaire was included with all correspondence.

Non-response to questionnaire studies are due to a variety of factors, including lack of interest / perceived importance of the subject matter, perceived time it would take to complete (affected by number and type of questions), how busy the respondent is, and cost and/or effort in returning the questionnaire (228-230).

By definition, all clinical directors of units participating in the OptiPrem study should have known about the study and given its aim (optimising place of care for babies born between 27-31 weeks), it could be assumed to be an important subject for neonatologists. However, in practice, when conducting the telephone conversation several clinical directors needed reminding regarding the details of OptiPrem. However, all respondents I spoke to regarding the study showed enthusiasm in supporting OptiPrem and believed WS2 was worth investigating. Therefore, one method of increasing the questionnaire response rate may have been to first contact the clinical directors to thank them for taking part in the study. The same opportunity could be taken to remind them what the study was about. This would lead to a discussion of the findings of my systematic review which shows there is currently a lack of evidence regarding outcomes for these preterm babies based on place of birth or care, and so highlights the importance of conducting this research. Following this method of increasing 'buy-in' and building a rapport, details of WS2 and the questionnaire pilot could be introduced before asking if they were willing to take part. Given the probable length of this conversation, first contacting the neonatal secretary to arrange a suitable date and time when the clinical director had at least a half hour slot free would work better than calling directly, in case of catching them at a busy time.

As previously discussed, several measures were taken to minimise the length of the questionnaire, including removal of questions for which I could obtain data from the Neonatal Data Analysis Unit (NDAU – see Section 6.1), and making all, apart from the last question on unit ethos, multiple choice. Apart from the last question, the rest interrogated unit policies and practices and so did not involve judgement questions where personal opinion is required on choosing what is considered the 'right' option; these questions can often take longer to answer. I requested clinical guidelines in advance of the telephone conversation because I believed this would allow me to complete a significant proportion of the questionnaire, however this did not turn out to be the case. For two of the three respondents sending me their guidelines caused a considerable delay

since they were not in a simple PDF format and required involvement from other personnel. This could be avoided in future. Because the questionnaire was filled in by me, respondents did not need to expend any time or energy doing this, nor in returning the questionnaire to me physically or electronically.

The second change category relates to the final question regarding the ethos of units. Units that fill in the questionnaire by themselves and do not write very much for this question would need to be re-contacted to request a telephone conversation to find out if their answer is accurate regarding the degree of their unit's engagement in quality improvement. Subsequent versions of this question would remove the included examples. In answer to the stem question some respondents can provide several examples of quality improvement projects their units are involved in, which may or may not overlap with the examples provided. Other respondents will answer the question by responding specifically to the five examples provided, sometimes just indicating whether they do or do not do it. This detracts from the purpose of this question, which is to identify differences in how units engage with quality improvement with an aim to use this to categorise units into two broad groups. If respondents are not provided with any examples this may better achieve this aim. There is a risk that without the examples the question may not be understood uniformly but given it would be answered during a telephone conversation this could easily be dealt with. To make the answers more objective, units could be asked to provide examples of quality improvement projects from a defined period of time (e.g., the last three years).

Given the purpose of the pilot was to assess the practicalities associated with using a questionnaire to collect data on quality of care, I have not gone into detail regarding the other questions which interrogated specific processes of care. In general, a consensus method could be used to review the comprehensive list of questions derived from clinical guidelines to highlight any significant omissions, as well as being used to evaluate each question against the MQC judging criteria to arrive at a final set of questions.

However, having conducted the pilot, and as discussed above, I have gained insight into why certain questions worked well compared to others. I would avoid questions that relate to practices which only affect one type of unit since although this can identify variation in practice and help categorise those units into groups, this reduces my sample size

considerably. This will make it more difficult to identify significant associations with outcomes that are already relatively infrequent for babies born at 27-31 weeks. Questions relating to practices which are not firmly established (i.e., where practice among senior clinicians is not uniform, or significant variation exists depending on the clinical context), increases the likelihood of the answer not reflecting practice within the unit, and so should also be avoided.

5.5 Conclusion

In conclusion, the main finding of the pilot was that using a questionnaire to collect data on quality of care faced significant problems that would be difficult to overcome, and so I should use an alternate source of data. In the subsequent section I discuss using data from NDAU and the National Neonatal Audit Programme (NNAP).

6 Using data from the Neonatal Data Analysis Unit (NDAU) to measure quality of care

An alternative data source for assessing quality of care is the Neonatal Data Analysis Unit (NDAU), which collects national data on routine care provided to babies in neonatal units within the UK (Figure 12). In this chapter I describe how NDAU manages data collected by the National Neonatal Research Database (NNRD). The process by which measures of quality of care were chosen from the NNRD data available to OptiPrem and publicly available NNAP data, is discussed. This leads onto the specific research question being asked in this PhD. The chapter ends with specification of the patient demographics, unit characteristics and outcomes that I will analyse.

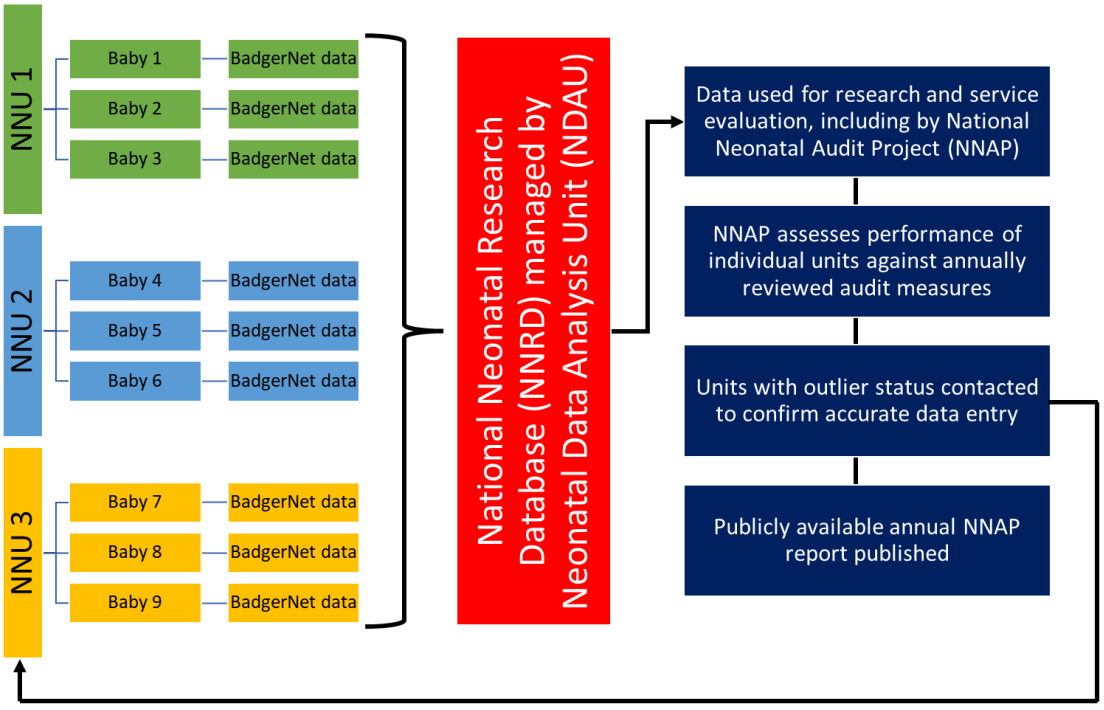


Figure 12 Schematic showing flow of data for individual patients from each neonatal unit to the National Neonatal Research Database (NNRD) and how the National Neonatal Audit Programme (NNAP) uses this data
NNU – neonatal unit

6.1 NDAU

Clevermed Ltd launched their electronic patient record system for neonatal units (BadgerNet) in 2004, which has since been adopted for use by nearly all units in England, Scotland and Wales. Details recorded for each baby include a predefined list of over 400

items called the Neonatal Dataset (NDS), which is an approved NHS Information Standard (ISB1595). This includes family demographics (e.g., parental age, address, occupation), pregnancy details (e.g., maternal health and pregnancy related complications, receipt of antenatal care, smoking and alcohol use), birth details (e.g., gestational age, birthweight, duration of rupture of membranes, congenital anomalies, Apgar scores, resuscitation required), daily care (e.g., ventilatory support, feeding method, antibiotics, radiological investigations, insertion of central lines), outcomes (e.g., retinopathy of prematurity, necrotising enterocolitis, infection, bronchopulmonary dysplasia, death), and two year neurodevelopmental follow-up. This data is inputted by the healthcare staff providing patient care (doctors and nurses).

Data is recorded from point of admission to discharge, or transfer to another place of care, and this is counted as a single episode of care. If the baby is transferred to another neonatal unit, and then back to the unit of birth (commonplace in a regionalised healthcare service), that would be three episodes of care, and for each baby these are linked by a unique BadgerID to provide an accurate chronological record of events from birth until final discharge destination (home, death, or to a place of care that does not use BadgerNet, e.g., paediatric ward, hospice). If babies are transferred between units, their electronic patient record can be accessed by the receiving unit, thereby allowing more seamless continuity of care. Locally, this data can be used for audit and quality improvement purposes. Individual patient data entered into BadgerNet is used to calculate reimbursement NHS Trusts receive for the neonatal care they provide by NHS England, thereby providing strong incentive for accurate recording of patient data.

Data from individual neonatal units is collated by NDAU, which was set up in 2007 (located at Chelsea and Westminster NHS Trust and managed by Imperial College, London). This is done on a quarterly basis, following which the data is cleaned and anonymised and used to construct the NNRD. Data quality is verified by examining patient population (using external data sources to ascertain number of eligible patients), completeness of data (e.g., demographics which are useful for risk-adjustment analyses), and accuracy (using consistency and range checks). During this process, if any potentially erroneous data is identified this is fed back to individual neonatal units to give them the opportunity to correct errors and complete missing data, which will be reanalysed in the next quarter. The clinical leads and Caldicott guardians of all 200 individual neonatal

units in England, Scotland and Wales have consented for their patient data to be used as part of the NNRD (and they form the UK Neonatal Collaborative). Therefore, the NNRD provides a national database of neonatal patients, containing data on over 1,000,000 babies and growing by roughly 100,000 annually (231). It has ethical approval (Research Ethics Committee reference 10/H0803/151) and can be used for national audits and service evaluation, assessing impact of regional quality improvement projects, and research (national and international).

Recently, a study was conducted to assess completeness and accuracy of NNRD data (2008-2014) (231). To assess completeness, a comparison was made with data collected by the Office of National Statistics (ONS), which has records on all livebirths. They found that by 2012, 100% of neonatal units in England were contributing data to the NNRD, and for >98% of their babies born between 25-33 weeks of gestation. The lower figures for babies born <25 weeks was assumed to be due to delivery room deaths, and for those born >33 weeks due to lack of involvement of the neonatal team for well, near-term babies. They also calculated the percentage of missing data for gestational age, sex, birthweight, antenatal steroids, mode of delivery, multiple birth and survival to discharge. For babies born <32 weeks of gestation (2012-2015), missing data for any of the seven parameters was <8%. To assess accuracy of NNRD data, a comparison was made with data for 1310 babies collected by the Probiotics in Preterm babies (PiPs) study; a regional multi-centre, blinded, randomised controlled trial (232). The PiPs trial recruited patients between 2010-2013 from 24 units in South-East England, collecting data using traditional Clinical Record Forms (CRF). NNRD data could be matched for 1258 babies for comparison of 44 item present in both databases (composed of demographic information, processes of care and outcomes). Using the PiPs data as the gold standard (i.e., assuming 100% accuracy), they looked at the degree of discordancy, with major discordancy defined as difference in binary items (e.g., whether gastrointestinal perforation occurred), or +/- $\geq$5 days difference for a continuous variable (e.g., antibiotic course duration). This was found for expected date of delivery, 5-minute Apgar, maternal ethnicity and LSOA (Lower Layer Super Output Area – small geographic areas linked to the Index of Multiple Deprivations 2010), mode of delivery, duration of high dependency care, central venous line in situ, antibiotic therapy, type of milk given on first day of milk feed, and receipt of supplementary oxygen on reaching 36 weeks postmenstrual age. Therefore, discordancy rates were <5% for 13/16 patient characteristics, 9/16 processes of care, and 10/11

outcomes. Considering outcomes, sensitivity of NNRD data was 50-100% and specificity was 86-100%.

6.1.1 National Neonatal Audit Programme (NNAP)

The Royal College of Paediatric and Child Health's (RCPCH) National Neonatal Audit Programme (NNAP) was setup in 2006, with an aim of assessing the quality of care provided within each UK neonatal unit. It is commissioned by the Healthcare Quality Improvement Partnership (HQIP) and funded by NHS England, and the Scottish and Welsh government. Quality of care is assessed by looking at adherence to annually reviewed standards or audit measures, which have been developed by a consensus method involving members of the NNAP project board, NDAU, and the wider neonatal community using the following criteria (201):

- *'Valid and accepted measures of a provider's quality of care*
- *Have clear relationships with quality of care*
- *Occur frequently enough to provide sufficient statistical power for analysis to identify outlying performance'*

'Adherence' to an audit measure may be determined by *'external sources, (research evidence, clinical judgment, audit data from elsewhere), or on internal sources, (such as average performance of all data providers to the audit, though may exclude the provider in question or outliers)'* (233). To determine this, the NNAP uses data from NDAU. If the data indicates a unit's performance is three standard deviations below the agreed upon standard, the data is validated after which the clinical lead is informed about potential outlier status. This gives them the opportunity to investigate whether there has been an error in data entry (incomplete, inaccurate or both), which will need rectifying prior to reanalysis by NDAU, or the outlier status is accurate, in which case the NNAP will notify the CQC (Care Quality Commission – an independent regulator of health and social care in England) or its equivalents in Scotland and Wales, and this information will be published in the publicly available annual NNAP report. The CQC will expect to see action plans of how individual units aims to address their outlier status, adherence to which will help avoid regulatory action (units with performance two standard deviations

below the standard are also listed as outliers, but do not go through the rest of the process described).

Therefore, for over a decade the NNAP audit measures have been used as a means of assessing quality of neonatal healthcare in the UK, by healthcare commissioners, providers and consumers.

6.2 Proposed measures of quality of care

6.2.1 Considerations in using data from the NNAP and NDAU

The OptiPrem study spans from 1st January 2014 until 31st December 2018. NDAU provided OptiPrem with data for babies born between 27-31 weeks of gestation for this time period. NNAP annual reports are available from 2014 to 2019 (from the publicly accessible RCPCH website: https://nnap.rcpch.ac.uk/annual-reports.aspx). However, for the purposes of WS2, in which I was interested in finding surrogate markers for measuring quality of care, it made sense to focus on the final year of the study – 2018. I did not plan to use a larger time scale since NNAP audit measures are reviewed annually and can and do change (i.e., in their definitions, or in terms of exclusion of old measures or inclusion of new measures). Furthermore, the healthcare environment of a neonatal unit can also vary quite significantly from one year to the next, e.g., with a different clinical director, or introduction of new local and/or national guidelines.

6.2.2 NDS variables to use as MQC

I chose a set of variables to use as measures of quality of care for this PhD from the Neonatal Dataset (NDS) (234). These were independent of the NNAP audit measures, and selection was based on three primary factors:

1. Availability of data
2. Presence of evidence linking variable to outcome(s)
    a. Since it was not feasible to use a Delphi approach to obtain consensus of expert opinion in identifying suitable measures of quality of care, by choosing measures with a strong evidence basis (see section 6.2.2.1), I can reasonably expect all units to be striving to provide this care (i.e., to be in agreement that this constitutes good quality of care)

3. Variables relating to process of care, structure or intermediate outcomes that occur in the peripartum period
    a. There is increasing recognition of the importance of providing optimal peripartum care to improve outcomes for preterm babies (235)
    b. Furthermore, since my analysis of outcomes would be by unit of birth, it made sense to look at aspects of care provided by those units, rather than care provided at a later time by which time the baby may have been transferred to a different unit

The list of variables I arrived at were:

1. Any dose of antenatal steroids given
2. Normal temperature measured within one hour of admission to the neonatal unit
3. Ratio/percentage of babies given non-invasive ventilation (NIV) out of all requiring ventilatory support on day one of life
4. Ratio/percentage of babies requiring intensive care (IC) provided with 1:1 nursing care on day one of life
5. Receipt of mother's milk on day 1 of life
6. Delayed cord clamping

From hereon I have referred to these as my non-NNAP MQC. Measures 1, 3, 5, 6 are processes of care; Measure 2 is an intermediate outcome and Measure 4 relates to structure.

This list of measures was vetted by EMB and TP and presented for further discussion to the OptiPrem study steering committee (SSC – composed of neonatologists, epidemiologists, statisticians, health economists, social scientists). They agreed with the chosen variables. The variables were also assessed (Appendix III, Table 50) using criteria for evaluating MQC as discussed in Section 4.4.

6.2.2.1 Evidence supporting choice of non-NNAP MQC

6.2.2.1.1 Antenatal steroids

From animal models we learn that exposure of the preterm fetus to corticosteroids accelerates pulmonary surfactant production, maturation of the alveolar blood gas

interface, parenchymal changes which enhance the structural integrity of the otherwise fragile preterm lungs, and causes cerebral vasoconstriction despite a hypercapnic challenge (236, 237).

Roberts et al. (238) conducted a Cochrane review in 2017, of randomised controlled trials comparing antenatal corticosteroid administration prior to preterm birth (spontaneous or elective) vs. placebo or no treatment. They included 30 studies (2 of which had low risk of bias across all domains) and 8158 infants (spanning birth at all gestational ages). Their meta-analyses found that treatment with antenatal steroids resulted in a significant reduction in:

- Perinatal death (n=6729; average RR 0.72, 95% CI 0.58-0.89)
- Neonatal death (n=7188; RR 0.69, 95% CI 0.59-0.81)
- RDS (n=7764; average RR 0.66, 95% CI 0.56-0.77)
- Moderate/severe RDS (n=1686; average RR 0.59, 95% CI 0.38-0.91)
- Need for mechanical ventilation (n=1368; RR 0.68, 95% CI 0.56-0.84)
- Duration of oxygen therapy (n=73; mean duration -2.86 days, 95% CI -5.51 to -0.21 days) and surfactant administration (n=3556; RR 0.68, 95% CI 0.51-0.90)
- IVH (n=6093; average RR 0.55, 95% CI 0.40-0.76)
- NEC (n=4702; RR 0.50, 95% CI 0.32-0.78)
- Systemic infections $\leq$48 hours of life (n=1753; RR 0.60, 95% CI 0.41-0.88)

There was no statistically significant difference in time requiring mechanical ventilation/CPAP (n=471; mean duration -1.91 days, 95% CI -4.59 to 0.76 days), air leak syndrome (n=2965; RR 0.76, 95% CI 0.32-1.80), and incidence of CLD (n=818; average RR 0.86, 95% CI 0.42-1.79).

Due to the strength of evidence, NICE guidelines specify that corticosteroids should be offered to all women between 24+0 and 33+6 weeks of pregnancy who are in '*suspected, diagnosed, or established preterm labour, are having a planned preterm birth or have P-PROM (prolonged preterm rupture of membranes)'* (174).

6.2.2.1.2 Delayed cord clamping

The definition of early versus delayed cord clamping is not uniform and includes time since birth (from 30 seconds to 5 minutes) or when cord pulsation ceases. Delaying clamping of the cord allows the baby to receive a placental blood transfusion due to uterine contractions and negative intrathoracic pressure as the baby takes its first breaths. Volumes of 80-100ml, or an additional 30% blood volume and 60% more red blood cells have been quoted (239-241). Especially in preterm babies, this results in increased haemodynamic stability in the immediate postnatal period (242), and in a small randomised trial on term babies was found to increase myelination at 4 months of age (243).

Fogarty et al. (244) conducted a systematic review and meta-analysis of studies comparing delayed versus early cord clamping for preterm babies of all gestational ages (18 studies, n=2834). Delayed clamping was defined as >30 seconds, and to prevent confounding they excluded studies in which they assessed >20% of infants in either arm to have had their cord milked or stripped. They only included randomised trials and performed their analysis by intention to treat. The meta-analysis showed delayed cord clamping was associated with a significant reduction in pre-discharge mortality (RR 0.68, 95% CI 0.52-0.90), with no heterogeneity and a symmetrical funnel plot. A sensitivity analysis using 9 studies at low risk of selection and attrition bias (judged as high quality), also showed a significant association between delayed cord clamping and reduced mortality (n=1233; RR 0.66, 95% CI 0.50-0.89). Delayed cord clamping was also associated with a significant increase in peak haematocrit and reduction in need for blood transfusion (RR 0.81, 95% CI 0.74-0.87), but did not affect incidence of CLD, IVH, NEC, or need for exchange transfusions (although the authors comment that these secondary analyses were underpowered, and amalgamation with future studies may alter results).

Due to this, and previous such evidence (245, 246), NICE guidelines and the Neonatal Life Support (NLS) course advocate delayed cord clamping for babies of all gestational ages who do not require immediate resuscitation or cutting of the cord for maternal reasons (174, 247, 248).

6.2.2.1.3 Normothermia after birth

At birth, babies transition from a warm wet environment to a cold dry one, and lose heat via evaporation, convection, conduction and radiation. Without adequate interventions core temperature can drop 2-3^{O}C within 30 minutes and there is a risk of hypothermia (249). In response to this, non-shivering thermogenesis takes place, which is metabolism of brown adipose tissue containing stores of fat and glycogen; heat being produced as a by-product of the biochemical reactions (250). When prolonged, in preterm babies who lack brown adipose tissue, and in babies with respiratory insufficiency, this leads to metabolic acidosis and hypoglycaemia (251, 252). Hypothermia induced pulmonary vasoconstriction can also lead to persistent pulmonary hypertension of the newborn (PPHN) (253). Preterm babies are at increased risk of heat loss at birth due to larger surface area to volume ratio, reduced ability for cutaneous vasoconstriction, and an immature skin barrier.

De Almeida et al. (254) conducted a multicentre (n=9) prospective cohort study of preterm babies (n=1764) born <34 weeks of gestation between 2010-2012 in Brazil. They defined hypothermia as axillary temperature <36.0^{O}C, and temperature was measured 5 minutes after birth and on NICU admission. Hypothermia at 5 minutes of age was strongly associated with hypothermia on NICU admission, which increased risk of early neonatal mortality (within 6 days of birth - OR 1.64, 95% CI 1.03-2.61).

Laptook et al. (255) conducted a similar multicentre (n=15) prospective cohort study, using multivariate logistic regression to determine associations between admission temperature and neonatal pre-discharge mortality for babies born <34 weeks of gestation (n=9031) between 2012-2013 in the U.S. Using WHO definition of normothermia (36.5-37.5^{O}C), they found decreasing odds of mortality with increasing admission temperature (OR 0.81, 95% CI 0.71-0.91) across both of their subgroups (babies born <29 weeks, and between 29-33 weeks).

Lyu et al. (256) carried out a retrospective observational multicentre study (n=29) of babies born <33 weeks (n=9833), admitted between 2010-2012 in Canada. Temperature was the first recording made within one hour of admission to NICU. Their primary outcome was a composite of mortality or any major neonatal morbidity, including grade III/IV IVH or PVL, ≥grade III ROP, ≥Bell grade II NEC, BPD, nosocomial infection

(defined as culture positive sepsis or meningitis at >48 hours of age). In their multivariate analysis in which admission temperature was used as a continuous variable and adjustment was made for confounding variables, a U-shaped relationship was found between admission temperature and the primary outcome, as well as duration of ventilation. The lowest rates of adverse outcomes occurred between 36.5-37.2^{O}C.

6.2.2.1.4 Early enteral feeding with breastmilk

Traditionally, enteral feeds were withheld from preterm babies for several days, especially in cases where it was assumed there was gut immaturity and so feeds would not only not be tolerated but would increase risk of developing NEC. This included extremely babies born extremely preterm and/or with IUGR and/or abnormal doppler studies of placental arteries. Feeds would also be withheld from babies suffering from significant RDS requiring ventilatory support. However, studies have shown that earlier introduction of enteral feeds with breastmilk in preterm babies promotes release of gastrin, gut motility, and establishment of gut microbiota (257-259). It results in quicker establishment of full enteral feeds (260), which reduces duration of parenteral nutrition and central venous catheters, with associated risks of cholestasis, catheter associated infection, and extravasation injuries (261).

Battersby et al. (123) conducted a prospective, whole population surveillance study between 2012-2013 (UK Neonatal Collaborative Necrotising Entercolitis - UKNC-NEC Study), on babies born <32 weeks of gestation, to look for associations between feeding practices and severe (Bell stage 2 or 3) or fatal NEC. Using a multivariate logistic regression model and propensity scoring, their analysis of 11939 babies found that commencement of enteral maternal milk feeds, with or without addition of bovine-origin products, in the first 7 days compared with after the 8th day of life, resulted in a relative risk of severe or fatal NEC of 0.69 (95% CI 0.60-0.78). Due to excluding Bell stage 1 NEC, this is likely an underestimation of the total effect size. 11523 (81.0%) were fed breastmilk during the first 2 days of life (9460 - 66.5% - their own mother's milk, 2063 - 14.5% - donor milk), but an analysis for this cohort compared to those fed after the 3rd day of life was not conducted.

The ADEPT (Abnormal Doppler Enteral Prescription Trial (260)) was a multicentre (n=54) randomised controlled trial comparing early (24-48 hours after birth) vs late (120-

144 hours after birth) introduction of enteral feeds to babies born <35 weeks who were small for gestational age (SGA) and had abnormal antenatal doppler studies indicative of IUGR. Feeds were increased according to birthweight, aiming to reach 150ml/kg/day over 9-13 days. From 2006-2009 they recruited 404 babies (400 required according to sample size calculation to show a 50% change in incidence of NEC with 60% power). Feeding commenced in the specified timeframes for 76-83%, and 78%-94% received an initial feed containing their mother's breastmilk. Their analysis revealed no difference in the incidence of any stage or severe NEC between groups, but earlier achievement of full enteral feeding in the early group (median age 18 days vs 21 days, Hazard Ratio 1.45, 95% CI 1.19-1.78 after adjustment for birthweight). Duration of parenteral nutrition and incidence of cholestasis were also significantly lower in the early feeding group.

6.2.2.1.5 Respiratory support at birth

Each breath a ventilator delivers risks overstretching of preterm alveoli (volutrauma) causing inflammation and contributing to development of chronic lung disease (CLD) (132). This is exacerbated by surfactant deficiency associated alveolar collapse at end expiration. With each breath, re-expansion causes shear stress (atelectrauma) and further inflammation. One method of reducing ventilation associated lung trauma is to minimise its use. This can be done through prophylactic use of non-invasive forms of ventilation (NIV), which provide end-expiratory pressure, thereby preventing alveolar collapse. In babies with sufficient respiratory drive, this can prevent the need for mechanical ventilation and surfactant.

Subramaniam et al. (262) conducted a Cochrane review examining evidence supporting use of nasal CPAP in a 'prophylactic' manner, i.e., within 15 minutes of birth regardless of the baby's respiratory status. Outcomes were need for mechanical ventilation, development of CLD, and mortality in very preterm (<32 weeks gestational age) and VLBW babies. Four studies looked at treatment failure (defined as need for assisted ventilation, rescue CPAP prior to mechanical ventilation, surfactant, or both) when comparing prophylactic CPAP to supportive care (e.g., supplemental oxygen by headbox or standard nasal cannula). Meta-analysis revealed a reduction in treatment failure (typical RR 0.66, 95% CI 0.45-0.98). Meta-analysis of three studies which compared prophylactic CPAP to mechanical ventilation revealed significant reduction in incidence

of CLD (n=2,150; typical RR 0.89, 95% CI 0.80-0.99) and combined outcome of CLD and death (n=2350; typical RR 0.89, 95% CI 0.81-0.97).

6.2.2.1.6 Nurse staffing

Evidence indicates a threshold of nurse-to-patient ratio is required for optimal outcomes. Hamilton et al. (168) used 1998-1999 data from the UK Neonatal Staffing Study (UKNSS), involving 54 units, 2636 babies (VLBW and/or <31 weeks of gestation at birth), and 35651 records of nursing shifts. They calculated the total number of registered nurses and expected number of nurses per shift. Values <1.0 indicated inadequate nurse staffing according to BAPM guidance at the time. They adjusted for unit volume, neonatal consultant availability, and created their own predictive mortality score for neonatal illness severity. Multivariate analysis showed that mortality was significantly linked to the ratio of specialist nurses per shift (OR 0.63, 95% CI 0.42-0.96). They found a reduction in odds of mortality of 48% when the ratio increased to 1.3-1.8 (compared to <1 – OR 0.52, 95% CI 0.33-0.83). This suggests that units with specialist nurse staffing at higher than the minimum requirements stipulated by guidelines at the time, increased odds of survival. Subsequent BAPM guidelines increased these minimum ratios for neonates requiring high dependency and intensive care.

Rogowski et al. (169) examined effects of nurse staffing on rates of nosocomial infection. The 2009 data was from 67 hospitals in the Vermont Oxford Network (VON), involving 3645 nurses and 8804 neonates. They calculated percentage of infants for whom understaffing occurred and number of nurses (in decimals) needed to meet guidelines. They found that 16-55% of units understaffed 25-50% of their patients, with only five adequately staffed units. In these units the predicted infection rate was 9%, but at the median understaffing level (0.89 nurse per infant), the predicted rate was 14%, rising to 21% at the 90th percentile for understaffing (0.78 nurse per infant). This meant that approximately one tenth of a nurse was associated with 40% higher odds of infection. This was despite this being a non-representative sample of NICU, containing proportionally more teaching hospitals and higher level NICU, as well as 40% having achieved nursing excellence awards (compared to 19% US average) – which will have biased results towards underestimating the effect of understaffing.

6.2.3 NNAP audit measures to use as MQC

Adherence with national guidance in the form of NNAP audit measures reflects organisational culture of neonatal units to improve outcomes and deliver good patient care. This involves both adherence with the audit measure in terms of fulfilling its requirements, and sufficient data completion to allow accurate monitoring of adherence level. Because these measures were to be used as surrogate markers, the data did not need to be specific to babies born between 27-31 weeks. It made more sense for it to relate to care provided to all babies, even if I was looking for associations with outcomes for babies born between 27-31 weeks of gestation. In this way, use of the NNAP audit measures fundamentally differed from analysis using my non-NNAP MQC, for which the data was specific to my cohort. It also therefore, made sense to use all NNAP audit measures unless excluded for specific reasons detailed below. 2018 data would be used (containing information regarding babies with a final discharge from neonatal care between 1st January 2018 to 31st December 2018) (263).

The complete list of audit measures available in the online annual report from the RCPCH website include (264):

1. *"Antenatal steroids*
    - *Are all mothers who deliver babies between 23 and 33 weeks gestation inclusive given any dose of antenatal steroids?*
2. *Magnesium sulphate*
    - *Is a mother who delivers a baby below 30 weeks gestational age given magnesium sulphate in the 24 hours prior to delivery?*
3. *Promoting normal temperature on admission for very preterm infants*
    - *What proportion of babies born at less than 32 weeks gestation, and who are admitted to a neonatal unit, have a first measured temperature of 36.5–37.5°C?*
4. *Consultation with parents*
    - *Is there a documented consultation with parents by a senior member of the neonatal team within 24 hours of a baby's first admission?*

5. *ROP screening*
    - *Does an admitted baby born weighing less than 1501g, or at gestational age of less than 32 weeks, undergo the first retinopathy of prematurity (ROP) screening in accordance with the NNAP interpretation of the current guideline recommendations?*
6. *Bloodstream infection*
    - *What proportion of babies have one or more episodes of bloodstream infection, characterised by one or more positive blood cultures taken after 72 hours of age?*
7. *Central line associated bloodstream infection*
    - *How many babies have a positive blood culture (any species) with a central line present, after the first 72 hours of life, per 1000 central line days?*
8. *Bronchopulmonary dysplasia*
    - *Does an admitted baby born at less than 32 weeks develop bronchopulmonary dysplasia (BPD)?*
9. *Mother's milk*
    - *Does a baby born at less than 33 weeks gestational age receive any of their own mother's milk at discharge to home from a neonatal unit?*
10. *Two-year follow-up*
    - *How many babies born at less than 30 weeks gestation received medical follow up at two years gestationally corrected age?*
11. *Parents on ward round*
    - *For all parents of babies with admissions of greater than 24 hours, did at least one parent attend a consultant ward round?*
12. *Necrotising enterocolitis*
    - *What proportion of live born babies born at less than 32 weeks gestation who were admitted to a neonatal unit met the NNAP surveillance definition for Necrotising enterocolitis (NEC) on one or more occasion?*

13. Minimising separation – term

- *For a baby born at gestational age greater than or equal to 37 weeks, who did not have any surgery or a transfer during any admission, how many special care or normal care days were provided when oxygen was not administered?*

14. Minimising separation - late preterm

- *For a baby born at 34-36 weeks gestational age, who did not have any surgery or a transfer during any admission, how many special care or normal care days were provided when oxygen was not administered?*

15. Nurse staffing on neonatal units

- *Measure one: What proportion of nursing shifts are numerically staffed according to guidelines and service specification?*
- *Measure two: What proportion of shifts staffed according to guidelines and service specification: qualification in speciality?*
- *Measure three: How many additional nursing shifts are required to be worked to meet guidelines and service specification?)"*

For the purposes of categorising units according to adherence, I excluded measures related to final outcome measures (mortality, NEC, BPD). Final outcome measures are not ideal for use as measures of quality of care. For a fair comparison between units, adjustment for all confounding factors at a patient level would be needed, and since the data was not from a randomised trial, there would be unknown confounders influencing outcomes. Furthermore, there is limited value in correlating one final outcome measure with another (i.e., units with higher incidence of NEC have higher mortality rates), with regards to providing units with advice on how to improve outcomes based on improving the quality of care provided. The exception to this was bloodstream infection, which can also be considered as an intermediate outcome measure relevant to other final outcomes measures (e.g., NEC, BPD, mortality).

6.2.3.1 Evidence supporting use of NNAP audit measures

I used engagement with national guidance (in the form of adherence with, and data completion for NNAP audit measures) to reflect organisational culture of neonatal units in improving outcomes and delivering good patient care.

There is evidence of association between aspects of organisational culture and healthcare outcomes. In the adult intensive care setting, units with lower mortality rates self-assessed their staff as being less dependent and more trusting, and their teams as more structured and organized than units with higher mortality rates (265). Huang et al. (266) found lower perceptions of management associated with increased mortality, and lower safety culture associated with increased length of hospital stay, also in an intensive care setting. Roch et al. (267) examined organisational factors which could affect nurses' perceptions of their ability to carry out caring practices for their patients. They found that high workload was the main factor in reducing time for caring practices, but team functioning (in terms of *'harmony, cooperation and role clarity'*) could help mitigate this. In the setting of paediatric primary care practices, '*group*' culture, valuing shared decision-making and teamwork (as opposed to focussing on rules, regulations, efficiency and authority in '*hierarchical*' or '*rational*' cultures), was associated with job satisfaction and perceived effectiveness for both clinicians and non-clinicians (268).

Within neonatal medicine, Pollack et al. (269) investigated whether there was a correlation with morbidity, mortality, length of stay and ventilator days for VLBW infants (n=522) cared for in eight NICU within Washington DC (adjusted for birthweight and SNAP score (270)). Questionnaires were completed by respiratory therapists, nurses and physicians. The score for nurses correlated with rates of intraventricular haemorrhage and periventricular leukomalacia, physicians with ROP, and respiratory therapists with mortality. A similar study from Canada (172) used the Quality Improvement Implementation Survey (QIIS) which classifies organizational culture into four types, hierarchical, developmental, rational and group. This was used at 18 NICU to look for associations with survival without major morbidity in babies born <29 weeks of gestation (n=1028). After adjusting for male gender, low 5 minute Apgar score, being born <26 weeks, SGA and high SNAP-II score (271), they found that higher group culture was associated with lower survival, whereas the opposite was true for hierarchical culture.

Yates et al. (272) examined the relationship between levels of missing data and performance in the context of The National Clinical Audit for Rheumatoid and Early Inflammatory Arthritis (NCAREIA) in 2014-2015. They found that of the 136 departments involved in the audit, 13 had high levels of missing data regarding disease

activity, and a statistically significant positive association was found between these departments and a delay in commencing treatment (OR 0.5, 95% CI 0.41-0.61, p<0.01).

Specifically, regarding data completion, there are many areas of medicine where it is clear good data completion aids, or is necessary in providing good quality of care, e.g., discharge letters of hospital patients (273). Without adequate and accurate details, physicians treating the patient in the community or on readmission to hospital will have an incomplete or incorrect history, thereby hampering their efforts to provide good quality of care. This is also the case for preterm babies in the UK for whom details of their stay (broken down by systems and including procedures and diagnoses) on BadgerNet (the electronic patient record) is used as their discharge letter. These details are also used by NHS England to calculate reimbursement for individual units, which impacts the future care they can provide. Furthermore, when units wish to conduct audits or quality improvement projects (on a regional or national level – as with the NNAP), it is often the same data stored on BadgerNet that is interrogated. Again, poor data completion will frustrate this process. When complaints/litigation occur, or reviews of patient care in morbidity and mortality meetings, an accurate and complete set of patient notes is required for the processes to fulfil their aims, which includes for the healthcare system as a whole to analyse and learn from previous care provided to help shape and improve future care. Therefore, while the relationship between data completion and quality of care might not be as direct as clinical practice and quality of care, it is nonetheless an important factor. This was the basis of including it as a surrogate marker to reflect the organisation culture of neonatal units in improving outcomes and delivering good patient care.

6.3 Specific research hypothesis and research question

My hypothesis was that by categorising units according to adherence with, and data completion for my non-NNAP MQC and NNAP audit measures, I anticipated finding associations with mortality and major morbidity outcomes for babies born between 27-31 weeks of gestation. **I.e., the question I sought to answer was *'do units with better adherence with, and data completion for my non-NNAP MQC and NNAP audit measures have better mortality and major morbidity outcomes for babies born between 27-31 weeks of gestation?'***

6.4 Demographics

I described my patient population using the below demographic and unit details. After categorising units by MQC, the demographic/unit profiles of comparator groups were compared to look for significant differences that could act as confounding factors when conducting analyses to look for associations with outcomes. For any such differences found, the need for further statistical analyses to adjust for confounding factors would be considered.

- Number of units and their designation
- Number of babies, average number of babies born per unit, and number of babies by each gestational week of birth
- Birthweight, gender, multiplicity, presence of major congenital anomalies
- Condition of baby at birth (cord base excess, Apgar score at 5 minutes of age, number of babies requiring resuscitation involving cardiac massage or adrenaline, worst base excess in first 24 hours of life)
- Socioeconomic factors (ethnic group and the index of multiple deprivation – IMD_Q)
- Health status of mother (pre-pregnancy, during pregnancy, and drug and alcohol use)

6.5 Outcomes

I would be looking for associations between my groupings of units and the following outcomes:

- Mortality (pre-discharge)
- Length of stay (LOS)

I had planned to look at severe morbidity outcomes (BPD, NEC requiring surgery, ROP grade III/IV and/or requiring treatment, neurological damage including IVH grade III/IV, PVL, porencephalic cysts, hydrocephalus), however, due to the impact of the Coronavirus pandemic, it was not possible to obtain this data in a timely fashion to allow this work to be done.

7 Methods

The following methods are described in this chapter:

- How I sorted the data supplied to OptiPrem by NDAU to arrive at my patient cohort.
- The process by which it was determined how many babies in each unit received the care specified by my non-NNAP MQC.
- The process to determine which of the NNAP audit measures were appropriate to interrogate individual unit compliance and data completion.
- The specific variables used to describe the demographic profile and unit characteristics of the comparison groups.
- The univariate and multivariate statistical tests used to look for associations with the pre-specified outcomes.

7.1 Modified OptiPrem dataset

From the NDS data NDAU provided to OptiPrem, I requested data for my chosen non-NNAP measures of quality of care (MQC) and demographic data, from the WS1 statistical team. The total cohort of babies was 29,703, born between 27+0 – 31+6 weeks of gestation, discharged from 1st January 2014 to 31st December 2018. 5638 of these babies were born in 2018 (Figure 13).

When sorting the data by demographics, there was a variable amount of missing data, ranging from <1% (e.g., for birthweight, gender) to 39% (e.g., for health problems prior to and during pregnancy). I did not exclude babies based on missing or obviously incorrect demographic data. This data is inputted for each baby by healthcare professionals on a day-to-day basis. It is not uncommon for certain data points to be missed out if the relevant information is not available at the time of data entry, especially on admission, or the unit is busy due to a sick baby and/or understaffing. Typographical errors can also occur when entering numbers and dates. This does not mean those babies did not exist or that all their data is invalid. Excluding all such babies would have significantly reduced the patient cohort and so this pragmatic approach was taken. The only exception was for five babies that had missing data for gestational age.

OptiPrem's data was for babies born between 27-31 weeks, who are meant to be born and cared for in LNU and NICU. However, there are occasions where birth occurs in SCU due to women presenting in preterm labour and deemed unsuitable for in-utero transfer. Such babies are transferred ex-utero, and so were present in the data since their second care episodes were in LNU or NICU. Since the variables I chose related to peripartum care up to the end of the first day of life, I excluded these babies since they would have received the majority or all of this care in a SCU. I was able to identify these babies by using the variable for level of unit of birth and excluding those born in SCU. For this reason, I also excluded babies whose data related to second or third care episodes (i.e., anything other than data for first care episodes). I confirmed the first recorded day of data was the first day of life for each baby by sorting the babies by number of minutes from birth, to recording of first daily data. Finally, I sorted the data by year of discharge, and excluded all except those discharged in 2018. The patient cohort I was left with was of 5038 babies.

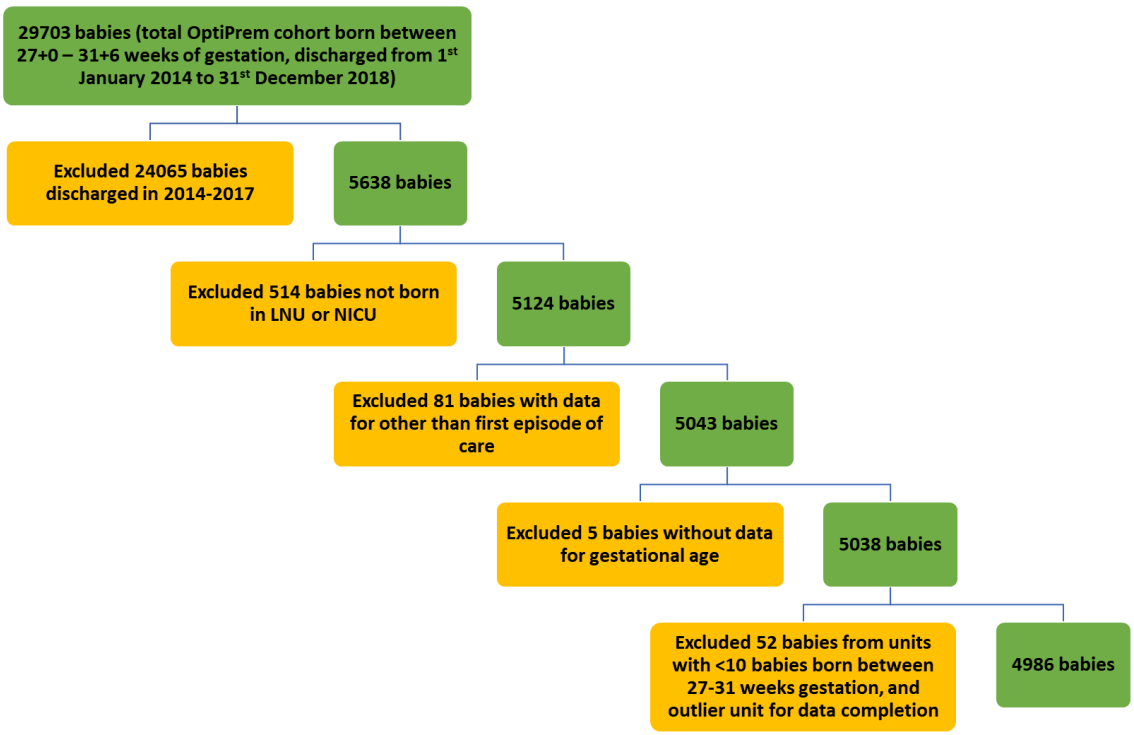


Figure 13 Flow diagram showing sequential exclusion of patients from OptiPrem patient cohort to arrive at cohort of patients used for WS2

For the next stage of data sorting for my chosen measures and drawing out the demographic data, I separated these babies by the 119 units they were born in. Of these, data for 5 units was excluded due to having <10 patients, and data for another unit

excluded due to having a high percentage of missing data and being an outlier. This left data for 4986 babies from 113 units (Figure 13).

7.1.1 Non-NNAP measures of quality of care

The six proposed non-NNAP measures of quality of care were:

- Measure 1 – any dose of antenatal steroids given
- Measure 2 – normal temperature recorded within one hour of admission
- Measure 3 - proportion of babies requiring ventilatory support given non-invasive ventilation (NIV) on day one of life
- Measure 4 - proportion of babies requiring intensive care provided with 1:1 nursing care on day one of life
- Measure 5 – receipt of mother's milk on day 1 of life
- Measure 6 – delayed cord clamping

For conciseness, from here on I have referred to these measures as:

- Measure 1 (steroids)
- Measure 2 (temperature)
- Measure 3 (ventilation)
- Measure 4 (nursing)
- Measure 5 (milk)
- Measure 6 (cord)

Each of these MQC were used to categorise units by adherence, into those that were in the top quartile versus those that were not in the top quartile (i.e. the bottom three quartiles). For conciseness, from here on I have referred to these as 'Group 1' and 'Group 2' respectively.

In the following subsections I have detailed how the data was sorted to calculate the proportion of babies receiving each measure within each unit. This is summarised in Table 11.

Non-NNAP measure of quality of care (non-NNAP MQC)		Variables related to measure	Value(s) inputted to calculate numerator		Value(s) inputted to calculate denominator		Exclusions	Value(s) inputted to signify missing data	
Measure 1	**Any dose of antenatal steroids given**	1) Any antenatal steroids given	Yes	[any value]	No	[any value other than 'yes']		Unknown [no value]	
		2) Complete course of steroids given	[any value]	Complete Incomplete	[any value other than 'complete' or 'incomplete']	None		Unknown [no value]	
Measure 2	**Normal temperature recorded within one hour of admission**	1) Admission time 2) Admission temperature time	[admission temp time – admission time ≤60 minutes]		[admission temp time – admission time >60 minutes] [no value for admission temp time]			[admission temp time – admission time >1440 minutes]	
		3) Admission temperature	36.5 – 37.5^{O}C		<36.5^{O}C >37.5^{O}C [no value]			<34^{O}C or >40^{O}C	
Measure 3	**Proportion of babies requiring ventilatory support given non-invasive ventilation (NIV) on day one of life**	1) Respiratory support	Non-invasive support	[no value]	Ventilation via ET tube Non-invasive support	[no value]	Resuscitation involving intubation [from numerator] Resuscitation involving cardiac massage or adrenaline	[no value]	
		2) Non-invasive respiratory support	[any value]	[any value]	[any value]	[any value]		[no value]	
Measure 4	**Proportion of babies requiring intensive care provided with 1:1 nursing care on day one of life**	1) BAPM level of care (2011)	Intensive care		Intensive care			[no value]	Intensive care
		2) One to one nursing	Yes		[any value]			[any value]	[no value]

Measure 5	**Receipt of mother's milk on day 1 of life**	1) Type of enteral feed	Suckling at the breast Mother's fresh expressed breast milk Mother's frozen expressed breast milk	[no value]	Nil by mouth Donor expressed breast milk Breast milk fortifier Formula Other	Length of stay (LOS) = 1 day Outcome = death	[no value]
		2) Feeding method	[any value]	Breast	[any value]		Bottle Cup Nasogastric tube Orogastric tube Gastrostomy Nasojejunal tube Other
Measure 6	**Delayed cord clamping**	1) Delayed cord clamping	Yes		Yes No	Apgar score at 1 minute = <7	[no value]

Table 11 Summary of how different variables from NDS were used to determine individual unit adherence with my non-NNAP MQC

7.1.1.1 Measure 1 - any dose of antenatal steroids given

There were two variables related to this measure:

1) Any antenatal steroids given
    - Values included 'yes', 'no', 'unknown' and no value inputted
2) Complete course of steroids given
    - Values included 'complete', 'incomplete', 'none', 'unknown' and no value inputted

Babies who received at least one dose of antenatal steroids were those who had 'yes' inputted for the first variable, regardless of the value inputted for the second variable, or 'complete' or 'incomplete' inputted for the second variable, regardless of the value inputted for the first variable.

Babies who did not receive at least one dose of antenatal steroids were those who had 'no' inputted for the first variable and did not have 'complete' or 'incomplete' inputted for the second variable, or 'none' inputted for the second variable but did not have 'yes' inputted for the first variable.

Babies for whom it was not possible to determine if they had received at least one dose of antenatal steroids (i.e., missing data) were those who had 'unknown' or no value inputted for the first variable and second variable.

7.1.1.2 Measure 2 - normal temperature measured within one hour of admission

There were three variables related to this measure:

1) Admission time
    - Value in minutes
2) Admission temperature time
    - Value in minutes
3) Admission temperature
    - Value in degrees centigrade

Using the first two variables I was able to work out how soon after admission the temperature was recorded. In combination with the third variable, babies with a normal temperature measured within one hour of admission were those with an admission temperature between 36.5 – 37.5^{O}C measured within $\leq$60 minutes of admission.

Babies who did not have a normal temperature and/or it was not measured within one hour of admission were those where the temperature was <36.5^{O}C or >37.5^{O}C or no value was inputted, and/or it was measured >60 minutes after admission or no value was inputted for admission temperature time.

Babies categorised as having missing data for this measure were those with a temperature recorded as <34^{O}C or >40^{O}C since these were implausible (and it is likely a typographical error had been made), and/or the admission temperature time minus the admission time was >1440 minutes (i.e., one day). This implied the incorrect date has been entered for admission temperature time, and it was not possible to determine if it was actually within 60 minutes of admission or not.

7.1.1.3 Measure 3 - proportion of babies requiring ventilatory support given non-invasive ventilation (NIV) on day one of life

There were two variables related to this measure:

1) Respiratory support
    - Values included 'no ventilation or CPAP', 'ventilation via endotracheal tube or tracheostomy', 'non-invasive support (including CPAP)', and no value inputted
2) Non-invasive respiratory support
    - Values included 'nasal CPAP (prong or mask)', 'BIPAP/SIPAP', 'high flow O_2 / air device', and no value inputted

Babies who had 'ventilation via endotracheal tube' or 'non-invasive support' inputted for the first variable, and babies who had no value inputted for the first variable and any value inputted for the second variable, formed my denominator, i.e., babies requiring ventilatory support on day one of life. This included babies for whom both values for the first variable were inputted, which meant they received both forms of ventilation on day one of life.

For my numerator, I used babies with 'non-invasive support' inputted for the first variable (regardless of the value inputted for the second variable), and babies with no value inputted for the first variable but any value inputted for the second variable.

I also wanted to include in the numerator those babies that were initially trialled on NIV and may have required intubation and invasive ventilation still within the first day of life, and so had both values inputted for the first variable. But this would also include those babies intubated and invasively ventilated and subsequently extubated onto NIV on day one of life, who may not have needed intubation in the first place and may have managed on NIV had it been trialled. I did not want to include these babies in the numerator and so to delineate, I included those babies with both values inputted for the first variable but excluded those who were intubated during resuscitation.
Babies who require extensive resuscitation immediately after birth are not candidates for NIV and I would reasonably expect them to be intubated and invasively ventilated, at least initially. Therefore, I excluded those babies requiring chest compressions or adrenaline during resuscitation.

Babies categorised as having missing data for this measure were those without values inputted for both variables.

7.1.1.4 Measure 4 - proportion of babies requiring intensive care provided with 1:1 nursing care on day one of life

There were two variables related to this measure:

1) BAPM level of care (2011)
    - Values included ‘intensive care’, high dependency care’, ‘special care’, ‘normal care’, and no value inputted
2) One to one nursing
    - Values included ‘no’, ‘yes’, and no value inputted

Babies requiring intensive care on day one of life formed the denominator for this measure, which were those with ‘intensive care’ inputted for the first variable.
Babies requiring intensive care on the first day of life provided with 1:1 nursing formed the numerator, which were those with ‘intensive care’ inputted for the first variable and ‘yes’ inputted for the second variable.
Babies for whom it was not possible to determine what level of care they required or whether they received 1:1 nursing (i.e., missing data), were those with no value inputted for the first variable (regardless of the value inputted for the second variable), and those

with ‘intensive care’ inputted for the first variable and no value inputted for the second variable.

7.1.1.5 Measure 5 - receipt of mother’s milk on day one of life

There were two variables related to this measure:

1) Type of enteral feed
    - Values included ‘nil by mouth’, ‘suckling at the breast’, ‘mother's fresh expressed breast milk’, ‘mother's frozen expressed breast milk’, ‘donor expressed breast milk’, ‘breast milk fortifier’, ‘formula’, ‘other’, and no value inputted
2) Feeding method
    - Values included ‘breast’, ‘bottle’, ‘cup’, ‘nasogastric tube’, ‘orogastric tube’, ‘gastrostomy’, ‘nasojejunal tube’, ‘other’, and no value inputted

Babies who received their mother’s milk on day one of life were those with ‘suckling at the breast’, ‘mother’s fresh expressed breast milk’ and ‘mother’s frozen expressed breast milk’ inputted for the first variable, regardless of the value for the second variable. Also included were those babies with no value inputted for the first variable and ‘breast’ inputted for the second variable.

Babies who did not receive their mother’s milk on day one of life were those who had any other value inputted for the first variable apart from the three mentioned, irrespective of the value for the second variable.

Because I might not expect babies who were born in a poor condition, or who became significantly unwell on day one of life to receive their mother’s milk on day one of life, I excluded those with a length of stay of 1 day and outcome of death.

Babies categorised as having missing data for this measure were those with no value inputted for the first variable and anything other than breast inputted for the second.

7.1.1.6 Measure 6 - delayed cord clamping

There was one variable related to this measure:

1) Delayed cord clamping
    - Values included ‘no’, ‘yes’, ‘unknown’, and no value inputted

Babies who had delayed cord clamping were those with ‘yes’ recorded for this variable. Babies who did not have cord clamping were those with ‘no’ recorded for this variable.

Because I would not expect babies who are born requiring resuscitation to have delayed cord clamping, babies with an Apgar score at one minute of <7 were excluded.

Babies for whom it was not possible to tell if they had delayed cord clamping (i.e., missing data) were those with ‘unknown’ or no value inputted for this variable.

7.1.2 Excluding measures based on degree of missing data

Following sorting of the data according to my non-NNAP MQC, I was able to analyse the degree of missing data for each measure (as described above). Measures 5 (milk) and 6 (cord) had >10% missing data, and so were excluded when categorising units according to adherence (Figure 14). Measures 1-4 (steroids, temperature, ventilation, nursing) had <1% missing data.

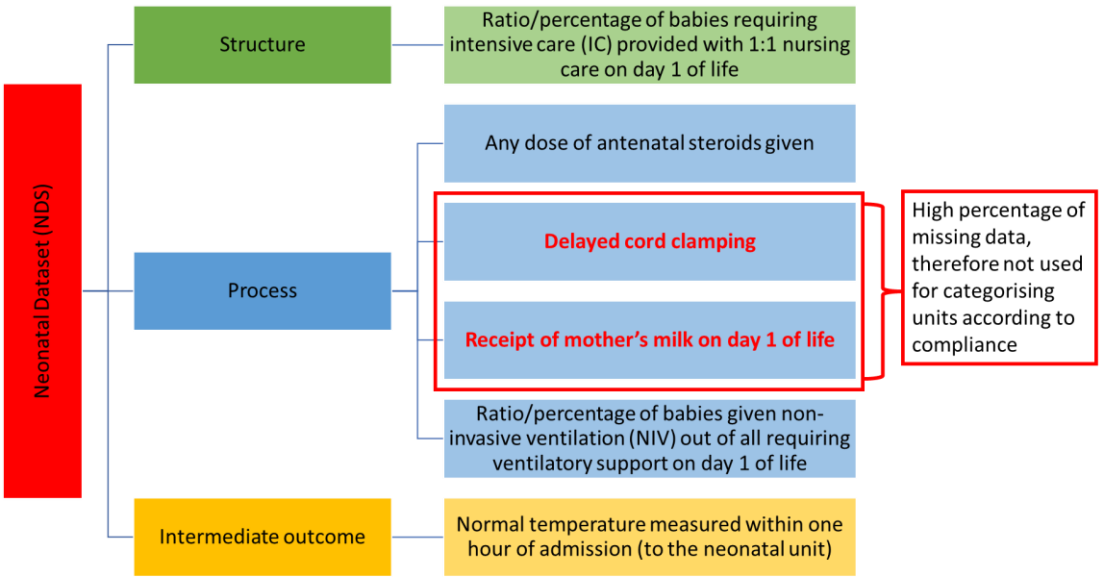


Figure 14 Schematic showing my different evidence-based measures of quality of care (MQC), classified according to structure, process and outcome, and exclusions due to levels of missing data

7.1.3 Categorising units according to non-NNAP measures of quality of care

7.1.3.1 Adherence

For Measures 1-4 (steroids, temperature, ventilation, nursing), I sorted the data by proportion of babies receiving the specified care (as detailed above) and identified units within Group 1 and 2. A combined list for Measures 1-4 was created in which units were organised hierarchically by the number of measures for which they were in Group 1 (Table 12). These were then separated into two groups; units in the top quartile for 2 or more measures, and units in the top quartile for <2 measures. Because I was using adherence with the combination of these measures as a marker of how willing neonatal units were to practice evidence-based medicine, I weighted the individual measures equally.

Adherence with non-NNAP MQC		**Categorisation of units**
Number of MQC	Units in the top quartile (n)	
4	1	Group 1 (units in top quartile for 2 or more measures) (n=33)
3	6	
2	26	
1	39	Group 2 (units in top quartile for less than 2 measures) (n=80)
0	41	

Table 12 Grouping of units by number of non-NNAP MQC for which they are in the top quartile, followed by separation into two groups for purposes of comparison

7.2 NNAP

NNAP data for 2018 was accessed and downloaded from the RCPCH website: https://nnap.rcpch.ac.uk/annual-reports.aspx. The initial download contained data for 184 units. Data for 65 units which were not LNU or NICU (i.e., SCU) were deleted, leaving the same 119 units as discussed in Section 7.1. The data were sorted by number of patients, and units with data for <10 patients or who requested the NNAP to hide their data for the audit measures I chose to use were excluded. These numbered 21 for adherence (leaving 98 units and 4594 patients), and 19 for data completion (leaving 100 units and 4722 patients). This included the one unit that opted out of the OptiPrem study. This threshold was chosen because a small number of data points increases variability making the result of any analysis less reliable. Conversely, if set too high this would lead to exclusion of a significantly large proportion of units, leading to a reduction in power.

Following this, for each of the proposed NNAP audit measures, the data was sorted by adherence and data completion.

7.2.1 Audit measures used for adherence

Audit measure	% missing data	
	Mean	Median
Parental presence at consultant ward rounds (for infants <7 days of age)	17.64	10.6
Follow-up at two years of age	12.75	5.25
NEC	5.64	3.65
Antenatal magnesium sulphate	2.56	0
Parent consultation within 24 hours of admission	1.61	0.15
On-time screening for ROP	0.95	0
BPD	0.41	0
Antenatal steroids	0.34	0
Promoting normal temperature on admission for very preterm babies	0.19	0
Breastmilk feeding at discharge home	0.15	0
Minimising inappropriate separation of mother and term baby	0.00	0
Minimising inappropriate separation of mother and late to moderate preterm baby	0.00	0

Table 13 Summary of missing data from 2018 NNAP audit measures (threshold set at 10%). Does not include audit measures relating to bloodstream infection due to absence of information on degree of missing data by neonatal unit (263)

I excluded measures with >10% levels of missing data (follow-up at two years of age, parental presence at consultant ward rounds – Table 13). I also excluded the two measures relating to bloodstream infection, due to a low level of data completion (only 119 of 179 units, 65.7%, provided assurance that 100% of positive blood culture data was submitted to the NNAP (175)). I have previously discussed the reason for excluding outcome measures (NEC, BPD).

This left a final list of NNAP audit measures to be used to categorise units according to adherence (Figure 15):

1. Antenatal steroids
    - NNAP standard 85%
2. Antenatal magnesium sulphate
    - Benchmarking measure
3. Promoting normal temperature on admission for very preterm babies
    - NNAP standard 90%
4. Minimising inappropriate separation of mother and late to moderate preterm baby
    - Benchmarking measure
5. Minimising inappropriate separation of mother and term baby
    - Benchmarking measure
6. Parent consultation within 24 hours of admission
    - NNAP standard 100%
7. On-time screening for ROP
    - NNAP standard 100%
8. Breastmilk feeding at discharge home
    - Benchmarking measure
9. Nurse staffing
    - NNAP standard 100%

The 'nurse staffing' audit measure contains several measures:

i. Number (%) of shifts with enough nurses to meet 'total nurses' element of service specification for the babies cared for on that shift
ii. Number (%) of shifts meeting the 'qualified in specialty' element of the service specification
iii. Additional number of nursing shifts which would need to be worked to staff all shifts to meet the 'total nurses' element of service specification
iv. Average additional number of nurses to meet service specification on all shifts

Out of these, the measure that I used to rank units was 'i'. Ideally, we would have used 'ii', since more important than just having an adequate number of nurses on shift is for the nurses to also have the appropriate training and expertise to look after the babies.

However, this was not possible because as per the method used by the NNAP (in that units where >50% of shifts are staffed with three registered nurses or fewer are excluded from the calculation for this measure and therefore no data is presented), eight units did not have any data available for this measure. It was also appropriate to use 'i' since that is what the NNAP uses to set its developmental standard (which is set at 100%). Since none of the units met this standard, the top quartile was selected.

Figure 15 Schematic showing the 2018 NNAP audit measures and basis for exclusion of specific measures when considering adherence

Units that met the audit threshold, or were in the top quartile for benchmarking measures, were amalgamated into a hierarchical list.

7.2.2 Audit measures used for missing data

I excluded measures which had no missing data (minimising inappropriate separation of mother and term, and late to moderate preterm babies, nurse staffing), since this would add no value to the analysis. I also excluded measures where the percentage of missing

data was <10% (antenatal steroids, BPD, breastmilk feeding at discharge home, promoting normal temperature on admission for very preterm babies), and for the remaining included measures, I only counted units as having missing data if their percentage of missing data was ≥10%, even if they were otherwise within the top quartile. Failure to exclude these measures/units would have increased the volatility of any analysis, where missing data for a very small number of patients (even one) could impact whether units are in the category of having missing data versus not having missing data. This arbitrary threshold of 10% was used for all analyses sensitive to degree of missing data. Bloodstream infection and central line associated bloodstream infection were excluded because information on degree of missing data by neonatal unit was not available.

This left a final list of NNAP audit measures to be used to categorise units according to data completion/missing data (Figure 16):

1. Antenatal magnesium sulphate
2. Parent consultation within 24 hours of admission
3. Parental presence at consultant ward rounds
4. On-time screening for ROP
5. NEC
6. Follow-up at two years of age

Since there were no standards/thresholds for acceptable levels of data completion, I selected the top quartile for missing data (versus the lower three quartiles for more complete data capture) as the measure against which to assess outcomes. For measures where less than a quarter of units had any missing data, I only included those units. This approach was necessary, as for several measures the majority of units had 100% data completion which meant that a 'top quartile' for data completion was not possible to define.

Figure 16 Schematic showing the 2018 NNAP audit measures and basis for exclusion of specific measures when considering data completion

7.2.3 Categorising units according to NNAP audit measures

Amalgamating these lists, I created two hierarchical lists, in which the 98-100 units were ordered by adherence (i.e., the number of audit measures for which they were in the top quartile/meeting the audit standard) and missing data (i.e., the number of audit measures for which they were in the top quartile for/had any missing data).

As can be seen in Tables 14 and 15, generally, the number of units in both lists increase as we go further down. Therefore, for the purposes of my analyses I created two groups within both lists. For adherence with audit measures, Group 1 included units meeting the threshold for/in the top quartile for adherence with 4 or more audit measures, and Group 2 included units meeting the threshold for/in the top quartile for adherence with <4 audit measures. For data completion, Group 1 included units in the top quartile for or with any missing data for 2 or more audit measures, and Group 2 included units in the top quartile

for or with any missing data for <2 audit measures. These divisions allowed for sufficient grouping of units to enable appropriate statistical analyses.

Adherence with NNAP audit measures		**Categorisation of units**
Number of audit measures	Units meeting threshold/in the top quartile (n)	
9	0	Group 1 (units meeting threshold / in top quartile for 4 or more measures) (n=21)
8	0	
7	2	
6	1	
5	4	
4	14	
3	37	Group 2 (units meeting threshold / in top quartile for less than 4 measures) (n=77)
2	25	
1	13	
0	2	

Table 14 Grouping of units by number of NNAP audit measures for which they are in the top quartile for compliance, followed by separation into two groups for purposes of comparison

Data completion for NNAP audit measures		**Categorisation of units**
Number of audit measures	Units in the top quartile/any missing data (n)	
6	0	Group 1 (units in top quartile / any missing data for 2 or more measures) (n=17)
5	1	
4	3	
3	3	
2	10	
1	39	Group 2 (units in top quartile / any missing data for less than 2 measures) (n=83)
0	44	

Table 15 Grouping of units by number of NNAP audit measures for which they are in the top quartile for data completion, followed by separation into two groups for purposes of comparison

To look for associations between NNAP data and outcome data provided to OptiPrem for babies born between 27-31 weeks of gestation, and to compare demographics of the comparator groups, the unit names by which the NNAP data was presented needed to be encoded to their corresponding unit codes by NDAU (Figure 17). In February 2021, each LNU and NICU (excluding the one unit that had already opted out of OptiPrem) was sent an update from OptiPrem including details of the analysis planned in WS2, and instructions of how to opt-out if they did not wish to give their consent. No units opted out and so all unit names were able to be encoded. This was done in groups of units and so it would not have been possible to use this data to de-anonymise NDAU data provided to OptiPrem.

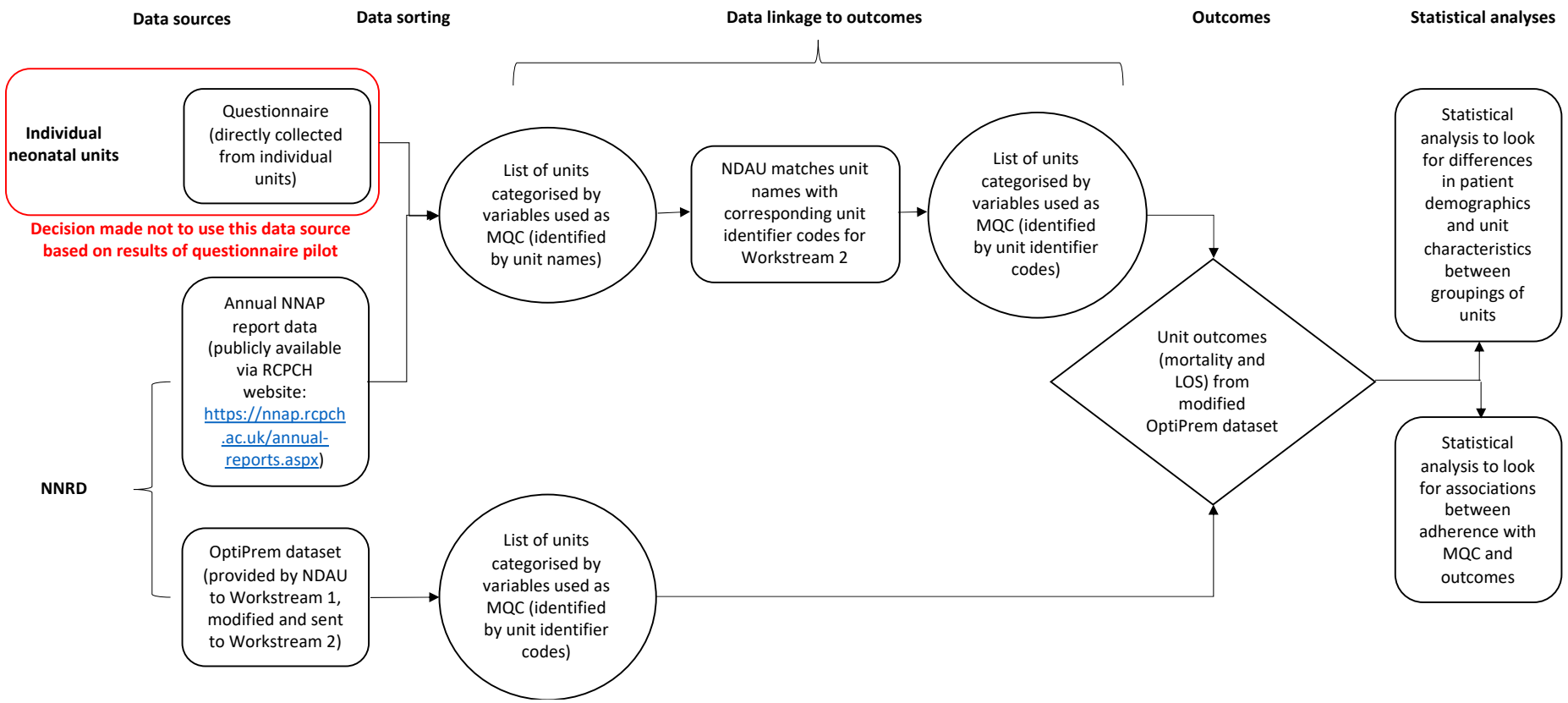


Figure 17 Schematic showing plan for data analyses, from data sources to statistical analyses
NNAP – National Neonatal Audit Programme, RCPCH – Royal College of Paediatrics and Child Health, NNRD – National Neonatal Research Database, NDAU – Neonatal Data Analysis Unit, MQC – Measures of quality of care, LOS – Length of stay

7.3 Demographics

I amalgamated details regarding the following demographic parameters (Figure 18), and the amount of missing data for each parameter:

- Unit characteristics:
    - Designation
        - LNU
        - NICU
            - Surgical NICU (i.e., provide care to babies requiring major surgery, including post-surgery care)
            - Non-surgical NICU
    - Number of babies by each gestational week (27-31)
- Neonatal characteristics:
    - General
        - Birthweight (mean)
        - Gender
        - Fetal number, grouped into:
            - 1
            - ≥2
    - Presence of significant congenital anomaly
    - Condition of baby in immediate postnatal period:
        - Cord base excess
        - Apgar score at 5 minutes (median)
        - Resuscitation involving cardiac massage or adrenaline
        - Worst base excess in first 24 hours of life
- Maternal characteristics:
    - Maternal health, grouped into:
        - No health problems (preceding and during pregnancy)
        - Any health problems (preceding and during pregnancy)
        - Drug and/or alcohol use

- Socioeconomic factors:
  - Ethnic group, split into five categories:
    - 1 – White
    - 2 – Mixed
    - 3 – Asian/Asian British
    - 4 – Black/Black British
    - 5 – Other
  - IMD_Q (index of multiple deprivation), split into quintiles:
    - 1 – Most deprived
    - 2
    - 3
    - 4
    - 5 – Least deprived

Following this, I combined the demographic data for Group 1 and Group 2 for my non-NNAP MQC and NNAP audit measures. This allowed comparison of the patient populations between the groupings (of units) I compared. Additional comparisons were made between these groupings, to allow for further comparison between different measures, including number of units, unit designations, and number of babies.

To understand whether the differences observed between the patient populations were significant, statistical tests were carried out. For unit designation, gestational week, gender, the IMD_Q, and resuscitation involving cardiac massage or adrenaline, the Chi-squared test was used. For comparing distribution of surgical versus non-surgical NICU the Chi-squared test was not used because a significant proportion of expected counts were <5, therefore the Fisher's exact test was used. For birthweight, the weighted two sample t-test was used (274). Missing data was variable. Any parameter with above the threshold level of missing data (10%) was excluded from statistical testing. This included Apgar score at 5 minutes (10.7%), cord base excess (60.5%), worst base excess in first 24 hours of life (100%), ethnic group (18.7%) and maternal health (38.8%).

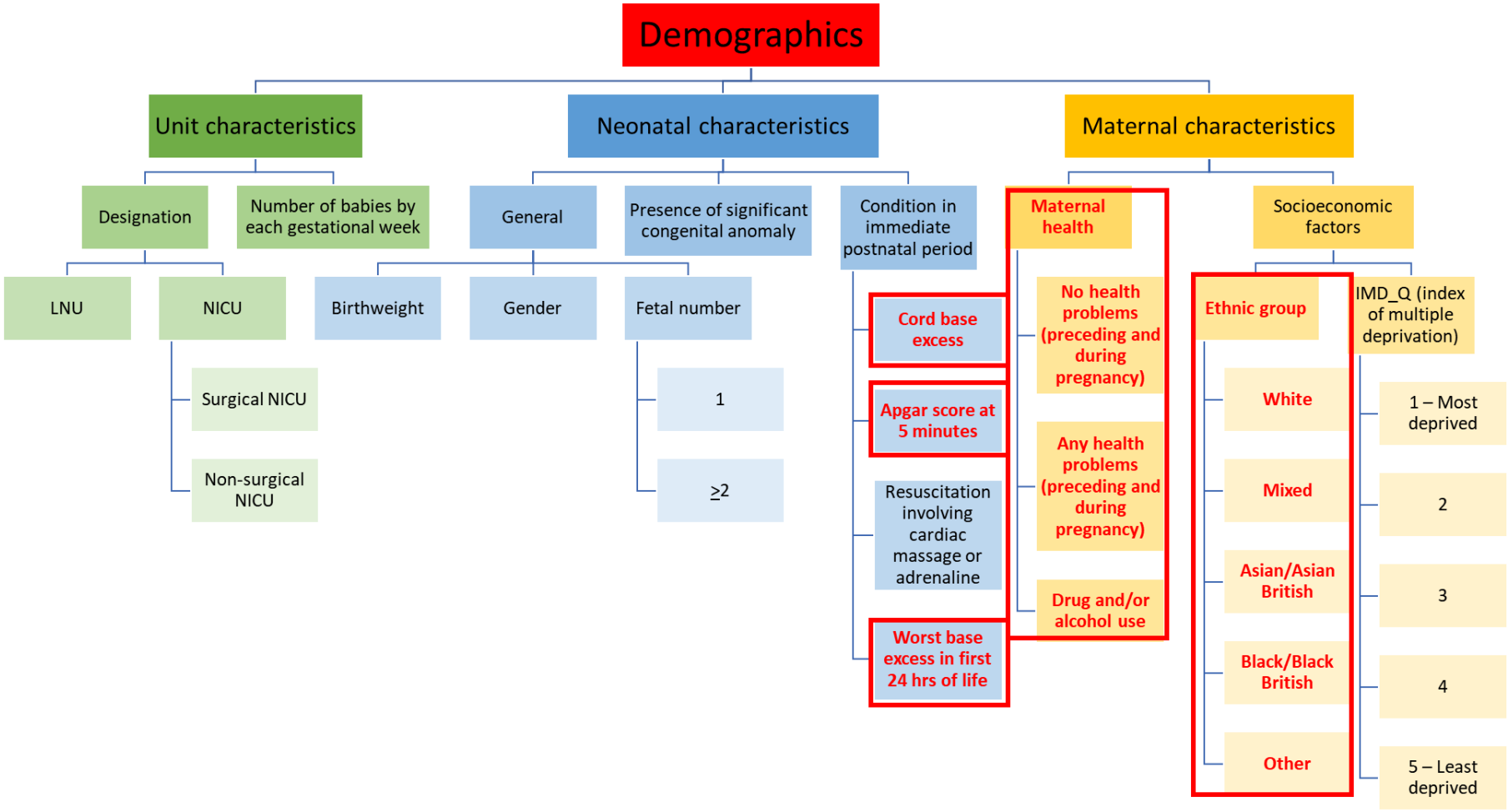


Figure 18 Schematic showing demographic parameters chosen for inclusion
Parameters in red text with red outline had >10% missing data and were excluded from statistical analyses

7.4 Looking for associations between outcomes and quality of care measures

7.4.1 Mortality and major morbidity outcome data

The outcomes I was looking for associations with were:

1. Pre-discharge mortality
2. Length of stay (LOS)

Mortality and length of stay (LOS) data was 100% complete. When analysing length of stay data, I excluded babies who died pre-discharge, since units which had a higher proportion of these babies would have an artificially reduced mean LOS.

I amalgamated pre-discharge mortality and length of stay data by gestational week, by gestational week and designation of unit of birth, and by gestational week and categorisation of units based on adherence and missing data for the NNAP audit measures, and adherence with my non-NNAP measures of quality of care. To analyse whether any difference in mortality between groups was statistically significant, I used the Chi-squared test. For length of stay data, I used the weighted two sample t-test. To check if the data was normally distributed it was plotted on a histogram.

7.4.2 Further analysis: multivariate analyses

Multivariate analyses were conducted for the following comparisons:

- Non-NNAP MQC:
    - Adherence with combination of non-NNAP MQC and mortality (n= 4364)
    - Adherence with combination of non-NNAP MQC and LOS (n= 4242)
- NNAP audit measures:
    - Adherence with NNAP audit measures and mortality (n= 4007)
    - Adherence with NNAP audit measures and LOS (n= 3894)

Analysis was conducted for the whole cohort of babies (i.e., 27-31 weeks), not by each gestational week of birth. This was because:

- The most important differences (in terms of clinical and statistical significance) were found for the whole cohort
- When splitting by gestational week the number of patients (especially at lower gestations where outcomes are more likely) reduced dramatically

This also meant I had a large patient group from which I was able to identify and exclude patients with unknown or incorrect demographic data for my variables of interest:

- Unit type
    - Categorical variable with three categories (LNU/non-surgical NICU/surgical NICU)
    - No missing data
- Birthweight
    - Continuous variable
    - Missing data and incorrect birthweight for 12 babies
- Gestational week
    - Continuous variable
    - No missing data (five babies without data already excluded)
- IMD_Q
    - Categorical variable with five categories (1-5)
    - Missing data for 173 babies
- Multiplicity
    - Categorical variable split into two categories (singletons, multiple pregnancy)
    - Missing data for six babies
- Gender
    - Categorical variable with two categories
    - Missing data for four babies
- Resuscitation involving cardiac massage/adrenaline
    - Categorical variable split into two categories (requiring significant resuscitation and not requiring significant resuscitation)
    - Missing data for 316 babies

The above numbers do not reflect the total number of exclusions for these variables, since some babies had missing or incorrect data for more than one of these variables (i.e., there were overlaps). Apart from the above, 137 babies had major congenital anomalies and were also excluded. When conducted analyses for LOS, 166 babies who died pre-discharge were also excluded.

For analyses involving the binary outcome variable, mortality, logistic regression was used. For analyses involving the continuous outcome variable, LOS, linear regression was used.

8 Results

In this chapter I describe the demographics of the patient cohort, broken down by gestational age and designation of unit, and the trends seen regarding adherence with my non-NNAP MQC. Following statistical testing, any significant differences found between the comparison groups regarding demographic profile and unit characteristics is presented. I also present the results of the univariate analysis looking for associations between adherence with my non-NNAP MQC and clinical outcomes. This is followed by the results of the univariate analysis looking for associations between adherence / data completion and outcomes for the NNAP audit measures. I compare the data for my non-NNAP MQC and NNAP audit measures. And finally, the results of the multivariate analyses is presented, alongside any transformation of data required to meet the assumptions of the tests.

8.1 Modified OptiPrem dataset

8.1.1 Describing demographics for entire cohort by gestational week of birth

When describing my cohort of babies born between 27-31 weeks of gestation within LNU and NICU and discharged in 2018, I have not excluded those born in units with <10 babies (as for the analysis for my non-NNAP measures of quality of care). Therefore, as described in Section 7.1, this cohort was of 5038 babies. They were born in 119 neonatal units, of which 44 were NICU (37%) and 75 were LNU (63%). Within NICU there was an equal split regarding those NICU that care for babies with significant surgical diagnoses and provide post-surgery care (called henceforth ‘surgical NICU’), and those that do not (called henceforth ‘non-surgical NICU’). Per unit, 64 babies born between 27-31 weeks were born in NICU compared to 30 in LNU (Figure 19).

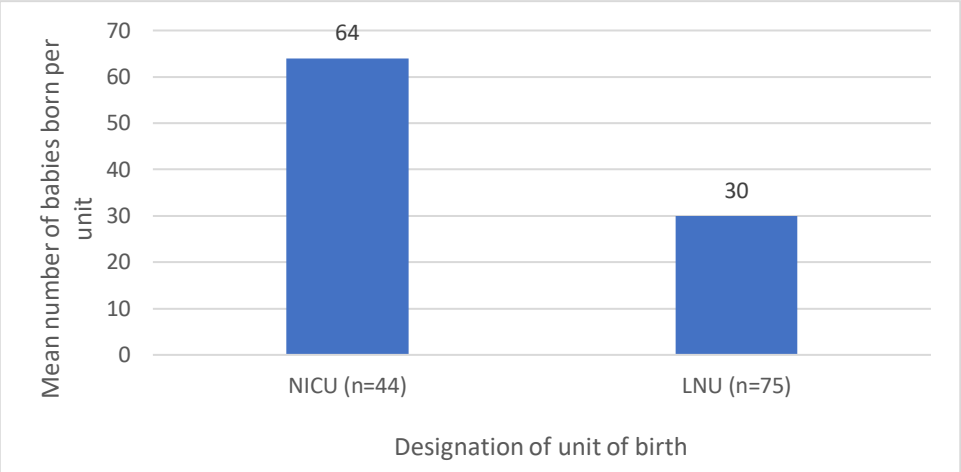


Figure 19 Number of babies born per unit by designation of unit of birth

Breaking down the cohort by gestation week of birth, this difference was largely due to a higher proportion of babies born at 27 and 28 weeks of gestation being born in NICU (70.3% and 59.1% vs. 29.7% and 40.9%, respectively, p=<0.01) (Table 16, Figure 21). In keeping with this, the birthweight of babies born in NICU was, on average, 75g lower than for babies born in LNU (p=<0.01) (Figure 20).

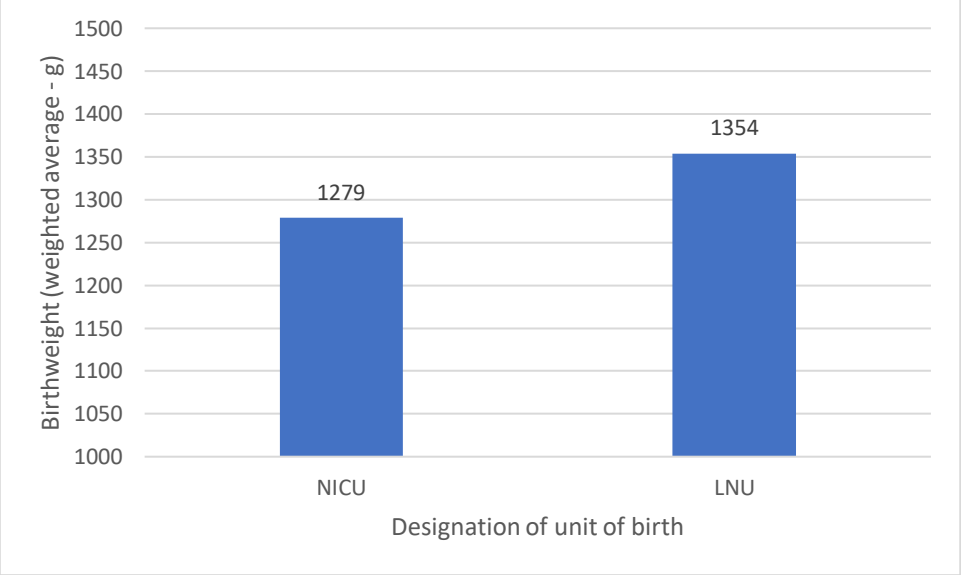


Figure 20 Birthweight (weighted average) by designation of unit of birth

I also observed an increase in the number of babies and their birthweight with increasing gestational age at birth, in keeping with the WHO preterm growth charts (https://www.rcpch.ac.uk/resources/uk-who-growth-charts-neonatal-infant-close-monitoring-nicm) (Table 16, Figure 22).

Weeks of gestation at birth	Number of babies		Mean birthweight (g)	SD (g)	Designation of unit of birth			
	n	(% of total cohort)			NICU	(% per gestational week)	LNU	(% per gestational week)
27	634	(12.6%)	968	(186)	446	(70.3%)	188	(29.7%)
28	775	(15.4%)	1091	(217)	458	(59.1%)	317	(40.9%)
29	919	(18.2%)	1227	(251)	485	(52.8%)	434	(47.2%)
30	1161	(23.0%)	1403	(272)	625	(53.8%)	536	(46.2%)
31	1549	(30.7%)	1549	(298)	791	(51.1%)	758	(48.9%)

Table 16 Data for number of babies, mean birthweight, and designation of unit of birth by gestational week of birth.
Missing data for birthweight n=12 (0.2%).

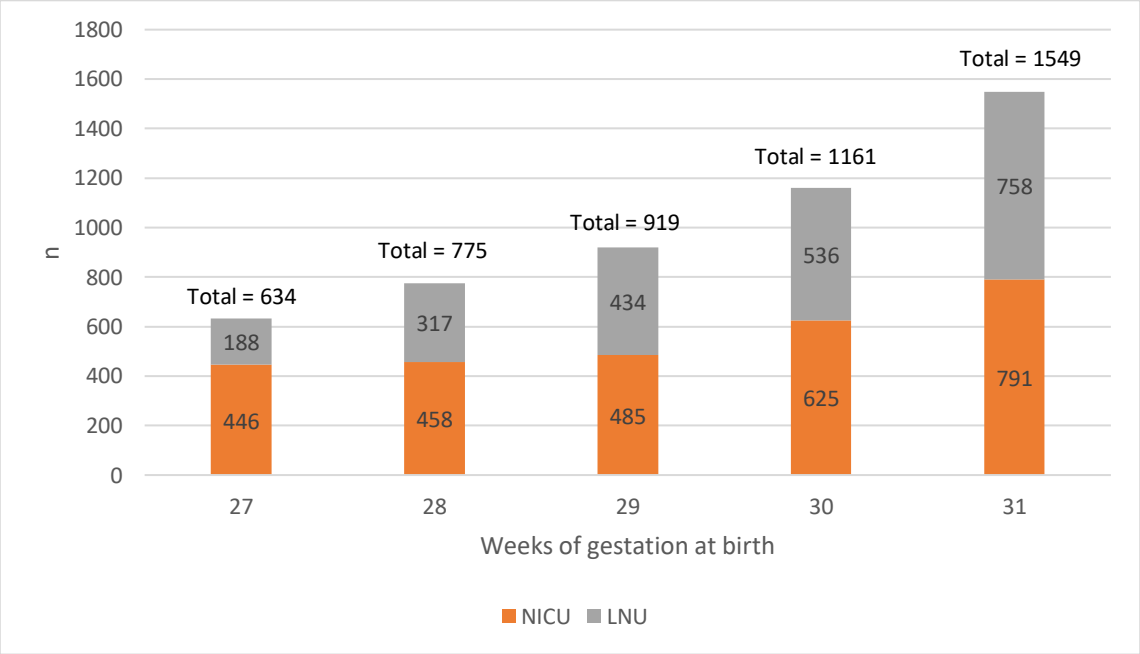


Figure 21 Number of babies born by gestational week and designation of unit of birth

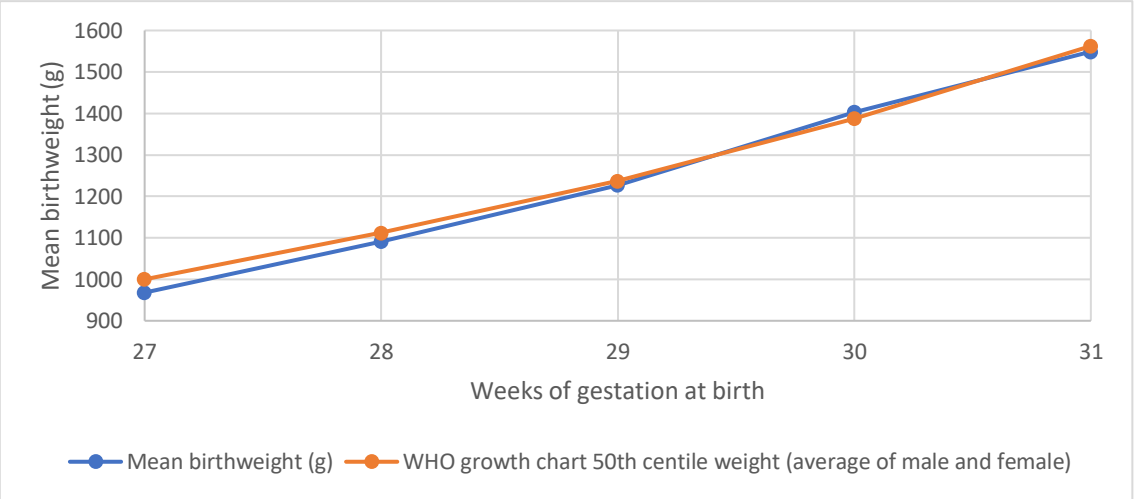
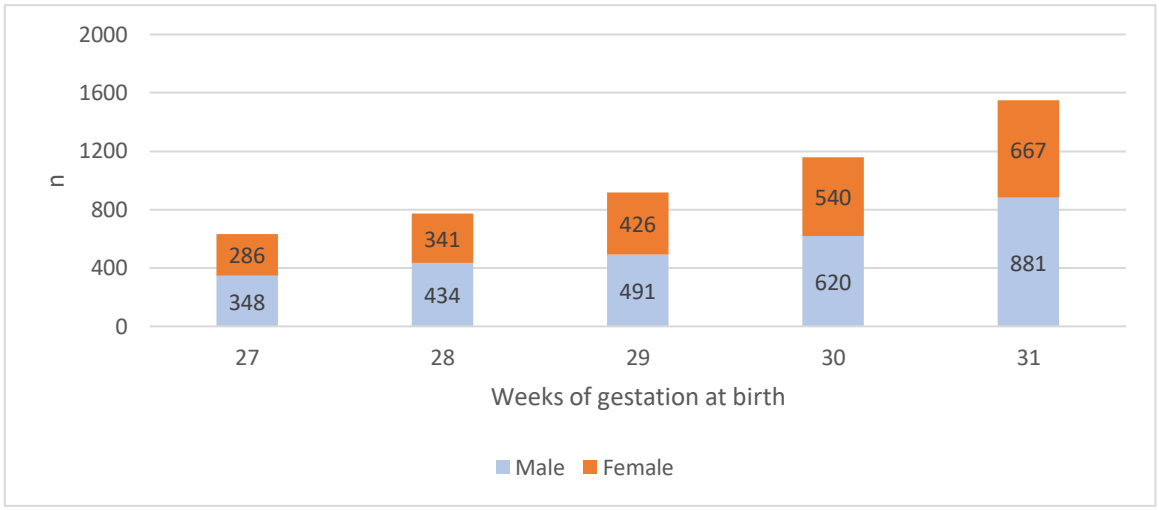


Figure 22 Mean birthweight by gestational week of birth

Breaking down the data by gender, there was a slight preponderance for males over females, which was largely consistent over the gestational age range (Table 17, Figure 23). By designation of unit of birth (NICU vs. LNU), the proportion of male to female babies in both types of units was identical (55% vs. 45%) (Figure 24). For multiplicity, as expected the majority of pregnancies involved singletons (Table 17, Figure 25). However, there was a slight trend towards reduced multiplicity with lower gestational age at birth. As a proportion of total births, significantly more multiple births were delivered in NICU compared to LNU (p=0.02) (Figure 26).

Weeks of gestation at birth	Gender				Multiplicity			
	Male	(%)	Female	(%)	1 fetus	(%)	≥2 fetuses	(%)
27	348	(54.9%)	286	(45.1%)	515	(81.2%)	119	(18.8%)
28	434	(56.0%)	341	(44.0%)	573	(73.9%)	200	(25.8%)
29	491	(53.4%)	426	(46.4%)	681	(74.1%)	236	(25.7%)
30	620	(53.4%)	540	(46.5%)	821	(70.1%)	340	(29.3%)
31	881	(56.9%)	667	(43.1%)	1100	(71.0%)	447	(28.9%)

Table 17 Data for gender and multiplicity by gestational week of birth.
Missing data for gender n=4 (0.1%), for multiplicity n=6 (0.1%).



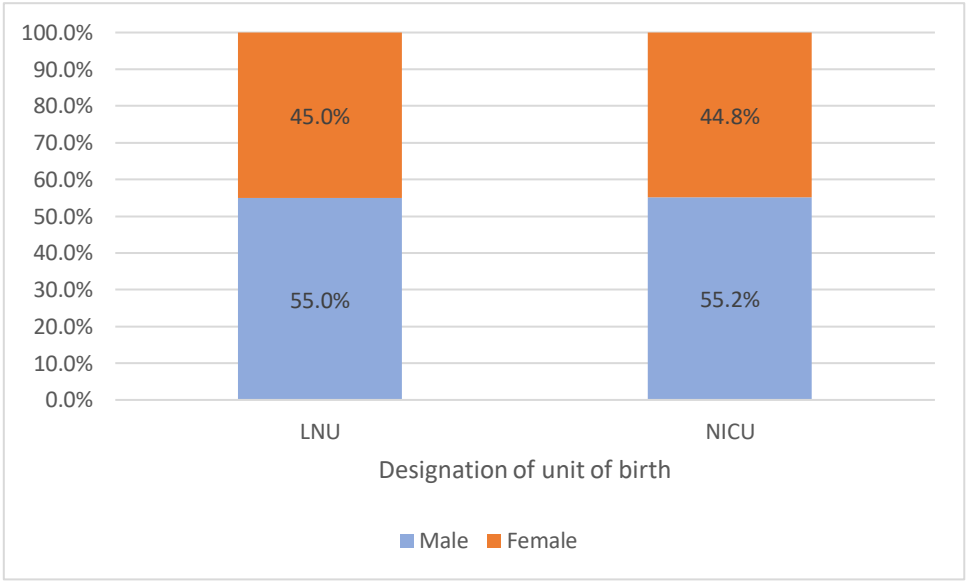


Figure 24 Proportion of male versus female babies by designation of unit of birth

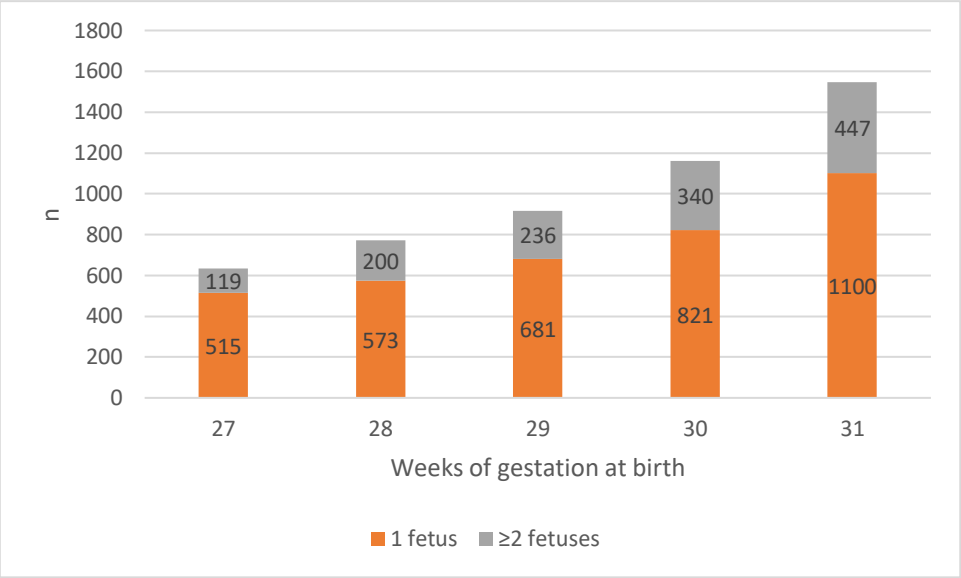


Figure 25 Number of babies by gestational week of birth and multiplicity

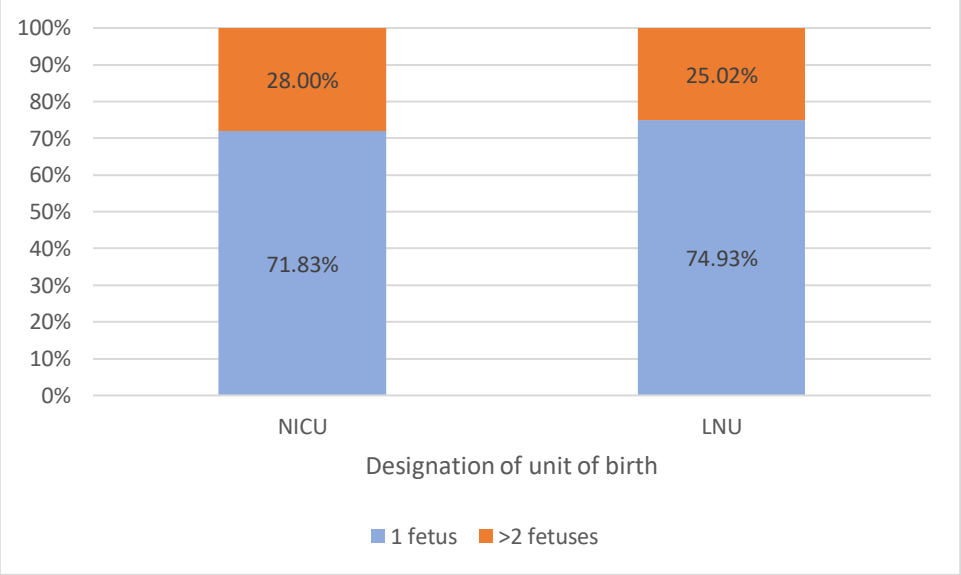


Figure 26 Proportion of singletons versus pregnancies with multiple foetuses by designation of unit of birth

2.7% of babies were born with a major congenital anomaly (for list of included diagnoses please see Appendix IV).

Breaking down the data by the IMD_Q (index of multiple deprivation score), the split was generally consistent across the gestational age range (Table 18, Figure 27), with a reducing proportion from 1 (most deprived) to 5 (least deprived).

Weeks of gestation at birth	IMD_Q (index of multiple deprivation score: 1=most deprived, 5=least deprived)									
	1	(%)	2	(%)	3	(%)	4	(%)	5	(%)
27	201	(31.7%)	141	(22.2%)	99	(15.6%)	91	(14.4%)	75	(11.8%)
28	234	(30.2%)	173	(22.3%)	138	(17.8%)	118	(15.2%)	84	(10.8%)
29	291	(31.7%)	210	(22.9%)	163	(17.7%)	118	(12.8%)	112	(12.2%)
30	371	(32.0%)	255	(22.0%)	199	(17.1%)	154	(13.3%)	137	(11.8%)
31	442	(28.5%)	370	(23.9%)	257	(16.6%)	206	(13.3%)	226	(14.6%)

Table 18 Data for IMD_Q by gestational week of birth
Missing data n=173 (3.4%).

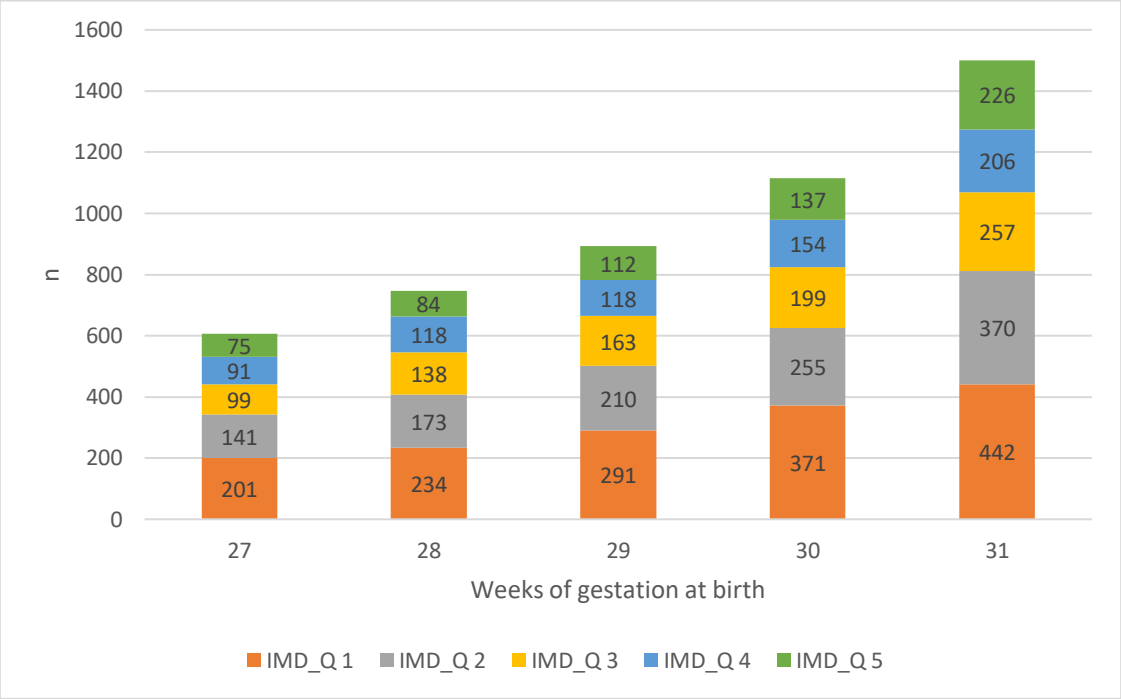


Figure 27 Number of babies by IMD_Q quintile and gestational week of birth

When analysing IMD_Q for my population of babies by designation of unit of birth, significantly more babies in the most deprived quintile (1), and two least deprived quintiles (4, 5) were born in NICU compared to LNU (p=<0.01) (Figure 28).

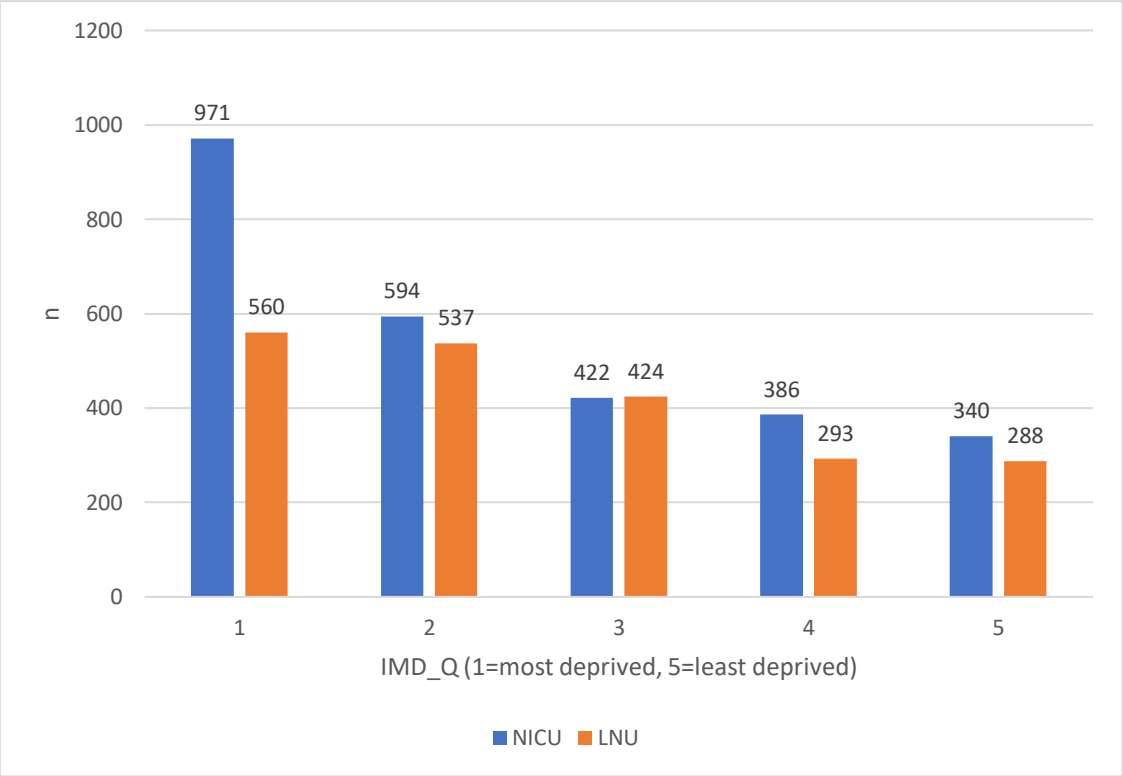


Figure 28 Number of babies by IMD_Q quintile and designation of unit of birth

8.1.2 Adherence with non-NNAP MQC across entire cohort by gestational week of birth

8.1.2.1 Measure 1 - receipt of any dose of antenatal steroids

Adherence with this measure was >90% and largely consistent across the gestational age range (Table 19, Figure 29).

Gestational week	Any dose of antenatal steroids	(%)
27	581	(92.1%)
28	717	(92.9%)
29	866	(94.3%)
30	1057	(91.4%)
31	1427	(92.5%)

Table 19 Data for any dose of antenatal steroids received across whole cohort by gestational week of birth.
Missing data n=18 (0.4%).

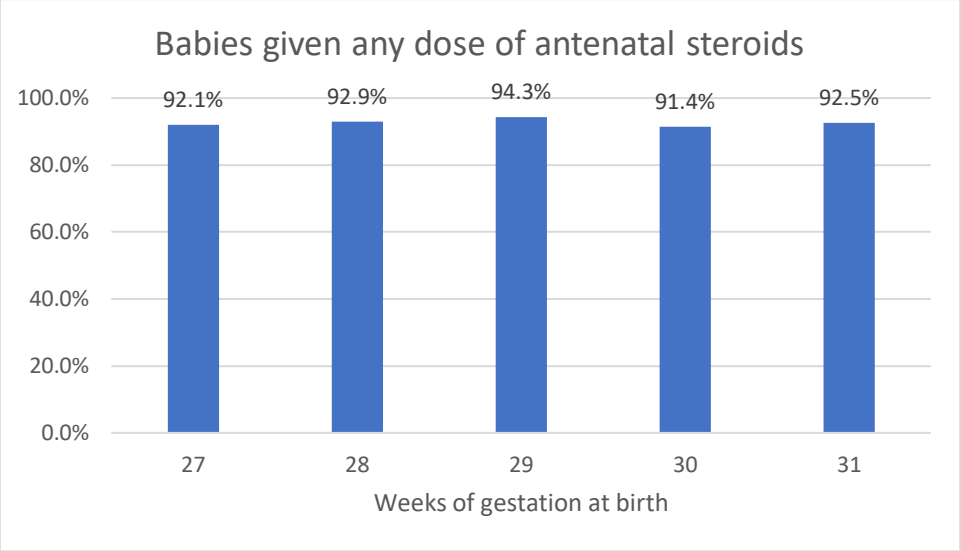


Figure 29 Proportion of babies given any dose of antenatal steroid by gestational week of birth

8.1.2.2 Measure 2 - normal temperature recorded within one hour of admission

67-68% of babies born at 27, 28 and 29 weeks of gestation had a normal temperature recorded within one hour of admission to the neonatal unit. For babies born at 30 and 31 weeks, this was slightly higher at 73-75% (Table 20, Figure 30).

Gestational week	Normal temperature within 1 hour of admission	(%)
27	420	(67.1%)
28	517	(67.6%)
29	612	(67.0%)
30	840	(73.0%)
31	1148	(74.6%)

Table 20 Data for normal temperature recorded within one hour of admission across whole cohort by gestational week of birth.
Missing data n=46 (0.9%)

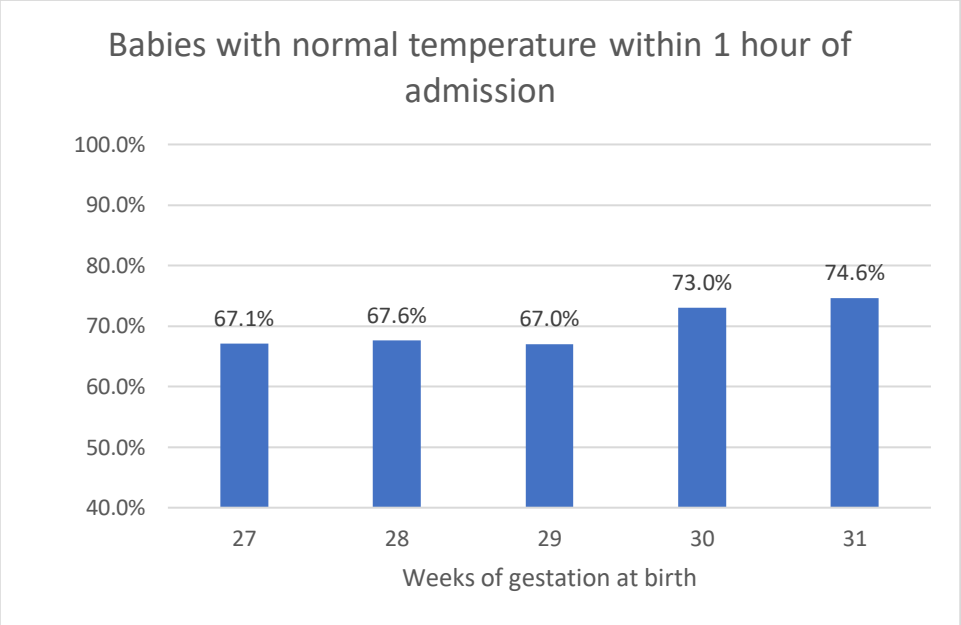


Figure 30 Proportion of babies with normal temperature within 1 hour of admission by gestational week of birth

8.1.2.3 Measure 3 - babies requiring ventilatory support on day one of life supported with non-invasive ventilation (NIV)

There was a nearly linear increase (from 32.8% to 79.2%), in the proportion of babies requiring ventilatory support on day one of life supported using NIV, from those born at 27 weeks through to those born at 31 weeks of gestation (Table 21, Figure 31).

Gestational week	Babies requiring ventilatory support	Babies supported with NIV (%)
27	597	196 (32.8%)
28	734	307 (41.8%)
29	856	470 (54.9%)
30	1046	749 (71.6%)
31	1249	989 (79.2%)

Table 21 Data for babies requiring ventilatory support on day one of life supported with NIV across whole cohort by gestational week of birth.
Missing data n=6 (0.1%).

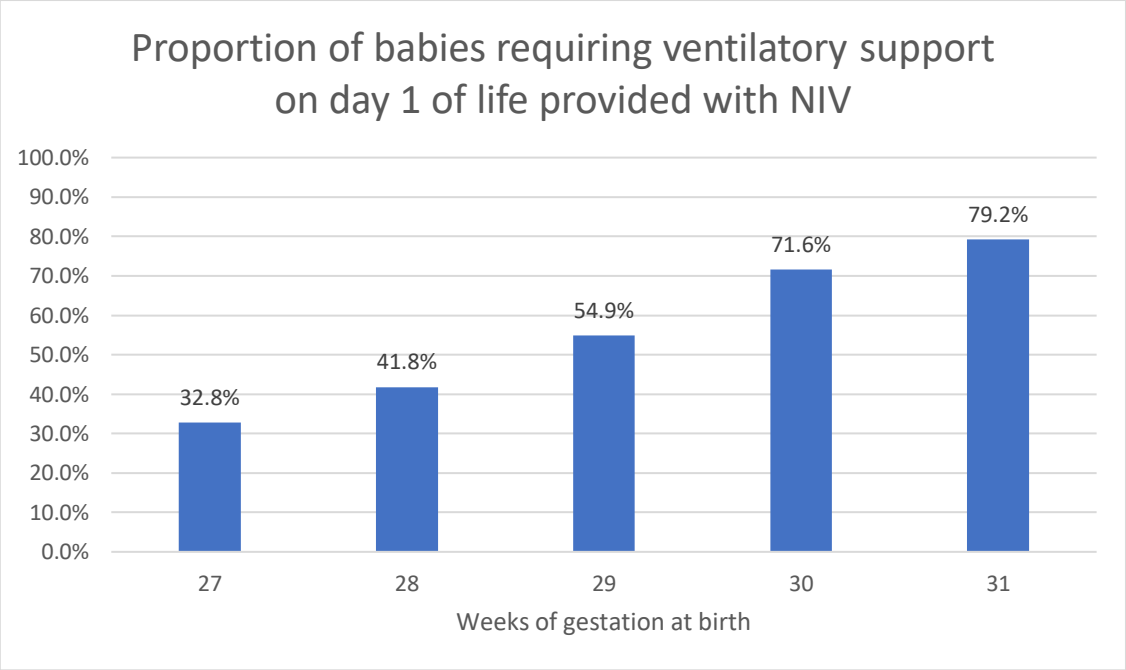


Figure 31 Proportion of babies requiring ventilatory support on day 1 of life provided with NIV by gestational week of birth

8.1.2.4 Measure 4 - babies requiring intensive care on day one of life provided with 1:1 nursing care

Adherence with this measure was generally low, with a maximum of 15.6% of babies born at 27 weeks of gestation requiring intensive care on day one of life receiving 1:1 care. This halved to 7.8% for babies born at 29 weeks of gestation, before slightly increasing to 10.1% at 31 weeks (Table 22, Figure 32).

Gestational week	Babies requiring intensive care	(%)	Babies requiring intensive care provided with 1:1 nursing care	(%)
27	596	(94.0)	93	(15.6)
28	689	(88.9)	87	(12.6)
29	753	(81.9)	59	(7.8)
30	706	(60.8)	63	(8.9)
31	661	(42.7)	67	(10.1)

Table 22 Data for babies requiring intensive care on day one of life provided with 1:1 nursing care across whole cohort by gestational week of birth.
Missing data n=16 (0.3%).

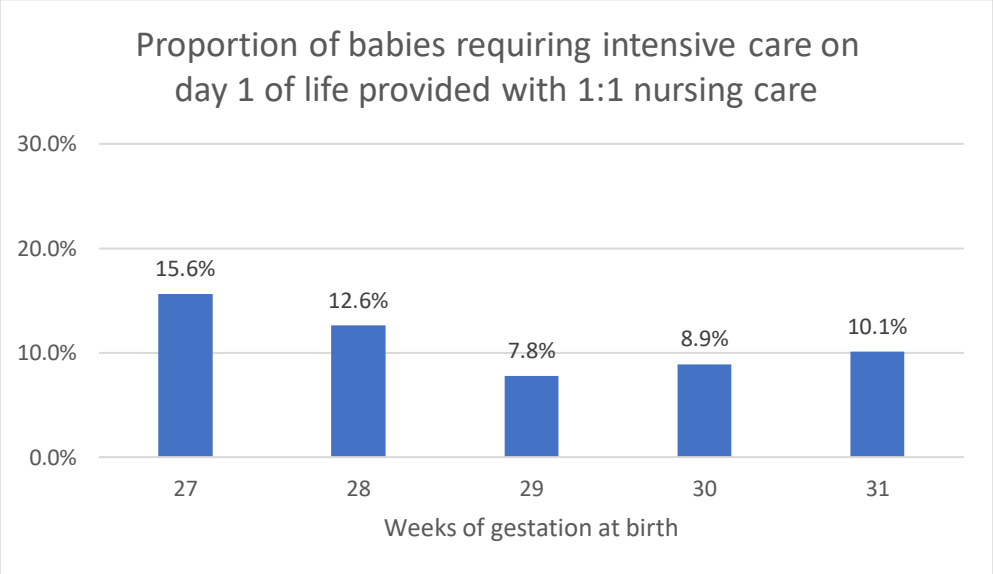


Figure 32 Proportion of babies requiring intensive care on day 1 of life provided with 1:1 nursing care by gestational week of birth

8.1.2.5 Excluded measures due to extent of missing data

8.1.2.5.7 Measure 5 - delayed cord clamping

There was a decreasing trend in adherence for delayed cord clamping, from 83.2% for babies born at 27 weeks of gestation to 70.6% for babies born at 31 weeks (Table 23, Figure 33).

Gestational week	Missing data	(%)	Delayed cord clamping	(% of babies without missing data who received delayed cord clamping)
27	214	(33.8%)	287	(83.2%)
28	291	(37.6%)	334	(81.5%)
29	348	(37.9%)	385	(77.8%)
30	431	(37.1%)	505	(76.4%)
31	577	(37.3%)	621	(70.6%)

Table 23 Data for delayed cord clamping across whole cohort by gestational week of birth. Missing data for whole cohort n=1861 (36.9%).

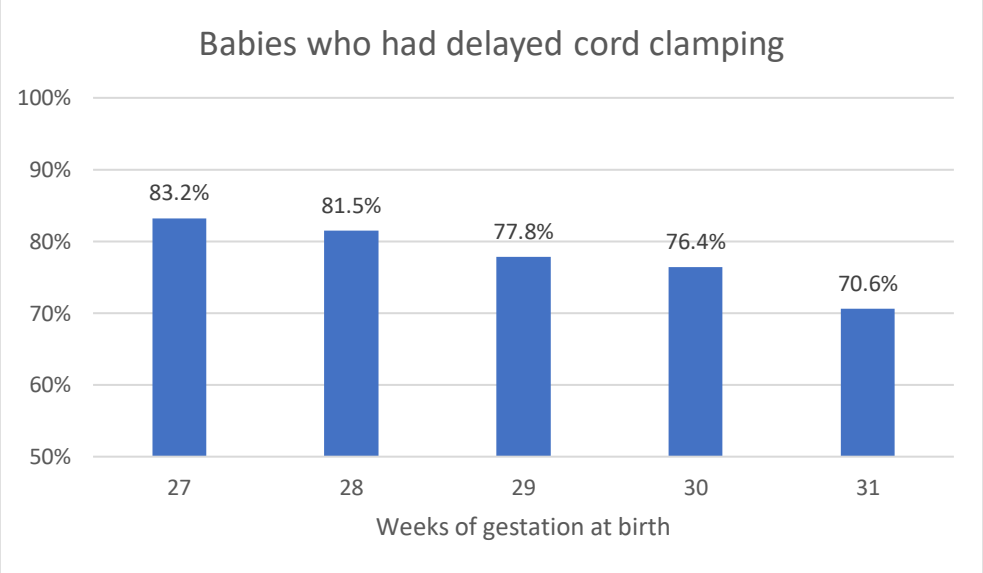


Figure 33 Proportion of babies who had delayed cord clamping by gestational week of birth

8.1.2.5.8 Measure 6 - receipt of mother's milk on day one of life

Adherence with this measure was generally low, but with an increasing trend from 9.1% of babies born at 27 weeks of gestation receiving their mother's milk on day one of life, to 15.0% for babies born at 31 weeks (Table 24, Figure 34).

Gestational week	Missing data	(%)	Receipt of mother's milk	(% of babies without missing data who received mother's milk)
27	91	(14.4%)	49	(9.1%)
28	116	(15.0%)	65	(10.0%)
29	142	(15.5%)	83	(10.8%)
30	168	(14.5%)	141	(14.2%)
31	226	(14.6%)	198	(15.0%)

Table 24 Data for receipt of mother's milk on day one of life across whole cohort by gestational week of birth.
Missing data for whole cohort n=743 (14.7%).

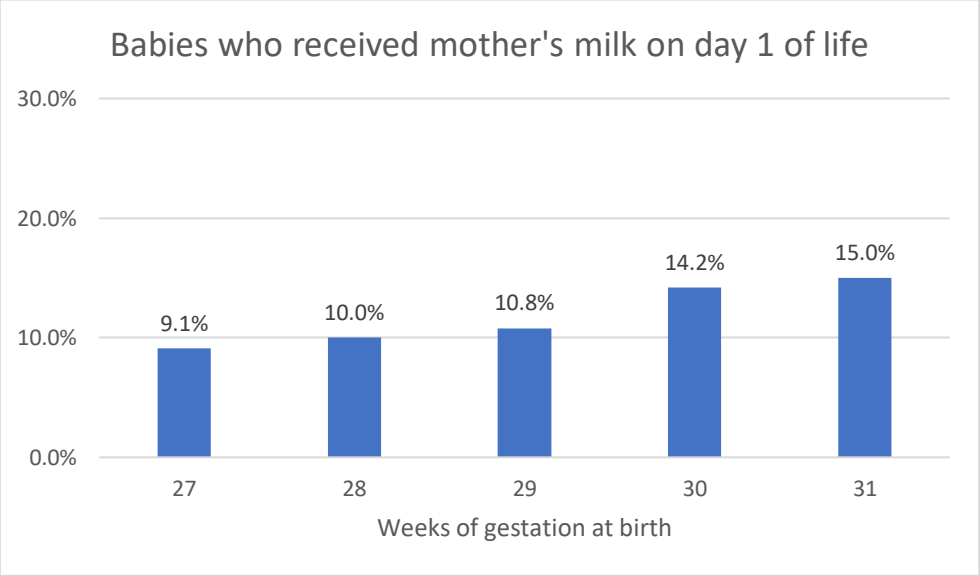


Figure 34 Proportion of babies who received mother's milk on day 1 of life by gestational week of birth

8.1.3 Comparing patient populations when categorising units based on adherence with non-NNAP MQC

8.1.3.1 Measure 1 - receipt of any dose of antenatal steroids

When categorising units by Measure 1 and comparing them and their populations, there was a significant difference in the IMD_Q (index of multiple deprivation score) (Table 25, Figure 35).

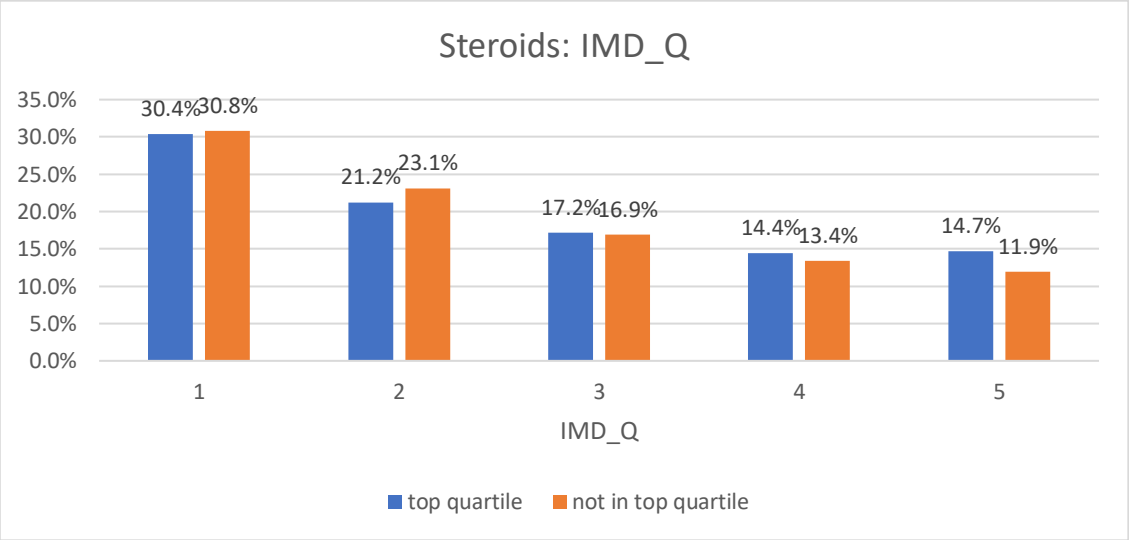


Figure 35 Proportion of babies in each IMD_Q quintile when categorising units according to Measure 1 of my non-NNAP MQC

Although the p value for the Chi squared test was <0.01, the maximum difference between observed and expected values across any of the IMD_Q categories in either grouping was ≤1%, and the direction of differences was not uniform. Therefore, the finding of statistical significance was likely secondary to the large dataset and of negligible clinical significance.

Measure 1 - receipt of any dose of antenatal steroids		Group 1: Top quartile		Group 2: Not in the top quartile		
Number of units		28		85		
Unit designations	**LNU**	17	(60.7%)	53	(62.4%)	p=1.00
	NICU (all)	11	(39.3%)	32	(37.6%)	
	NICU (surgical)	4	(14.3%)	17	(20.0%)	p=0.49
	NICU (non-surgical)	7	(25.0%)	15	(17.6%)	
Number of babies		1177	(23.6%)	3809	(76.4%)	
Birthweight (weighted average - g)		1299		1316		p=0.24
Unknown or incorrect birthweight		n=12 (0.2%)				
Gestational week	**27**	143	(12.1%)	487	(12.8%)	p=0.59
	28	195	(16.6%)	571	(15.0%)	
	29	205	(17.4%)	707	(18.6%)	
	30	279	(23.7%)	869	(22.8%)	
	31	355	(30.2%)	1175	(30.8%)	
Gender	**male**	662	(56.2%)	2084	(54.7%)	p=0.36
	female	514	(43.7%)	1722	(45.2%)	
	Unknown gender	n=4 (0.08%)				
Multiplicity	**1**	857	(72.8%)	2792	(73.3%)	p=0.77
	≥2	318	(27.0%)	1013	(26.6%)	
	Unknown multiplicity	n=6 (0.1%)				
Significant congenital anomalies		30	(2.5%)	107	(2.8%)	p=0.63
Apgar score at 5min (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=535 (10.7%)				
IMD_Q	**1 (most deprived)**	358	(30.4%)	1173	(30.8%)	p=<0.01
	2	250	(21.2%)	881	(23.1%)	
	3	202	(17.2%)	644	(16.9%)	
	4	170	(14.4%)	509	(13.4%)	
	5 (least deprived)	173	(14.7%)	455	(11.9%)	
	Unknown IMD_Q	n=171 (3.4%)				
Resuscitation involving cardiac massage or adrenaline		33	(2.8%)	102	(2.7%)	p=0.81
Unknown resuscitation status		n=316 (6.3%)				

Table 25 Comparison of patient populations when categorising units according to quality-of-care Measure 1 (receipt of any dose of antenatal steroids)

8.1.3.2 Measure 2 - normal temperature recorded within one hour of admission

When categorising units by Measure 2 and comparing them and their populations, there was a significant difference in the IMD_Q and resuscitation involving cardiac massage or adrenaline (Table 26, Figures 36 and 37).

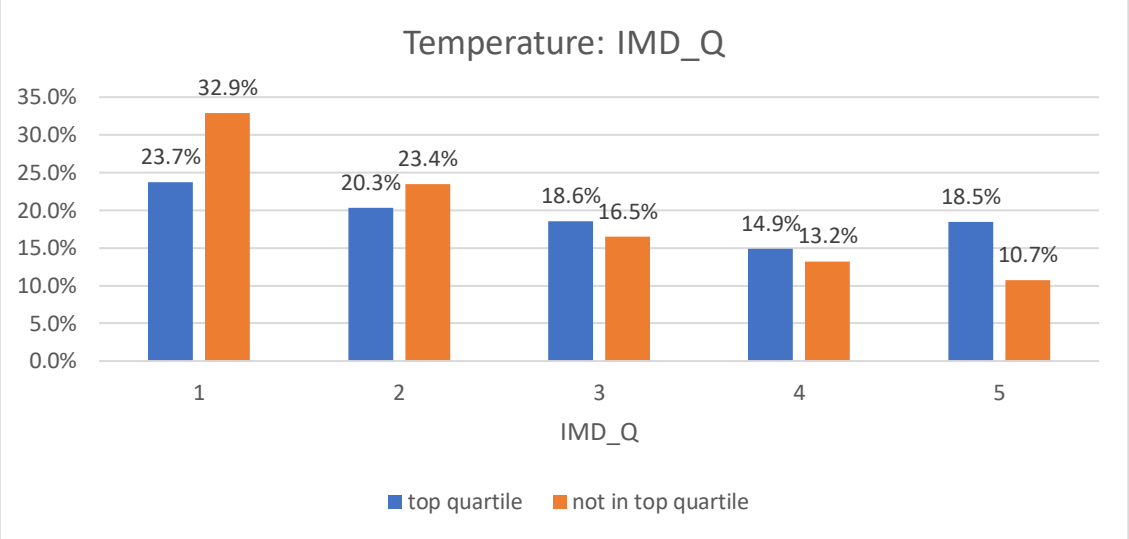


Figure 36 Proportion of babies in each IMD_Q quintile when categorising units according to Measure 2 of my non-NNAP MQC

For IMD_Q, units within the top quartile had a significantly greater proportion of babies in categories 3, 4 and 5, with a reduction in the proportion of babies in categories 1 and 2. I.e., there was a significant trend towards a less deprived population of patients in units that were in the top quartile, compared to units that were not in the top quartile.

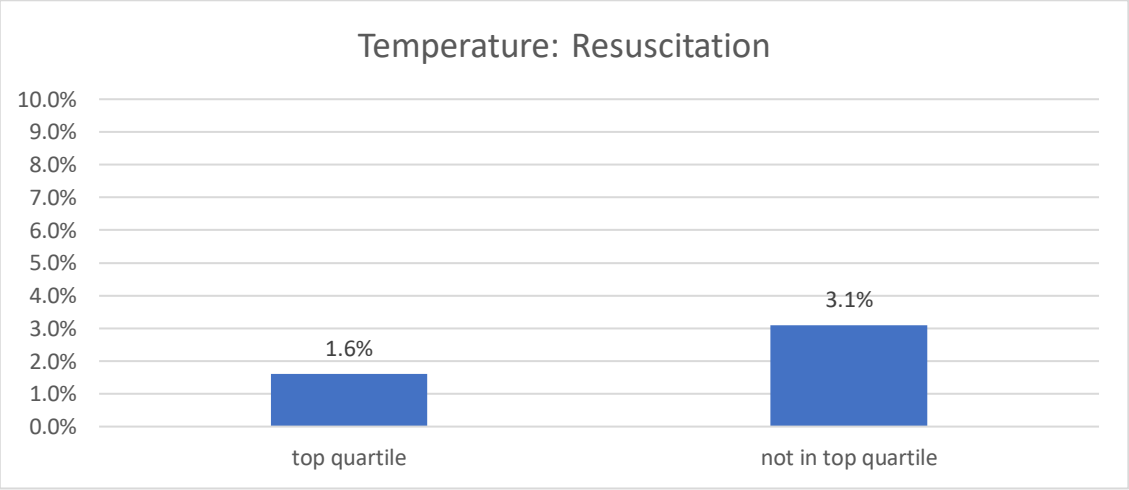


Figure 37 Proportion of babies requiring significant resuscitation when categorising units according to Measure 2 of my non-NNAP MQC

Nearly twice as many babies required resuscitation including cardiac massage or adrenaline in units that were not in the top quartile for this measure, compared to units that were in the top quartile. However, the absolute difference was only 1.5%.

Measure 2 - normal temperature recorded within one hour of admission		Group 1: Top quartile		Group 2: Not in the top quartile		
Number of units		29		84		
Unit designations	**LNU**	21	(72.4%)	49	(58.3%)	p=0.18
	NICU (all)	8	(27.6%)	35	(41.7%)	
	NICU (surgical)	5	(17.2%)	16	(19.0%)	p=0.46
	NICU (non-surgical)	3	(10.3%)	19	(22.6%)	
Number of babies		1202	(24.1%)	3784	(75.9%)	
Birthweight (weighted average - g)		1315		1311		p=0.90
Unknown or incorrect birthweight		n=12 (0.2%)				
Gestational week	**27**	149	(12.4%)	481	(12.7%)	p=0.96
	28	188	(15.6%)	578	(15.3%)	
	29	212	(17.6%)	700	(18.5%)	
	30	281	(23.4%)	867	(22.9%)	
	31	372	(30.9%)	1158	(30.6%)	
Gender	**Male**	691	(57.5%)	2055	(54.3%)	p=0.05
	Female	510	(42.4%)	1726	(45.6%)	
	Unknown gender	n=4 (0.08%)				
Multiplicity	**1**	888	(73.9%)	2761	(73.0%)	p=0.52
	≥2	312	(26.0%)	1019	(26.9%)	
	Unknown multiplicity	n=6 (0.1%)				
Significant congenital anomalies		29	(2.4%)	108	(2.9%)	p=0.41
Apgar score at 5 minutes (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=535 (10.7%)				
IMD_Q	**1 (most deprived)**	285	(23.7%)	1246	(32.9%)	p=<0.01
	2	144	(12.0%)	887	(23.4%)	
	3	223	(18.6%)	623	(16.5%)	
	4	179	(14.9%)	500	(13.2%)	
	5 (least deprived)	222	(18.5%)	406	(10.7%)	
	Unknown IMD_Q	n=171 (3.4%)				
Resuscitation involving cardiac massage or adrenaline		19	(1.6%)	116	(3.1%)	p=<0.01
Unknown resuscitation status		n=316 (6.3%)				

Table 26 Comparison of patient populations when categorising units according to quality of care Measure 2 (normal temperature recorded within one hour of admission)

8.1.3.3 Measure 3 - babies requiring ventilatory support on day one of life supported with non-invasive ventilation (NIV)

When categorising units by Measure 3 and comparing them and their populations, there was a significant difference in the IMD_Q, and resuscitation involving cardiac massage or adrenaline (Table 27, Figure 38 and 39).

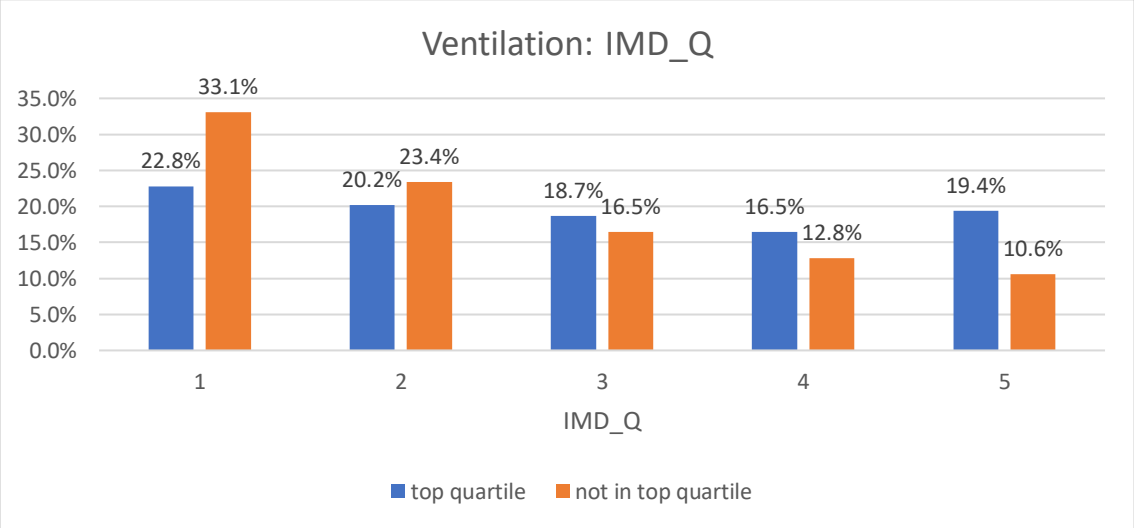


Figure 38 Proportion of babies in each IMD_Q quintile when categorising units according to Measure 3 of my non-NNAP MQC

For the IMD_Q, units within the top quartile had a significantly greater proportion of babies in categories 4 and 5, with a reduction in the proportion of babies in categories 1 and 2. Therefore, similar to the previous measure, there was a significant trend towards a less deprived population of patients in units that were in the top quartile, compared to units that were not in the top quartile.

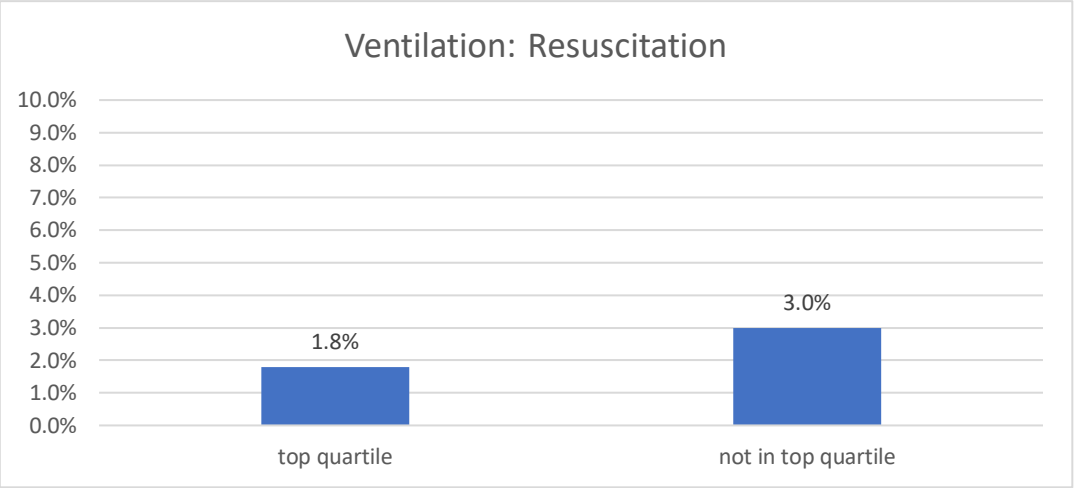


Figure 39 Proportion of babies requiring significant resuscitation when categorising units according to Measure 3 of my non-NNAP MQC

More babies required resuscitation including cardiac massage or adrenaline in units that were not in the top quartile for this measure, compared to units that were in the top quartile. While this reached statistical significance, the absolute difference was only 1.2%.

Measure 3 - babies requiring ventilatory support on day one of life supported with non-invasive ventilation		Group 1: Top quartile		Group 2: Not in the top quartile		
Number of units		28		85		
Unit designations	**LNU**	21	(75.0%)	49	(57.6%)	p=0.07
	NICU (all)	7	(25.0%)	36	(42.4%)	
	NICU (surgical)	4	(14.3%)	17	(20.0%)	p=0.70
	NICU (non-surgical)	3	(10.7%)	19	(22.4%)	
Number of babies		1141	(22.9%)	3845	(77.1%)	
Birthweight (weighted average - g)		1331		1306		p=0.11
Unknown or incorrect birthweight		n=12 (0.2%)				
Gestational week	**27**	136	(11.9%)	494	(12.8%)	p=0.14
	28	158	(13.8%)	608	(15.8%)	
	29	199	(17.4%)	713	(18.5%)	
	30	267	(23.4%)	881	(22.9%)	
	31	381	(33.4%)	1149	(29.9%)	
Gender	**male**	639	(56.0%)	2107	(54.8%)	p=0.46
	female	501	(43.9%)	1735	(45.1%)	
	Unknown gender	n=4 (0.08%)				
Multiplicity	**1**	849	(74.4%)	2800	(72.8%)	p=0.30
	≥2	291	(25.5%)	1040	(27.0%)	
	Unknown multiplicity	n=6 (0.1%)				
Significant congenital anomalies		24	(2.1%)	113	(2.9%)	p=0.13
Apgar score at 5min (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=535 (10.7%)				
IMD_Q	**1 (most deprived)**	260	(22.8%)	1271	(33.1%)	p=<0.01
	2	230	(20.2%)	901	(23.4%)	
	3	213	(18.7%)	633	(16.5%)	
	4	188	(16.5%)	491	(12.8%)	
	5 (least deprived)	221	(19.4%)	407	(10.6%)	
	Unknown IMD_Q	n=171 (3.4%)				
Resuscitation involving cardiac massage or adrenaline		21	(1.8%)	114	(3.0%)	p=0.04
Unknown resuscitation status		n=316 (6.3%)				

Table 27 Comparison of patient populations when categorising units according to quality of care Measure 3 (babies requiring ventilatory support on day one of life supported with non-invasive ventilation)

8.1.3.4 Measure 4 - babies requiring intensive care on day one of life provided with 1:1 nursing care

When categorising units by Measure 4 and comparing them and their populations, there was a significant difference in unit designation and the IMD_Q (Table 28, Figures 40 and 41).

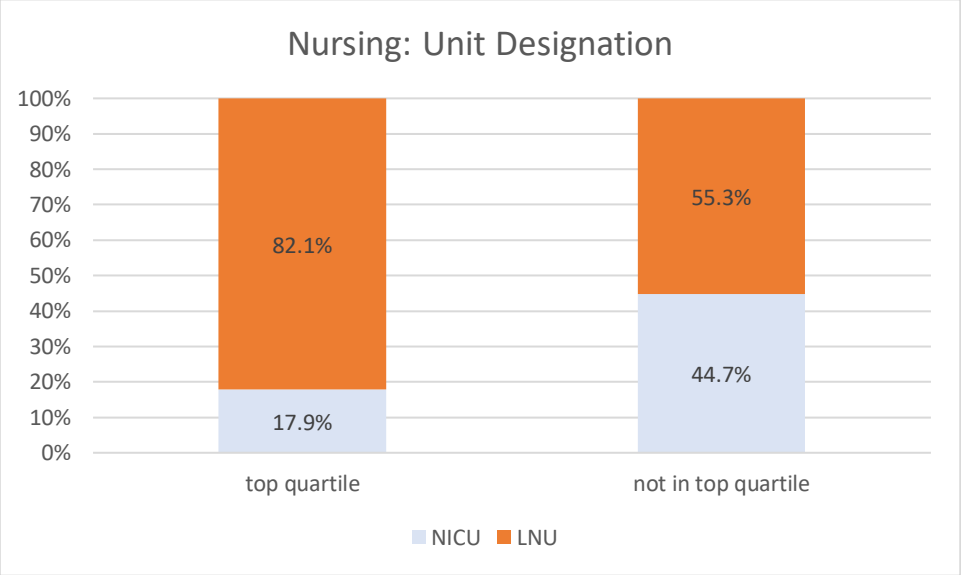


Figure 40 Proportion of babies by designation of unit of birth when categorising units according to Measure 4 of my non-NNAP MQC

The ratio of LNU:NICU was significantly greater for units within the top quartile compared to units not within the top quartile.

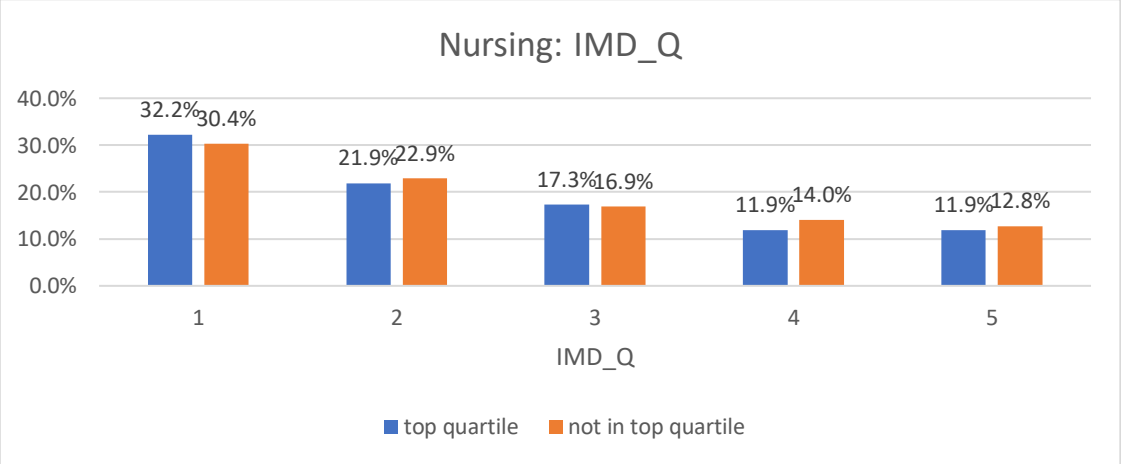


Figure 41 Proportion of babies in each IMD_Q quintile when categorising units according to Measure 4 of my non-NNAP MQC

Similar to the first measure (receipt of any dose of antenatal steroids), the Chi squared test produced a p value denoting significant difference (0.04), but the maximum difference between observed and expected values across any of the IMD_Q categories in either grouping was $\leq$0.7%, and the direction of differences was not uniform. Therefore, the finding of statistical significance was likely secondary to the large dataset and is of negligible clinical significance.

Measure 4 - babies requiring intensive care on day one of life provided with 1:1 nursing care		Group 1: Top quartile		Group 2: Not in the top quartile		
Number of units		28		85		
Unit designations	**LNU**	23	(82.1%)	47	(55.3%)	p=<0.01
	NICU (all)	5	(17.9%)	38	(44.7%)	
	NICU (surgical)	4	(14.3%)	17	(20.0%)	p=0.19
	NICU (non-surgical)	1	(3.6%)	21	(24.7%)	
Number of babies		964	(19.3%)	4022	(80.7%)	
Birthweight (weighted average - g)		1329		1308		p=0.15
Unknown or incorrect birthweight		n=12 (0.2%)				
Gestational week	**27**	111	(11.5%)	519	(12.9%)	p=0.58
	28	139	(14.4%)	627	(15.6%)	
	29	180	(18.7%)	732	(18.2%)	
	30	224	(23.2%)	924	(23.0%)	
	31	310	(32.2%)	1220	(30.3%)	
Gender	**Male**	530	(55.0%)	2216	(55.1%)	p=0.93
	Female	433	(44.9%)	1803	(44.8%)	
	Unknown gender	n=4 (0.08%)				
Multiplicity	**1**	715	(74.2%)	2934	(72.9%)	p=0.45
	≥2	248	(25.7%)	1083	(26.9%)	
	Unknown multiplicity	n=6 (0.1%)				
Significant congenital anomalies		26	(2.7%)	111	(2.8%)	p=0.91
Apgar score at 5 minutes (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=535 (10.7%)				
IMD_Q	**1 (most deprived)**	310	(32.2%)	1221	(30.4%)	p=0.04
	2	211	(21.9%)	920	(22.9%)	
	3	167	(17.3%)	679	(16.9%)	
	4	115	(11.9%)	564	(14.0%)	
	5 (least deprived)	115	(11.9%)	513	(12.8%)	
	Unknown IMD_Q	n=171 (3.4%)				
Resuscitation involving cardiac massage or adrenaline		34	(3.5%)	101	(2.5%)	p=0.08
Unknown resuscitation status		n=316 (6.3%)				

Table 28 Comparison of patient populations when categorising units according to quality of care Measure 4 (babies requiring intensive care on day one of life provided with 1:1 nursing care)

8.1.3.5 Combined measures of quality of care (non-NNAP)

When using a combination of Measures 1-4 (steroids, temperature, ventilation, nursing) to categorise units and compare them and their populations, there was a significant difference in unit designation, the number of babies with significant congenital anomalies, and the IMD_Q (Table 29, Figures 42, 43 and 44).

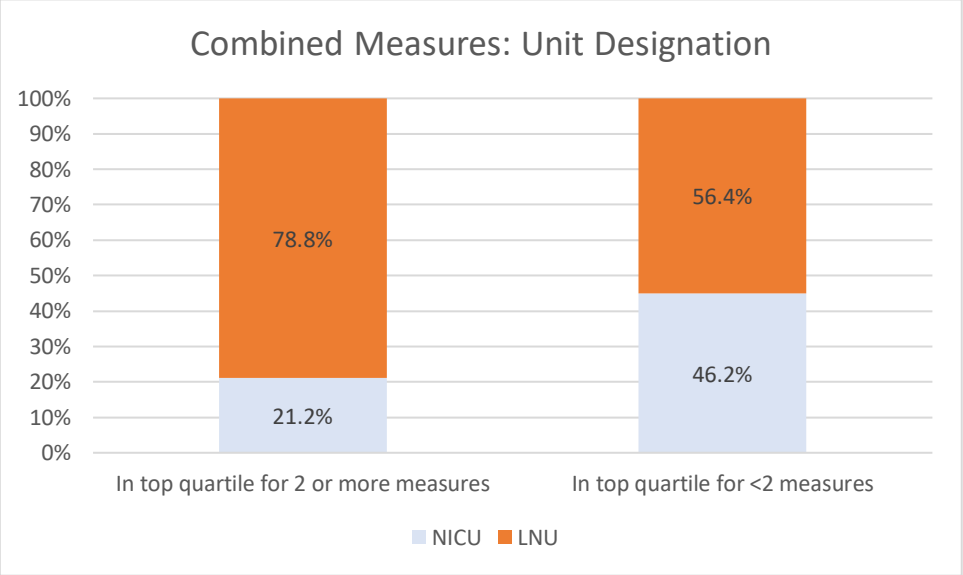


Figure 42 Proportion of babies by designation of unit of birth when categorising units according to the combined Measures of my non-NNAP MQC

The ratio of LNU:NICU was significantly greater for units within the top quartile for two or more measures compared to units within the top quartile for <2 measures.

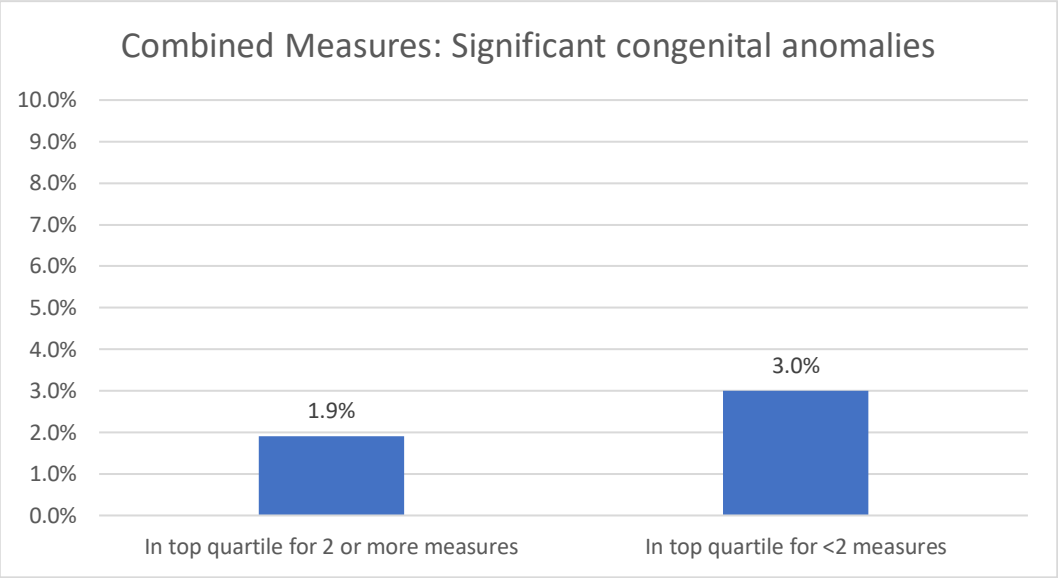


Figure 43 Proportion of babies with significant congenital anomalies when categorising units according to the combined Measures of my non-NNAP MQC

The proportion of babies with significant congenital anomalies was significantly higher in units in the top quartile for <2 measures compared to units in the top quartile for 2 or more measures.

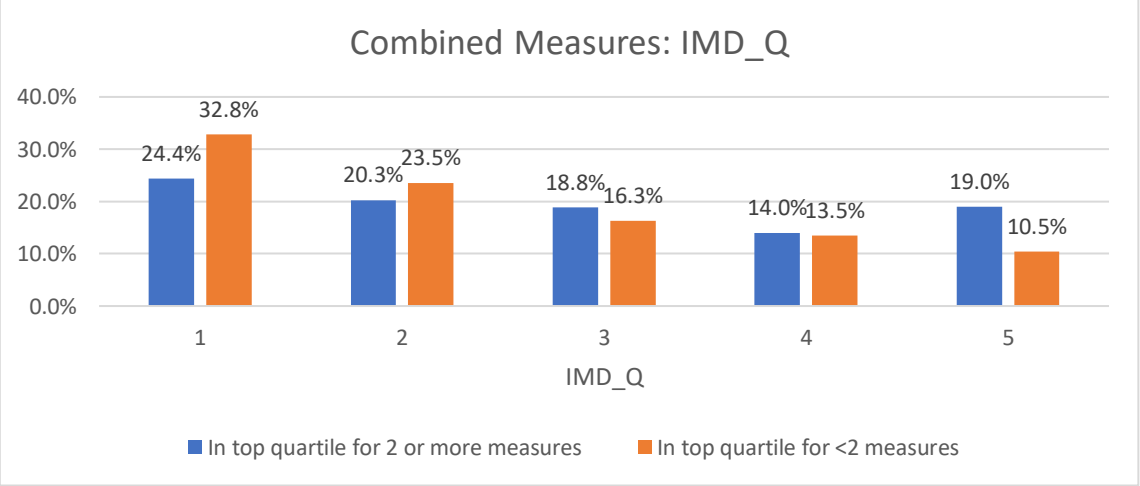


Figure 44 Proportion of babies in each IMD_Q quintile when categorising units according to the combined Measures of my non-NNAP MQC

For the IMD_Q, units within the top quartile for two or more measures had a significantly greater proportion of babies in category 5, with a reduction in the proportion of babies in category 1 and 2. I.e., there was a significant trend towards a less deprived population of patients in units within the top quartile for two or more measures, compared to within the top quartile for <2 measures.

Combined measures of quality of care (non-NNAP)		Group 1: In top quartile for 2 or more measures		Group 2: In top quartile for <2 measures		
Number of units		33		80		
Unit designations	**LNU**	26	(78.8%)	44	(55.0%)	p=0.01
	NICU (all)	7	(21.2%)	36	(45.0%)	
	NICU (surgical)	3	(9.1%)	18	(22.5%)	p=1.00
	NICU (non-surgical)	4	(12.1%)	18	(22.5%)	
Number of babies		1254	(25.2%)	3732	(74.8%)	
Average number of babies born per unit		38		47		p=0.08
Birthweight (weighted average - g)		1330		1306		p=0.08
Unknown or incorrect birthweight		n=12 (0.2%)				
Gestational week	**27**	145	(11.6%)	479	(13.0%)	p=0.76
	28	200	(15.9%)	566	(15.2%)	
	29	231	(18.4%)	674	(18.2%)	
	30	288	(23.0%)	850	(23.0%)	
	31	390	(31.1%)	1128	(30.5%)	
Gender	**Male**	718	(57.3%)	2028	(54.3%)	p=0.07
	Female	535	(42.7%)	1701	(45.6%)	
	Unknown gender	n=4 (0.08%)				
Multiplicity	**1**	938	(74.9%)	2711	(72.6%)	p=0.13
	≥2	314	(25.0%)	1017	(27.3%)	
	Unknown multiplicity	n=6 (0.1%)				
Significant congenital anomalies		24	(1.9%)	113	(3.0%)	p=0.04
Apgar score at 5 minutes (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=535 (10.7%)				
IMD_Q	**1 (most deprived)**	306	(24.4%)	1225	(32.8%)	p=<0.01
	2	254	(20.3%)	877	(23.5%)	
	3	236	(18.8%)	610	(16.3%)	
	4	175	(14.0%)	504	(13.5%)	
	5 (least deprived)	238	(19.0%)	390	(10.5%)	
	Unknown IMD_Q	n=171 (3.4%)				
Resuscitation involving cardiac massage or adrenaline		25	(2.0%)	110	(2.9%)	p=0.07
Unknown resuscitation status		n=316 (6.3%)				

Table 29 Comparison of patient populations when categorising units according to combination of quality of care Measures 1-4 (steroids, temperature, ventilation, nursing)

8.1.4 Associations between adherence with non-NNAP MQC and outcomes using univariate analyses

8.1.4.1 Pre-discharge mortality

8.1.4.1.9 Pre-discharge mortality by gestational week of birth

Pre-discharge mortality decreased from 7.7% for babies born at 27 weeks of gestation, to 1.5% for babies born at 31 weeks of gestation (Table 30, Figure 45). The fall in mortality was greatest from 27 to 29 weeks, with a nearly threefold reduction. For the entire cohort it was 3.3%.

Gestational week at birth	Mortality (pre-discharge)	Patient total	(%)
27	49	634	(7.7%)
28	42	775	(5.4%)
29	25	919	(2.7%)
30	26	1161	(2.2%)
31	24	1549	(1.5%)
27-31	166	5038	(3.3%)

Table 30 Pre-discharge mortality by gestational week of birth

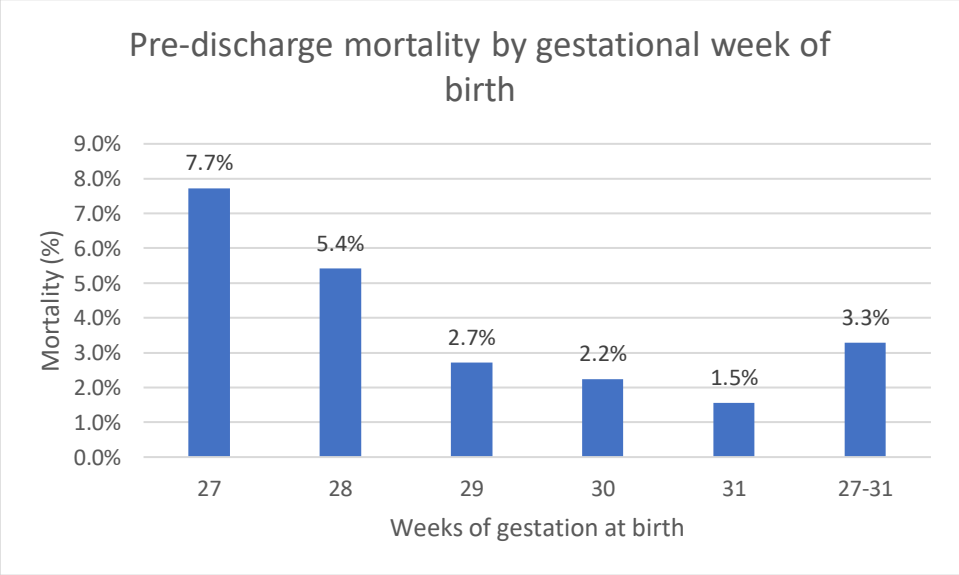


Figure 45 Pre-discharge mortality by gestational week of birth

8.1.4.1.10 Mortality by designation of unit of birth

When separating pre-discharge mortality for babies by gestational week, and designation of unit of birth (NICU vs. LNU), a consistent trend was seen for increased mortality across the gestational age range, and this was strongly statistically significant for the whole

cohort (Table 31, Figure 46). However, by each gestational week, this difference was only statistically significant for babies born at 31 weeks.

Gestational week at birth	NICU			LNU			
	Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
27	36	446	(8.1%)	13	184	(7.1%)	p=0.68
28	30	457	(6.6%)	12	309	(3.9%)	p=0.12
29	16	485	(3.3%)	9	427	(2.1%)	p=0.28
30	18	625	(2.9%)	8	523	(1.5%)	p=0.13
31	18	791	(2.3%)	6	739	(0.8%)	p=0.02
27-31	118	2804	(4.2%)	48	2182	(2.2%)	p=<0.01

Table 31 Pre-discharge mortality for babies by gestational week of birth and designation of unit of birth

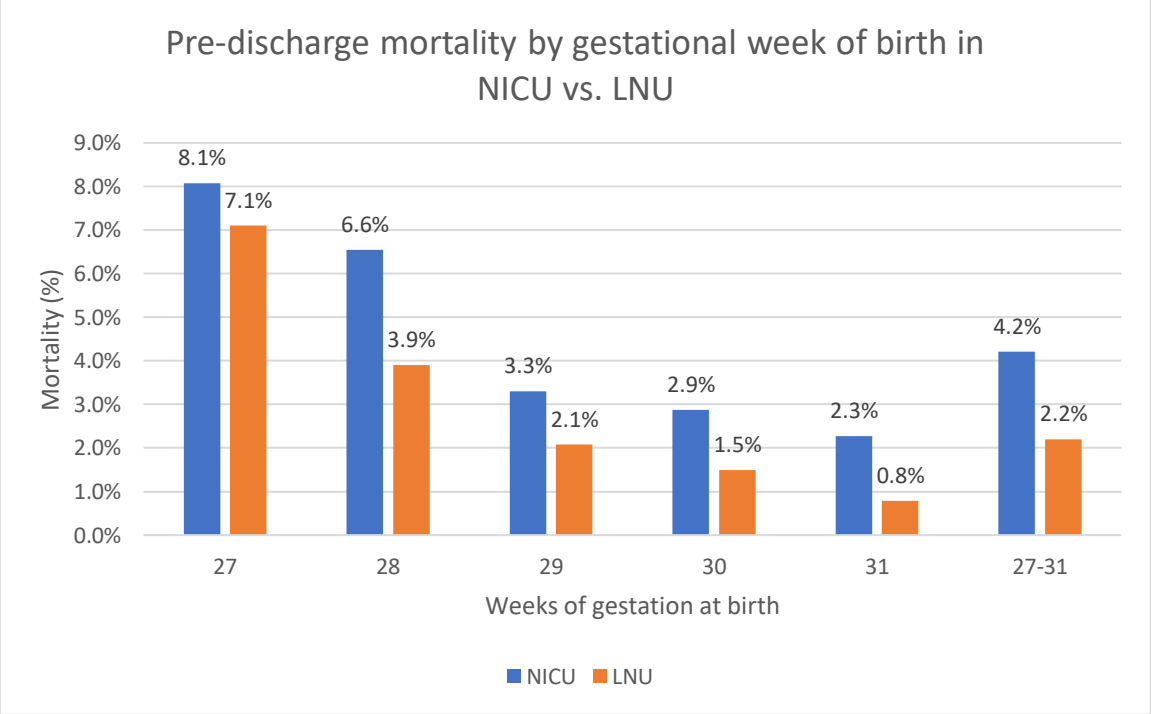


Figure 46 Pre-discharge mortality by gestational week and designation of unit of birth

When comparing mortality in surgical NICU versus non-surgical NICU, it was significantly higher for babies born at 29 and 31 weeks, and for the whole cohort (Table 32, Figure 47).

Gestational week at birth	Surgical NICU			Non-surgical NICU			
	Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
27	17	212	(8.0%)	19	234	(8.1%)	p=0.97
28	19	227	(8.4%)	11	230	(4.8%)	p=0.13
29	13	273	(4.5%)	3	212	(1.4%)	p=0.04
30	10	322	(3.1%)	8	303	(2.6%)	p=0.73
31	14	422	(3.3%)	4	369	(1.1%)	p=0.04
27-31	73	1456	(5.0%)	45	1348	(3.3%)	p=0.03

Table 32 Pre-discharge mortality for babies by gestational week of birth in surgical NICU versus non-surgical NICU

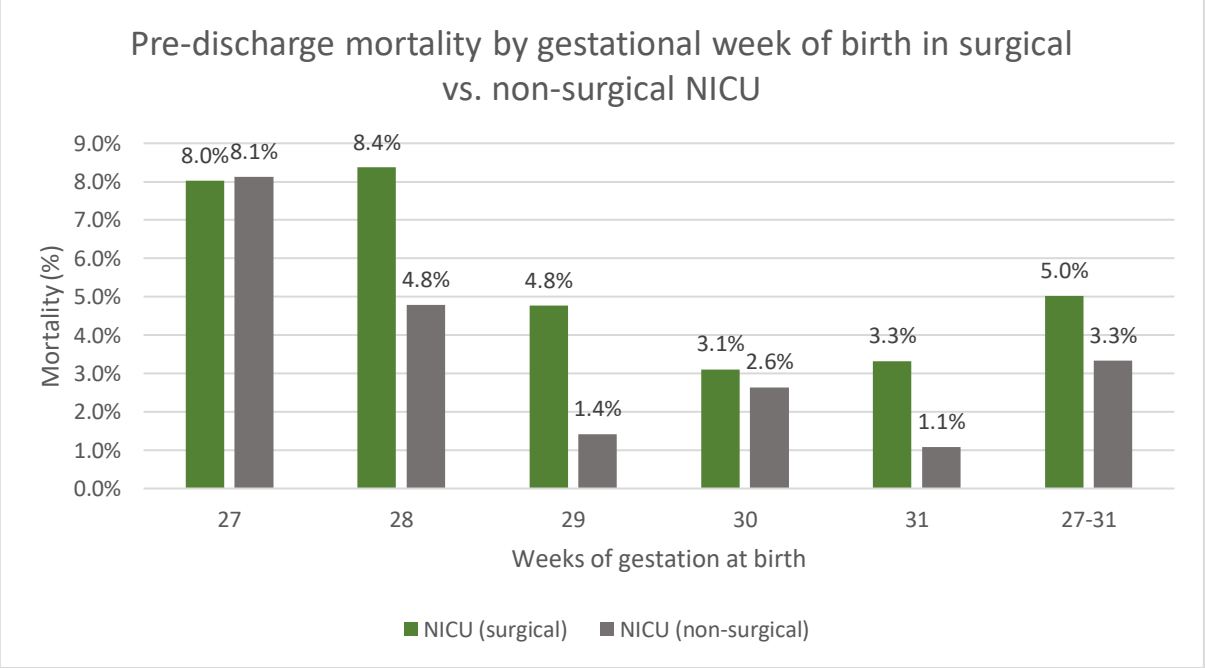


Figure 47 Pre-discharge mortality by gestational week of birth for surgical versus non-surgical NICU

When comparing non-surgical NICU with LNU, I found a significantly higher mortality for whole cohort but not for any specific gestational week (Table 33, Figure 48).

Gestational week at birth	Non-surgical NICU			LNU			
	Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
27	19	234	(8.1%)	13	184	(7.1%)	p=0.70
28	11	230	(4.8%)	12	309	(3.9%)	p=0.62
29	3	212	(1.4%)	9	427	(2.1%)	p=0.55
30	8	303	(2.6%)	8	523	(1.5%)	p=0.27
31	4	369	(1.1%)	6	739	(0.8%)	p=0.65
27-31	45	1348	(3.3%)	48	2182	(2.2%)	p=0.04

Table 33 Pre-discharge mortality for babies by gestational week of birth in non-surgical NICU versus LNU

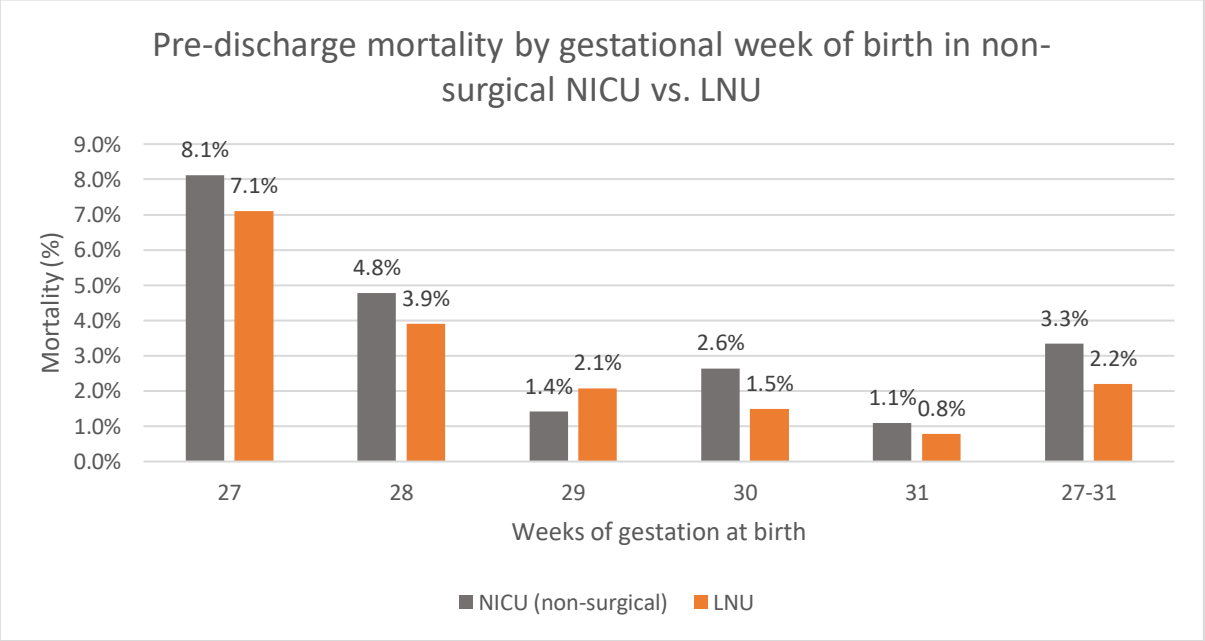


Figure 48 Pre-discharge mortality by gestational week of birth for non-surgical NICU versus LNU

8.1.4.1.11 Mortality by adherence with non-NNAP MQC

Comparing pre-discharge mortality of Group 1 versus Group 2 for adherence with Measures 1-4 and the combination of Measures, the only significant difference found was an increase in mortality for Measure 4, for babies born at 27 weeks of gestation (Table 34 and 32, Figures 49 - 53).

Non-NNAP measure of quality of care	Gestational week at birth	Group 1: Units in top quartile for adherence			Group 2: Units not in top quartile for adherence			
		Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
Measure 1 - receipt of any dose of antenatal steroids	27	12	143	(8.4%)	37	487	(7.6%)	p=0.76
	28	9	195	(4.6%)	33	571	(5.8%)	p=0.55
	29	2	205	(1.0%)	23	707	(3.3%)	p=0.08
	30	5	279	(1.8%)	21	869	(2.4%)	p=0.55
	31	3	355	(0.8%)	21	1175	(1.8%)	p=0.21
	27-31	31	1177	(2.6%)	135	3809	(3.5%)	p=0.13
Measure 2 - normal temperature recorded within one hour of admission	27	15	149	(10.1%)	34	481	(7.1%)	p=0.25
	28	13	188	(6.9%)	29	578	(5.0%)	p=0.33
	29	4	212	(1.9%)	21	700	(3.0%)	p=0.39
	30	4	281	(1.4%)	22	867	(2.5%)	p=0.28
	31	3	372	(0.8%)	21	1158	(1.8%)	p=0.18
	27-31	39	1202	(3.2%)	127	3784	(3.4%)	p=0.85
Measure 3 - babies requiring ventilatory support on day one of life supported with non-invasive ventilation	27	7	136	(5.1%)	42	494	(8.5%)	p=0.21
	28	9	158	(5.7%)	33	608	(5.4%)	p=0.90
	29	5	199	(2.5%)	20	713	(2.8%)	p=0.83
	30	8	267	(3.0%)	18	881	(2.0%)	p=0.36
	31	4	381	(1.0%)	20	1149	(1.7%)	p=0.35
	27-31	33	1141	(2.9%)	133	3845	(3.5%)	p=0.36
Measure 4 - babies requiring intensive care on day one of life provided with 1:1 nursing care	27	15	111	(13.5%)	34	519	(6.6%)	p=0.02
	28	3	139	(2.2%)	39	627	(6.2%)	p=0.06
	29	6	180	(3.3%)	19	732	(2.6%)	p=0.59
	30	7	224	(3.1%)	19	924	(2.1%)	p=0.34
	31	8	310	(2.6%)	16	1220	(1.3%)	p=0.11
	27-31	39	964	(4.0%)	127	4022	(3.2%)	p=0.17

Table 34 Pre-discharge mortality for babies by gestational week of birth when categorising units based on adherence with non-NNAP measures of quality of care

Non-NNAP measure of quality of care	Gestational week at birth	Group 1: In top quartile for 2 or more measures			Group 2: In top quartile for <2 measures			
		Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
Combined measures of quality of care	27	16	145	(11.0%)	33	485	(6.8%)	p=0.11
	28	9	200	(4.5%)	33	566	(5.8%)	p=0.49
	29	5	231	(2.2%)	20	681	(2.9%)	p=0.54
	30	4	288	(1.4%)	22	860	(2.6%)	p=0.25
	31	2	390	(0.5%)	22	1140	(1.9%)	p=0.05
	27-31	36	1254	(2.9%)	130	3732	(3.5%)	p=0.30

Table 35 Pre-discharge mortality for babies by gestational week of birth when categorising units based on adherence with non-NNAP measures of quality of care

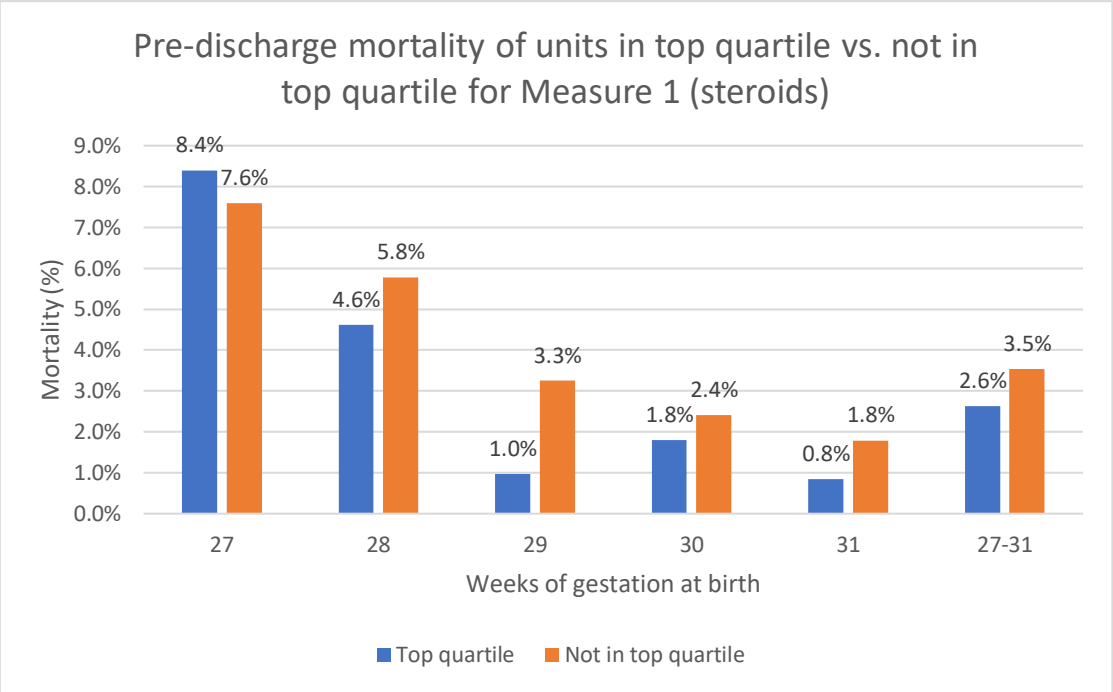


Figure 49 Pre-discharge mortality by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 1 (steroids) of my non-NNAP MQC

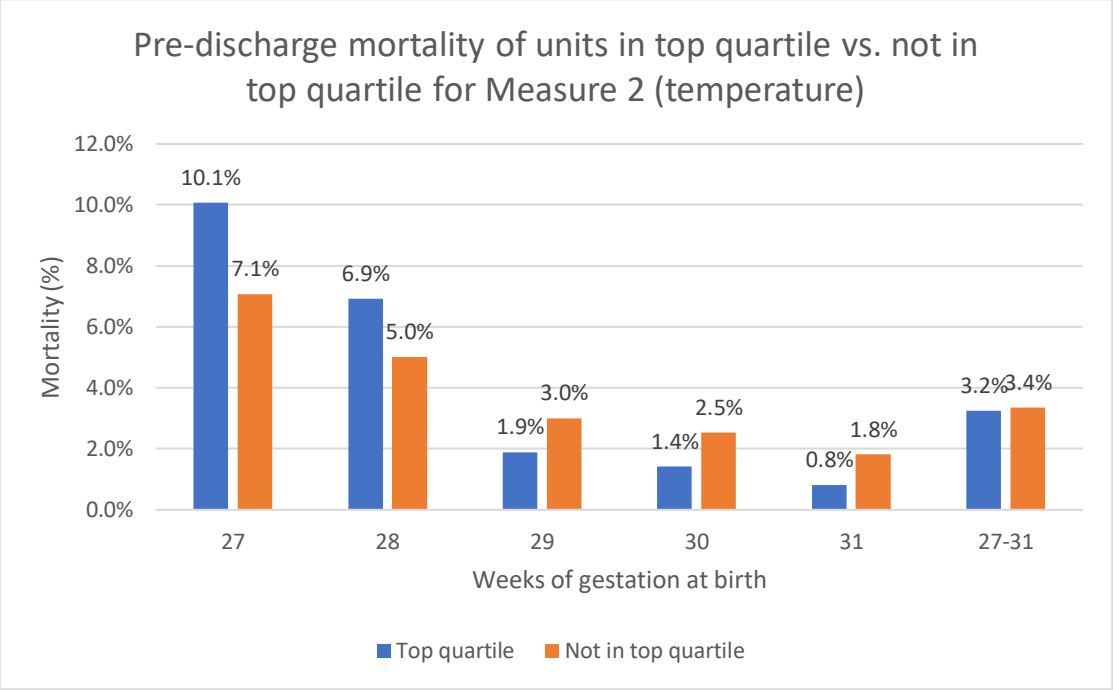


Figure 50 Pre-discharge mortality by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 2 (temperature) of my non-NNAP MQC

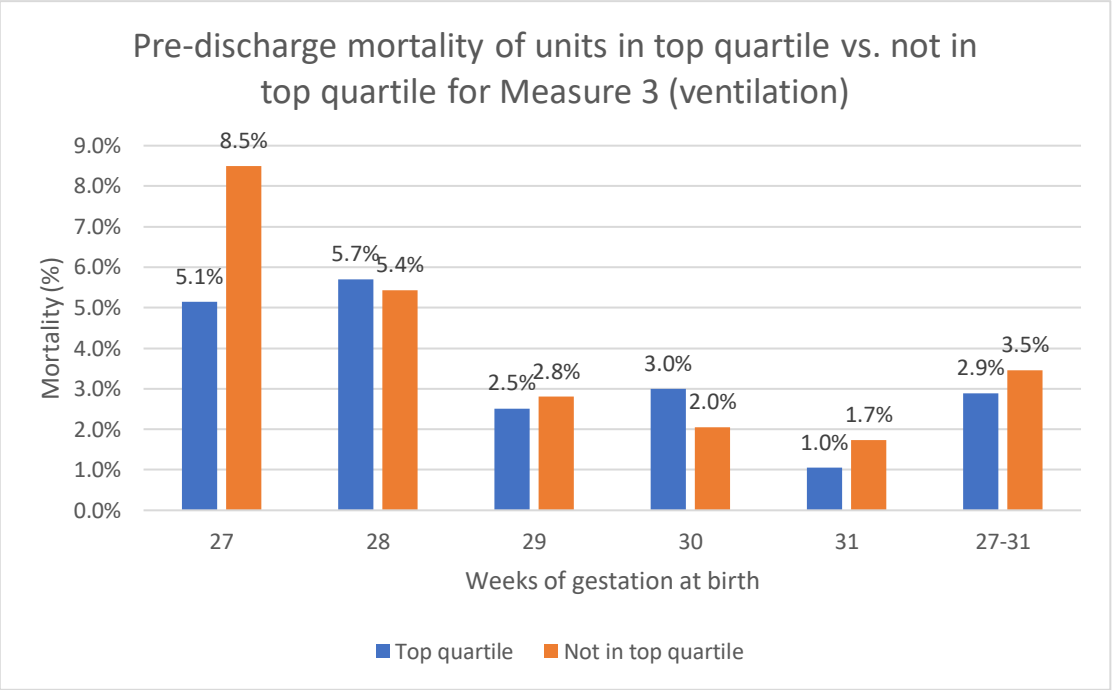


Figure 51 Pre-discharge mortality by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 3(ventilation) of my non-NNAP MQC

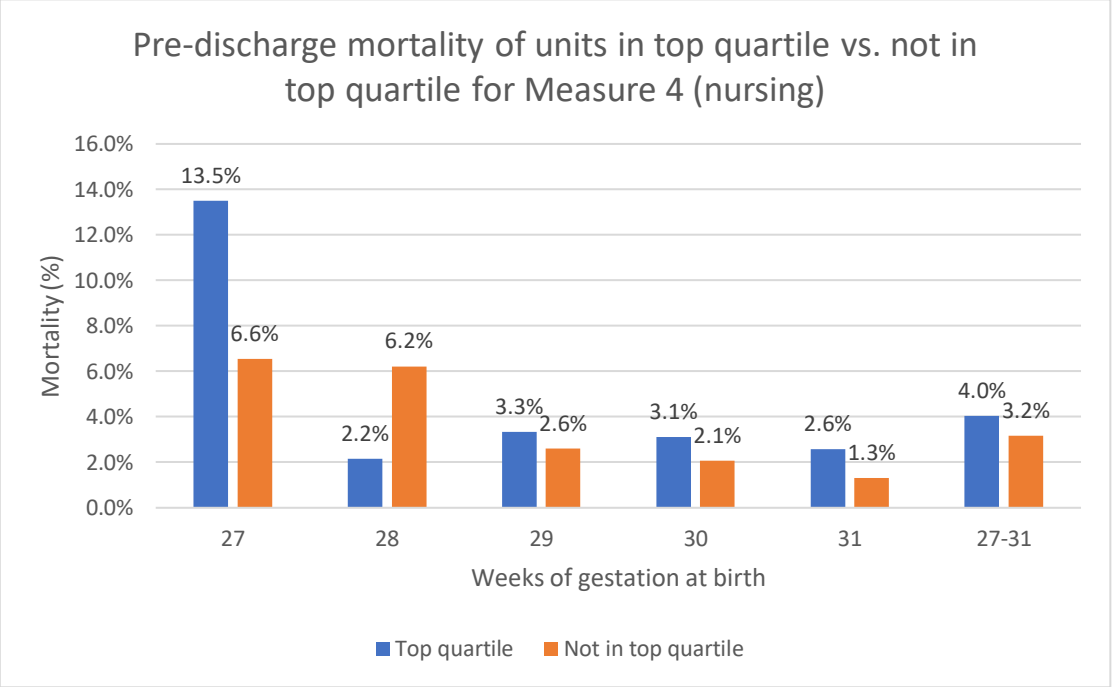


Figure 52 Pre-discharge mortality by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 4 (nursing) of my non-NNAP MQC

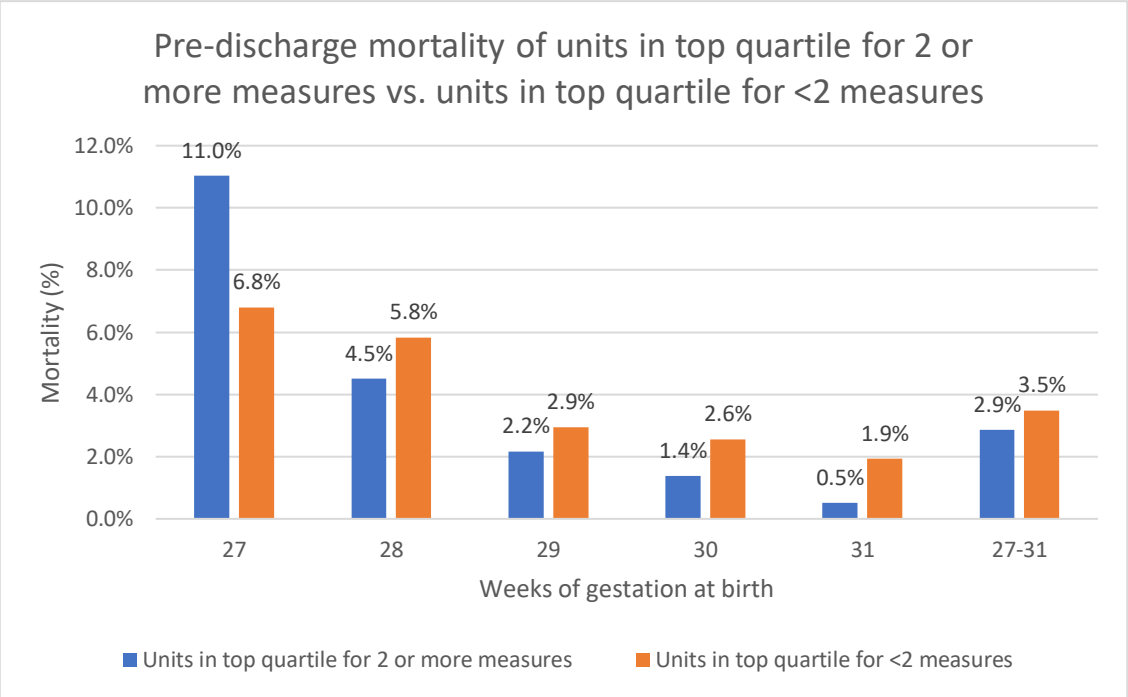


Figure 53 Pre-discharge mortality by gestational week of birth for units in top quartile for 2 or more measures versus units in top quartile for <2 measures for the combined Measures (steroids, temperature, ventilation, nursing) of my non-NNAP MQC

8.1.4.2 Length of stay (LOS)

8.1.4.2.12 LOS by gestational week and designation of unit of birth

LOS more than doubled, from a mean of 36 days for babies born at 31 weeks of gestation, to 84 days for babies born at 27 weeks of gestation (Table 36, Figure 54). For the entire cohort it was 54 days. When separating LOS for babies by gestational week, and designation of unit of birth, a consistent trend was seen for increased LOS in NICU vs. LNU across the gestational age range (Figure 55). This was statistically significant for the whole cohort, as well as for babies born at 27, 30 and 31 weeks of gestation.

Gestational week at birth	Mean length of stay (LOS – days)				
	All units (SD)	NICU	LNU	Difference (days, 95% CI)	
27	84 (27)	86	80	6.6 (2.5 – 10.6)	p=<0.01
28	71 (25)	72	70	2.2 (-2.9 – 7.3)	p=0.40
29	58 (23)	60	57	3.1 (-0.8 – 6.9)	p=0.10
30	47 (18)	49	46	3.1 (0.5 – 5.6)	p=0.02
31	36 (15)	38	35	2.9 (0.4 – 5.4)	p=0.02
27-31	54 (26)	57	50	6.9 (4.9 – 9.0)	p=<0.01

Table 36 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth and designation of unit of birth

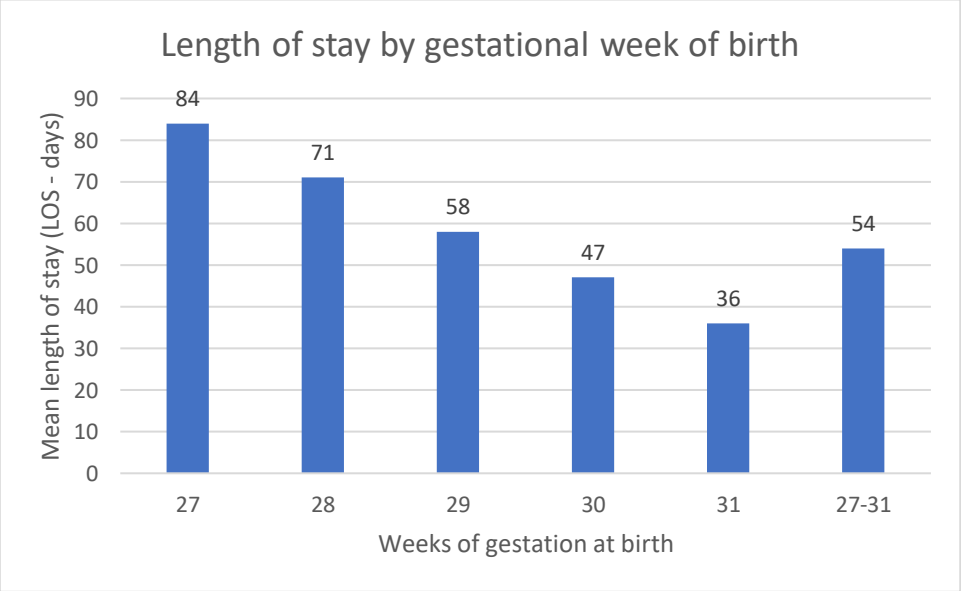


Figure 54 Length of stay by gestational week of birth

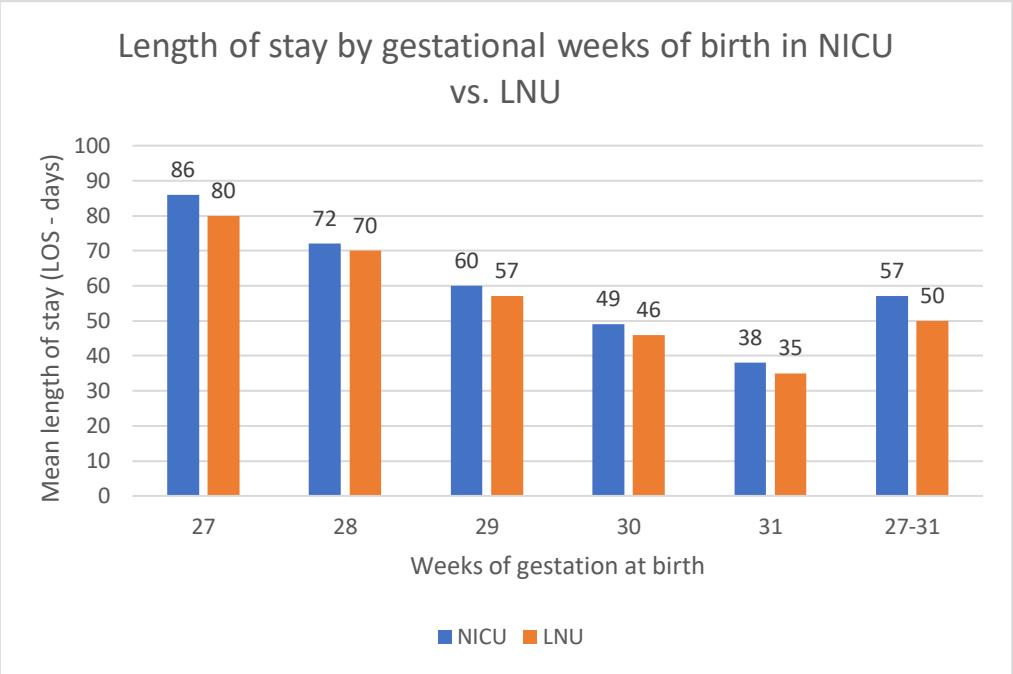


Figure 55 Length of stay by gestational week and designation of unit of birth

Comparing LOS for babies born in surgical NICU versus non-surgical NICU, it was significantly longer for babies born at 31 weeks of gestation but not the whole cohort (Table 37, Figure 56).

Gestational week at birth	Mean length of stay (LOS – days)			
	Surgical NICU	Non-surgical NICU	Difference (days, 95% CI)	
27	86	85	0.9 (-5.6 – 7.4)	p=0.78
28	74	71	3.5 (-4.0 – 10.9)	p=0.34
29	61	59	1.4 (-5.3 – 8.1)	p=0.68
30	50	47	3.3 (-0.5 – 7.0)	p=0.08
31	40	36	4.3 (0.3 – 8.4)	p=0.03
27-31	58	56	1.8 (-1.9 – 5.4)	p=0.32

Table 37 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth in surgical NICU versus non-surgical NICU

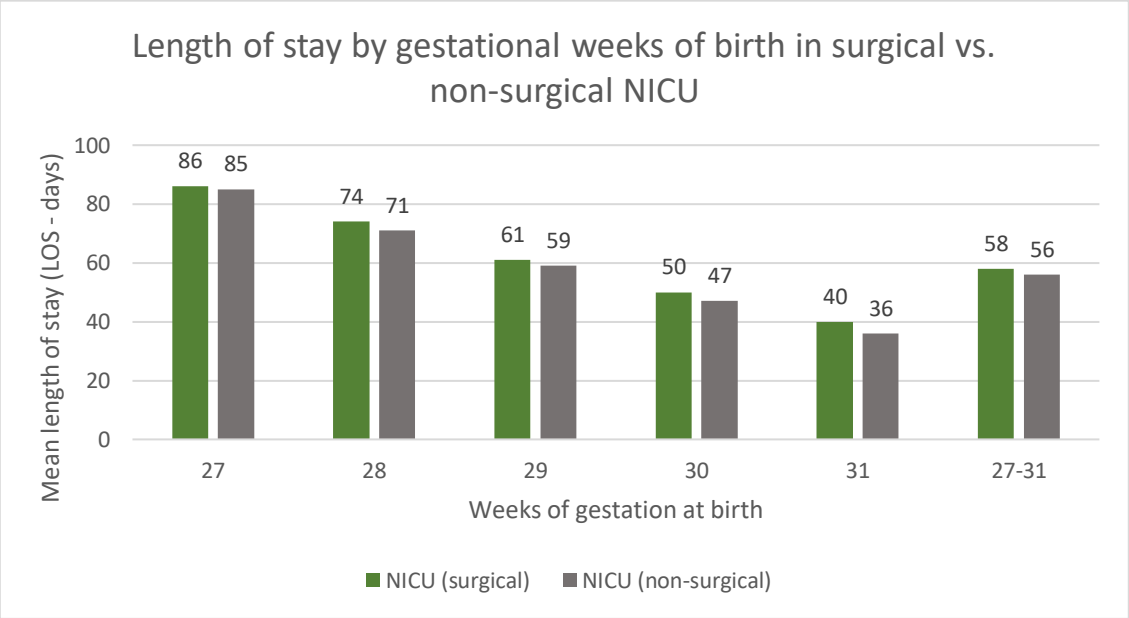


Figure 56 Length of stay by gestational week of birth for surgical versus non-surgical NICU

Comparing LOS for babies born in non-surgical NICU versus LNU, it was significantly longer for the whole cohort but not any specific gestational week (Table 38, Figure 57).

Gestational week at birth	Mean length of stay (LOS – days)			
	Non-surgical NICU	LNU	Difference (days, 95% CI)	
27	85	80	5.4 (-1.2 – 11.9)	p=0.10
28	71	70	1.0 (-4.6 – 6.6)	p=0.71
29	59	57	2.5 (-1.7 – 6.8)	p=0.23
30	47	46	1.3 (-2.0 – 4.6)	p=0.42
31	36	35	1.1 (-1.5 – 3.6)	p=0.40
27-31	56	50	6.0 (3.2 – 8.7)	p=<0.01

Table 38 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth in non-surgical NICU versus LNU

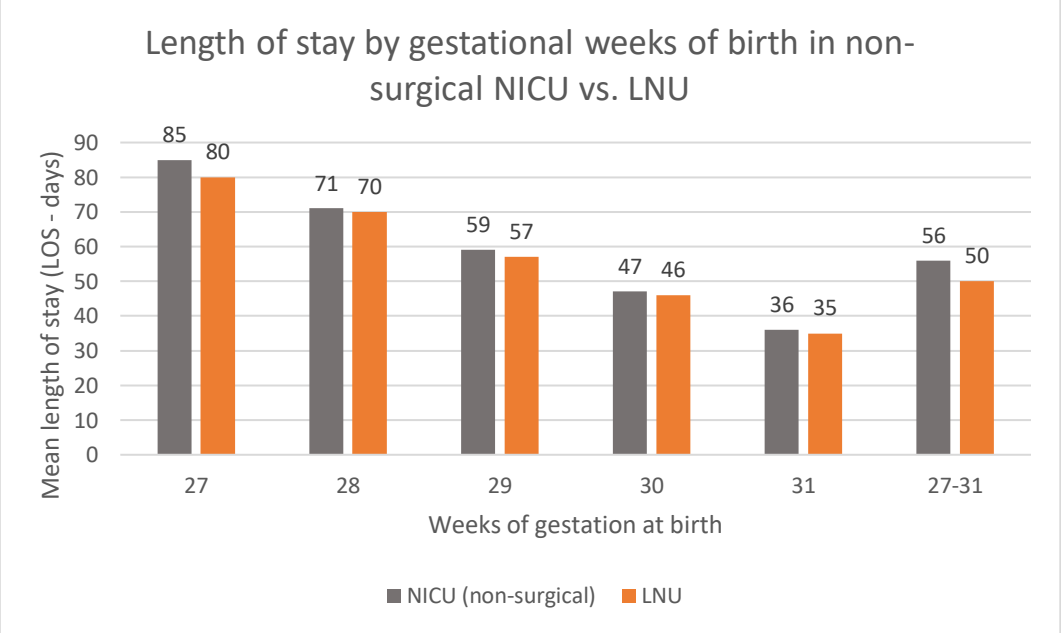


Figure 57 Length of stay by gestational week of birth for non-surgical NICU versus LNU

8.1.4.2.13 LOS by adherence with non-NNAP MQC

Comparing LOS of Group 1 versus Group 2 for adherence with Measures 1-4, several significant differences were found (Table 39 and 40, Figures 58 - 62).

For Measure 3 (ventilation), LOS for the whole cohort of babies in Group 1 was significantly less than for Group 2 (difference in weighted mean LOS 3.5 days, 95% CI 0.6 – 6.3, p=0.01). By each gestational week of birth, the difference in LOS for babies born at 27 weeks was also significant (difference in weighted mean LOS 5.3 days, 95% CI 0.15 – 10.37, p=0.04).

Non-NNAP measure of quality of care	**Gestational week at birth**	**Mean length of stay (LOS – days)**			
		Group 1: Units in top quartile for adherence	**Group 2: Units not in top quartile for adherence**	**Difference (days, 95% CI)**	
Measure 1 - receipt of any dose of antenatal steroids	27	82	85	2.3 (-2.8 – 7.5)	p=0.36
	28	74	70	4.0 (-1.1 – 9.2)	p=0.11
	29	59	58	0.2 (-4.2 – 4.6)	p=0.93
	30	49	47	1.8 (-1.2 – 4.7)	p=0.22
	31	37	36	0.1 (-2.6 – 2.7)	p=0.96
	27-31	55	54	0.9 (-2.0 – 3.8)	p=0.53
Measure 2 - normal temperature recorded within one hour of admission	27	81	85	3.5 (-1.6 – 8.6)	p=0.16
	28	70	72	2.0 (-3.2 – 7.2)	p=0.43
	29	58	58	0.2 (-4.0 – 4.5)	p=0.91
	30	47	47	0.1 (-2.9 – 3.1)	p=0.94
	31	35	37	1.9 (-0.7 – 4.4)	p=0.14
	27-31	53	54	1.7 (-1.1 – 4.6)	p=0.22
Measure 3 - babies requiring ventilatory support on day one of life supported with non-invasive ventilation	27	80	85	5.3 (0.2 – 10.4)	p=0.04
	28	67	72	4.8 (-0.3 – 10.0)	p=0.06
	29	56	59	2.5 (-1.8 – 6.8)	p=0.23
	30	47	47	0.2 (-2.8 – 3.2)	p=0.91
	31	35	37	1.7 (-0.9 – 4.3)	p=0.19
	27-31	51	55	3.5 (0.6 – 6.3)	p=0.01
Measure 4 - babies requiring intensive care on day one of life provided with 1:1 nursing care	27	82	84	2.9 (-2.4 – 8.2)	p=0.26
	28	72	71	0.6 (-4.7 – 5.8)	p=0.82
	29	58	59	0.9 (-3.3 – 5.2)	p=0.65
	30	46	47	1.3 (-1.7 – 4.3)	p=0.38
	31	35	37	2.0 (-0.5 – 4.6)	p=0.11
	27-31	52	54	2.3 (-0.5 – 5.2)	p=0.10

Table 39 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth when categorising units based on adherence with non-NNAP measures of quality of care

For the combined measures (steroids, temperature, ventilation, nursing), LOS for the whole cohort of babies in Group 1 was significantly less than for Group 2 (difference in weighted mean LOS 3.1 days, 95% CI 0.4 – 5.8, p=0.02). By each gestational week of birth, the difference in LOS for babies born at 27 weeks (difference in weighted mean LOS 5.6 days, 95% CI 0.7 – 10.4, p=0.02), and 31 weeks (difference in weighted mean LOS 3.4 days, 95% CI 1.0 – 5.8, p=<0.01), was also significant.

Non-NNAP measure of quality of care	Gestational week at birth	Mean length of stay (LOS – days)			
		Group 1: In top quartile for 2 or more measures	Group 2: In top quartile for <2 measures	Difference (days, 95% CI)	
Combined measures of quality of care (steroids, temperature, ventilation, nursing)	27	80	85	5.6 (0.7 – 10.4)	p=0.02
	28	69	72	2.4 (-2.5 – 7.4)	p=0.32
	29	57	59	1.4 (-2.7 – 5.4)	p=0.50
	30	47	47	0.9 (-1.9 – 3.8)	p=0.51
	31	34	37	3.4 (1.0 – 5.8)	p=<0.01
	27-31	52	55	3.1 (0.4 – 5.8)	p=0.02

Table 40 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth when categorising units based on adherence with non-NNAP measures of quality of care

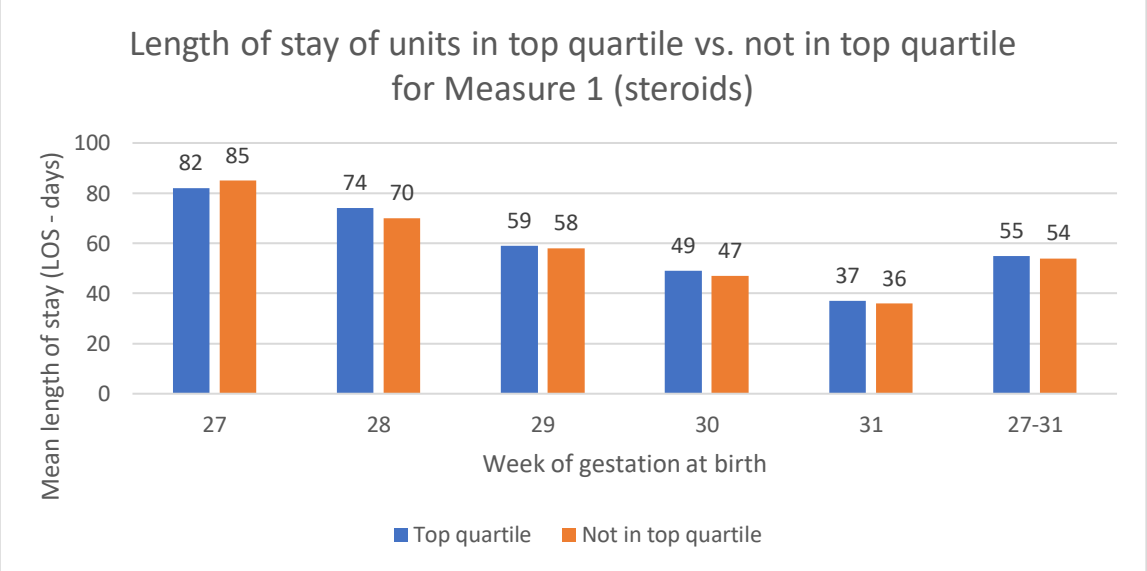


Figure 58 Length of stay by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 1 (steroids) of my non-NNAP MQC

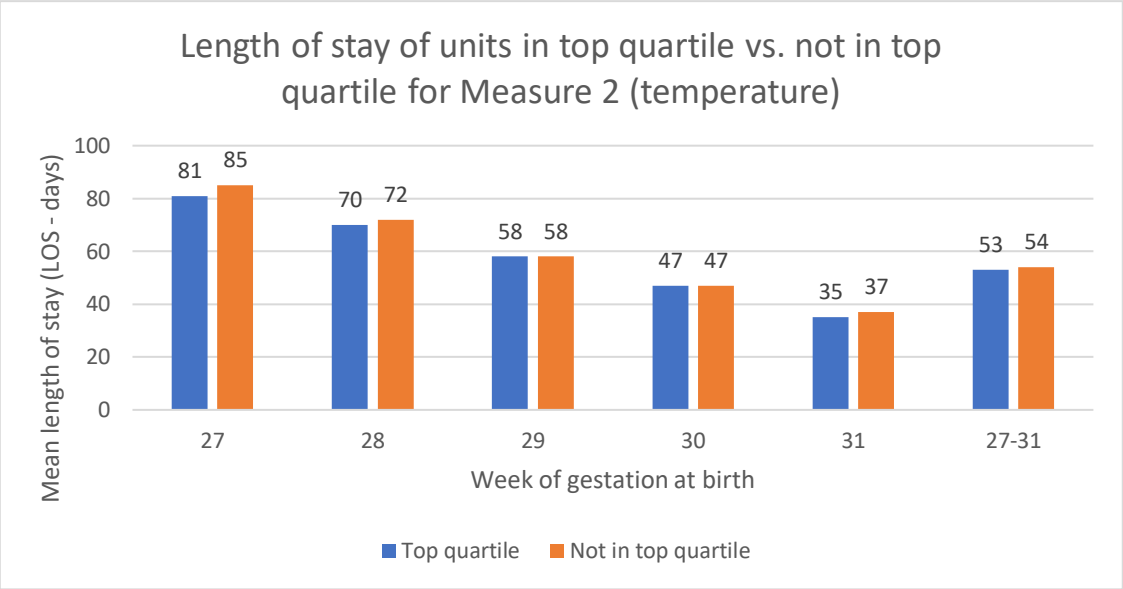


Figure 59 Length of stay by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 2 (temperature) of my non-NNAP MQC

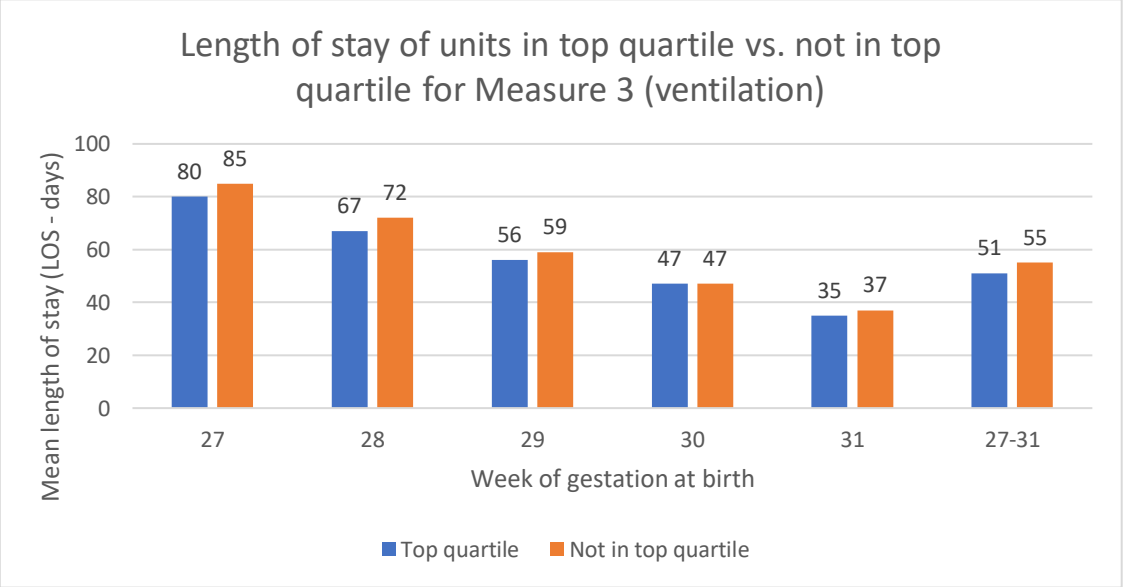


Figure 60 Length of stay by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 3 (ventilation) of my non-NNAP MQC

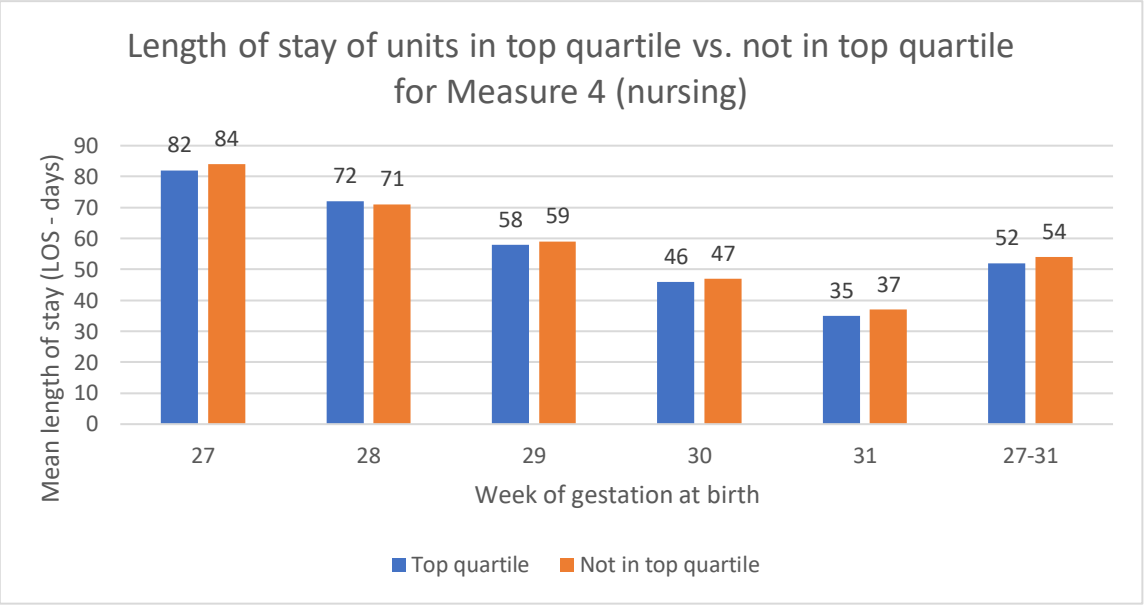


Figure 61 Length of stay by gestational week of birth for units in top quartile (group 1) versus units not in top quartile (group 2) for Measure 4 (nursing) of my non-NNAP MQC

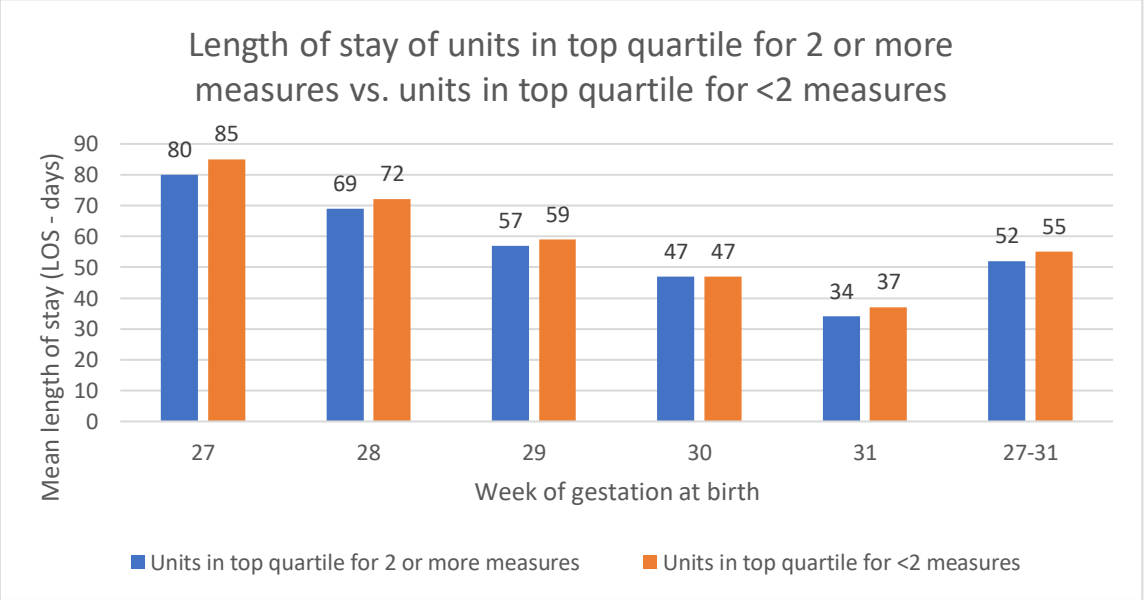


Figure 62 Length of stay by gestational week of birth for units in top quartile for 2 or more measures versus units in top quartile for <2 measures for the combined Measures (steroids, temperature, ventilation, nursing) of my non-NNAP MQC

Graphical representation of these results (not displayed and discussed above) can be found in Appendix V.

8.2 NNAP audit measures

8.2.1 Comparing patient populations when categorising units based on adherence with audit measures

Comparing units and babies (born between 27-31 weeks of gestation) within Group 1 and Group 2 following categorisation of units by adherence with NNAP audit measures, the only significant difference was in the IMD_Q (Table 41).

Adherence with NNAP audit measures		Group 1: Meeting threshold/in top quartile for 4 or more measures		Group 2: Meeting threshold/in top quartile for <4 measures		
Number of units		21		77		
Unit designations	**LNU**	15	(71.4%)	41	(53.2%)	p=0.14
	NICU (all)	6	(28.6%)	36	(46.8%)	
	NICU (surgical)	2	(9.5%)	18	(23.4%)	p=0.67
	NICU (non-surgical)	4	(19.0%)	18	(23.4%)	
Number of babies		889	(19.4%)	3705	(80.6%)	
Birthweight (weighted average - g)		1314		1305		p=0.94
Unknown or incorrect birthweight		n=12 (0.3%)				
Gestational week	**27**	102	(11.5%)	498	(13.4%)	p=0.14
	28	124	(13.9%)	592	(16.0%)	
	29	162	(18.2%)	680	(18.4%)	
	30	207	(23.3%)	833	(22.5%)	
	31	294	(33.1%)	1102	(29.7%)	
Gender	**male**	481	(54.1%)	2065	(55.7%)	p=0.37
	female	408	(45.9%)	1636	(44.2%)	
	Unknown gender	n=4 (0.1%)				
Multiplicity	**1**	658	(74.0%)	2700	(72.9%)	p=0.55
	≥2	231	(26.0%)	999	(27.0%)	
	Unknown multiplicity	n=6 (0.55%)				
Significant congenital anomalies		18	(2.0%)	113	(3.0%)	p=0.08
Apgar score at 5min (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=6 (0.1%)				
IMD_Q	**1 (most deprived)**	188	(21.1%)	1234	(33.3%)	p=<0.01
	2	214	(24.1%)	799	(21.6%)	
	3	173	(19.5%)	610	(16.5%)	
	4	141	(15.9%)	492	(13.3%)	
	5 (least deprived)	155	(17.4%)	435	(11.7%)	
	Unknown IMD_Q	n=153 (3.3%)				
Resuscitation involving cardiac massage or adrenaline		18	(2.0%)	107	(2.9%)	p=0.16
Unknown resuscitation status		n= 303 (6.6%)				

Table 41 Comparison of patient populations when categorising units according to adherence with NNAP audit measures

8.2.2 Comparing patient populations when categorising units based on missing data for audit measures

Comparing units and babies (born between 27-31 weeks of gestation) within Group 1 and Group 2 following categorisation of units by missing data for NNAP audit measures, the only significant differences were in gender and the IMD_Q (Table 42).

Missing data for NNAP audit measures		Group 1: In top quartile/any missing data for 2 or more measures		Group 2: In top quartile/any missing data for <2 measures		
Number of units		17		83		
Unit designations	**LNU**	10	(58.8%)	47	(56.6%)	p=1.00
	NICU (all)	7	(41.2%)	36	(43.4%)	
	NICU (surgical)	4	(23.5%)	17	(20.5%)	p=0.70
	NICU (non-surgical)	3	(17.6%)	19	(22.9%)	
Number of babies		891	(18.9%)	3831	(81.1%)	
Birthweight (weighted average - g)		1306		1308		p=0.99
Unknown or incorrect birthweight		n=12 (0.3%)				
Gestational week	**27**	97	(10.9%)	512	(13.4%)	p=0.26
	28	146	(16.4%)	588	(15.3%)	
	29	166	(18.6%)	701	(18.3%)	
	30	196	(22.0%)	880	(23.0%)	
	31	286	(32.1%)	1150	(30.0%)	
Gender	**male**	460	(51.6%)	2139	(55.8%)	p=0.02
	female	430	(48.3%)	1689	(44.1%)	
	Unknown gender	n=4 (0.1%)				
Multiplicity	**1**	662	(74.3%)	2793	(72.9%)	p=0.40
	≥2	228	(25.6%)	1033	(27.0%)	
	Unknown multiplicity	n=6 (0.1%)				
Significant congenital anomalies		29	(3.3%)	104	(2.7%)	p=0.35
Apgar score at 5min (weighted average of medians)		9		9		
Unknown Apgar score at 5 minutes		n=535 (10.7%)				
IMD_Q	**1 (most deprived)**	300	(33.7%)	1177	(30.7%)	p=0.01
	2	223	(25.0%)	849	(22.2%)	
	3	126	(14.1%)	660	(17.2%)	
	4	99	(11.1%)	539	(14.1%)	
	5 (least deprived)	113	(12.7%)	473	(12.3%)	
	Unknown IMD_Q	n=163 (3.5%)				
Resuscitation involving cardiac massage or adrenaline		18	(2.0%)	111	(2.9%)	p=0.15
Unknown resuscitation status		n=300 (6.4%)				

Table 42 Comparison of patient populations when categorising units according to missing data for NNAP audit measures

8.2.3 Comparing NNAP data for adherence and completion

Comparing units categorised according to adherence and missing data for NNAP audit measures, out of the 21 units in Group 1 for adherence, 18 were in Group 2 for missing data (one unit excluded for missing data, two units in Group 1). Similarly, out of the 17 units in Group 1 for missing data, 12 were in Group 2 for adherence (three units excluded for adherence, two units in Group 1). I.e., only two units were present in both Group 1 for adherence and missing data. While the unit lists for adherence and missing data were not identical (due to differential exclusion criteria), they were largely the same (with 96 units found in both lists).

8.2.4 Associations between NNAP audit measures and outcomes using univariate analyses

8.2.4.1 Pre- discharge mortality by adherence with NNAP audit measures

Comparing pre-discharge mortality of Group 1 versus Group 2 for adherence with NNAP audit measures, it was significantly lower for the whole cohort of babies born between 27-31 weeks, but not for any single gestational week (Table 43, Figure 63).

Gestational week at birth	Group 1: Meeting threshold/in top quartile for 4 or more measures			Group 2: Meeting threshold/in top quartile for <4 measures			
	Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
27	7	102	(6.8%)	41	498	(8.2%)	p=0.62
28	6	124	(4.8%)	33	592	(5.6%)	p=0.75
29	3	162	(1.9%)	19	680	(2.8%)	p=0.49
30	2	207	(1.0%)	20	833	(2.4%)	p=0.20
31	2	294	(0.7%)	22	1102	(2.0%)	p=0.10
27-31	20	889	(2.2%)	135	3705	(3.6%)	p=0.04

Table 43 Pre-discharge mortality for babies by gestational week of birth when categorising units based on adherence with NNAP audit measures

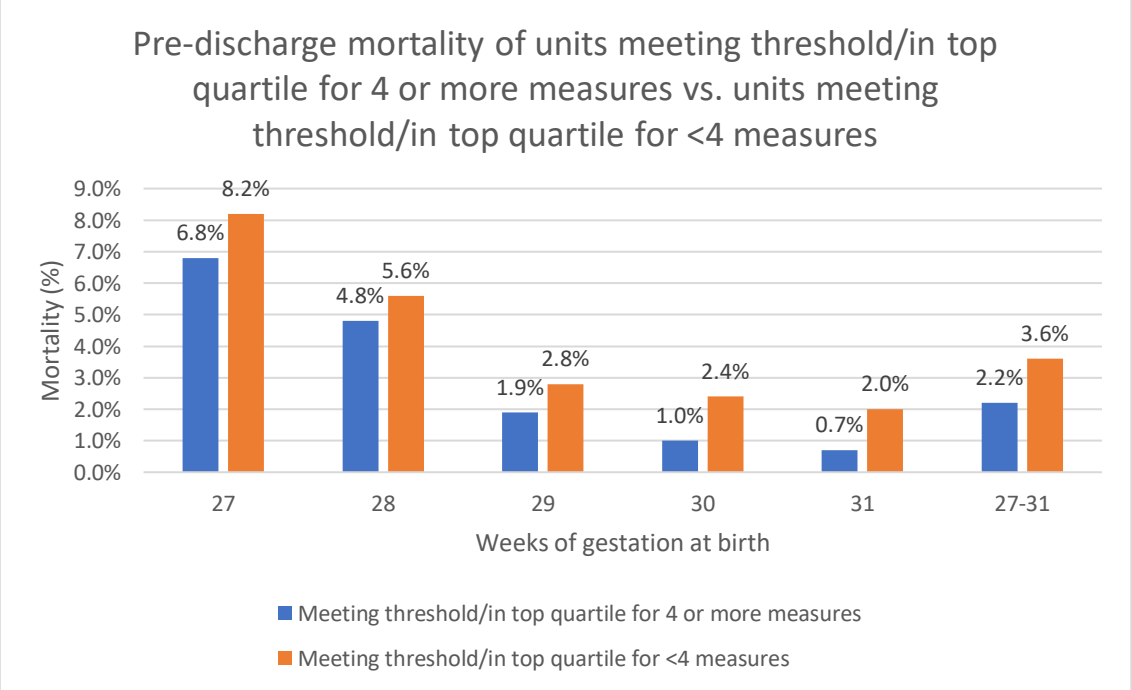


Figure 63 Pre-discharge mortality by gestational week of birth for units meeting threshold/in top quartile for 4 or more NNAP audit measures versus units meeting threshold/in top quartile for <4 measures

8.2.4.2 Pre-discharge mortality by missing data for NNAP audit measures

Comparing pre-discharge mortality of Group 1 versus Group 2 for missing data for NNAP audit measures, no significant difference found was for the whole cohort, or by each gestational week of birth (Table 44, Figure 64).

Gestational week at birth	Group 1: In top quartile/any missing data for 2 or more measures			Group 2: In top quartile/any missing data for <2 measures			
	Mortality (pre-discharge)	Patient total	(%)	Mortality (pre-discharge)	Patient total	(%)	
27	8	97	(8.2%)	39	512	(7.6%)	p=0.83
28	9	146	(6.2%)	33	588	(5.6%)	p=0.71
29	6	166	(3.6%)	17	701	(2.4%)	p=0.40
30	7	196	(3.6%)	15	880	(1.7%)	p=0.10
31	6	286	(2.1%)	18	1150	(1.6%)	p=0.47
27-31	36	891	(4.0%)	122	3831	(3.2%)	p=0.21

Table 44 Pre-discharge mortality for babies by gestational week of birth when categorising units based on missing data for NNAP audit measures

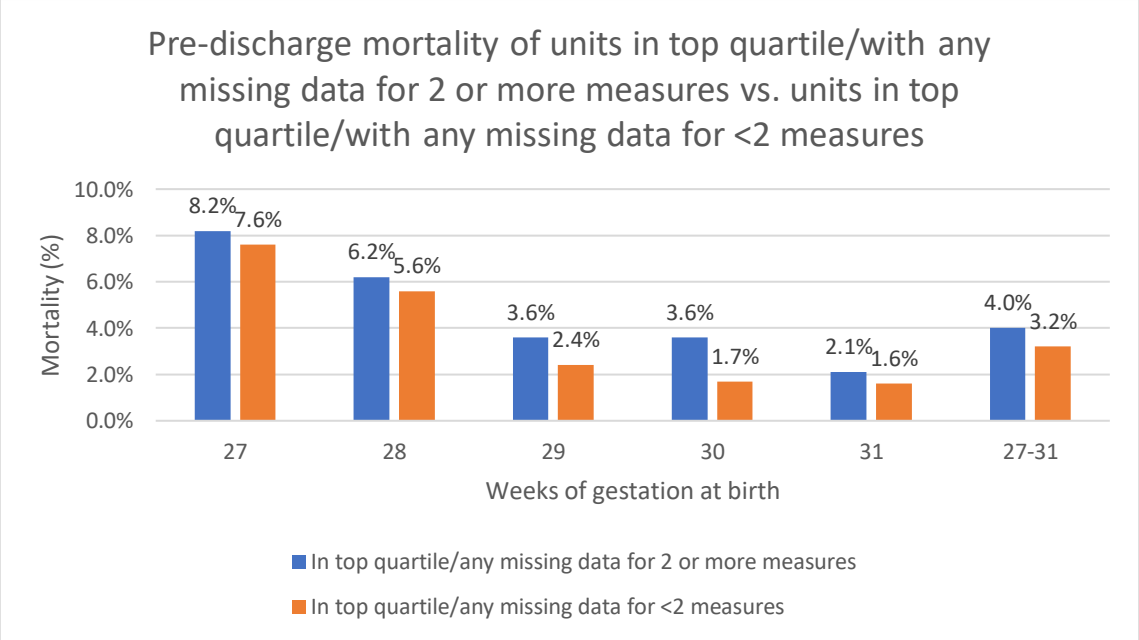


Figure 64 Pre-discharge mortality by gestational week of birth for units in top quartile/with any missing data for 2 or more NNAP audit measures versus units in top quartile/with any missing data for <2 measures

8.2.4.3 Length of stay (LOS) by adherence with NNAP audit measures

Comparing mean LOS (days) of Group 1 versus Group 2 for adherence with NNAP audit measures, it was significantly lower for the whole cohort of babies born between 27-31 weeks (difference in weighted mean LOS 3.7 days, 95% CI 0.6 – 6.8, p=0.02). By each gestational week of birth, it was significantly lower for babies born at 30 weeks (difference in weighted mean LOS 3.3 days, 95% CI 0.0 – 6.6, p=0.04) (Table 45, Figure 65).

Gestational week at birth	Mean length of stay (LOS – days)			
	Group 1: Meeting threshold/in top quartile for 4 or more measures	Group 2: Meeting threshold/in top quartile for <4 measures	Difference (days, 95% CI)	
27	84	84	0.3 (-5.5 – 6.0)	p=0.93
28	70	72	1.6 (-4.3 – 7.4)	p=0.58
29	57	59	1.5 (-3.4 – 6.5)	p=0.53
30	45	48	3.3 (0.0 – 6.6)	p=0.04
31	35	37	2.4 (-0.6 – 5.4)	p=0.11
27-31	52	55	3.7 (0.6 – 6.8)	p=0.02

Table 45 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth when categorising units based on adherence with NNAP audit measures

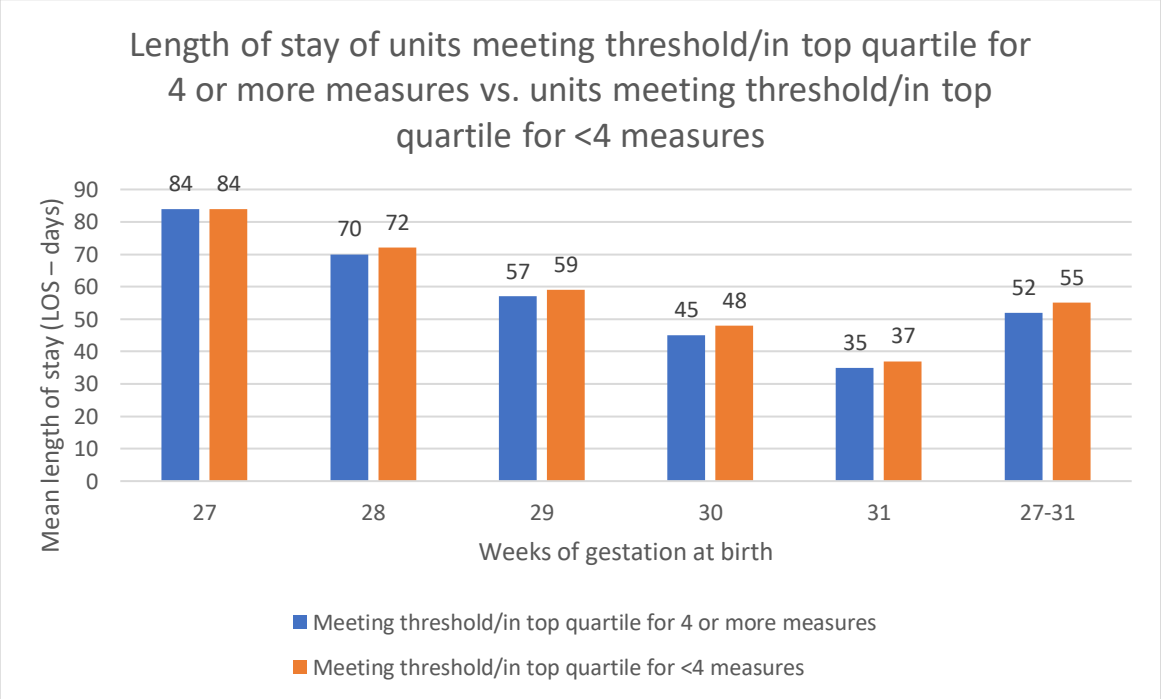


Figure 65 Length of stay by gestational week of birth for units meeting threshold/in top quartile for 4 or more NNAP audit measures versus units meeting threshold/in top quartile for <4 measures

8.2.4.4 LOS by missing data for NNAP audit measures

Comparing LOS of Group 1 versus Group 2 for missing data for NNAP audit measures, no significant difference found was for the whole cohort, or by each gestational week of birth (Table 46, Figure 66).

Gestational week at birth	Mean length of stay (LOS – days)			
	Group 1: In top quartile/any missing data for 2 or more measures	Group 2: In top quartile/any missing data for <2 measures	Difference (days, 95% CI)	
27	80	85	4.8 (-1.4 – 11.1)	p=0.12
28	67	73	5.4 (-0.9 – 11.7)	p=0.09
29	58	59	0.4 (-5.0 – 5.7)	p=0.88
30	46	47	1.0 (-2.7 – 4.6)	p=0.59
31	37	37	0.3 (-3.0 – 3.5)	p=0.87
27-31	52	55	2.4 (-1.1 – 5.9)	p=0.16

Table 46 Length of stay (LOS – weighted mean rounded to nearest day) for babies by gestational week of birth when categorising units based on missing data for NNAP audit measures

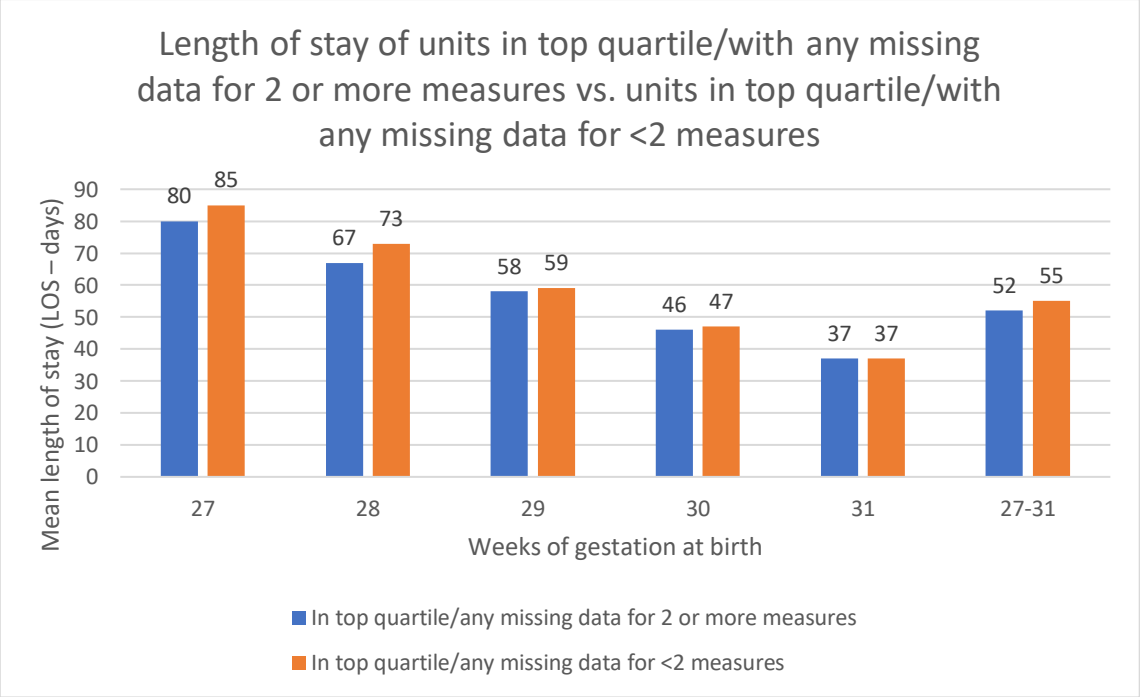


Figure 66 Length of stay by gestational week of birth for units in top quartile/with any missing data for 2 or more NNAP audit measures versus units in top quartile/with any missing data for <2 measures

8.3 Comparing data for NNAP audit measures and my non-NNAP MQC

There was no significant overlap when comparing Group 1 for adherence with my non-NNAP MQC (combined Measures 1-4 – steroids, temperature, ventilation, nursing), with Group 1 for adherence with NNAP audit measures. However, both datasets had a different total number of units - 98 for NNAP, all of which were included in the 113 units for NDAU. Of the 21 units in Group 1 for the NNAP audit measures, 10 were in Group 1 for the combined Measures 1-4 of my non-NNAP MQC, and 11 were not.

8.4 Multivariate analyses

8.4.1 Transformation of data to meet assumptions of multivariate analysis

Before conducting the multivariate analyses, I checked my data to see if it met the assumptions for the tests. For logistic regression, there were the following assumptions:

1. Large sample size
    a. Yes (>10 x number of confounding variables)
2. Response (outcome) variable is binary
    a. Yes
3. Observations are independent
    a. Yes
4. No multicollinearity among explanatory (confounding) variables
    a. Yes, checked by examining Pearson correlations between variables (Appendix VI)
5. No extreme outliers
    a. Yes (since all babies with incorrect or missing data for confounding variables were excluded)
6. Linear relationship between explanatory (confounding / independent) variables and the logit of the response (outcome / dependent) variable
    a. Box Tidwell plot (Appendix VI) showed BW had a non-linear relationship with outcome variable (mortality).

So my data would meet the assumptions for logistic regression, BW was changed from a continuous variable to categorical variable using internationally accepted classification of normal BW (>2500g), LBW (1500-2499g), VLBW (1000-1499g) and ELBW (≤999g).

For linear regression, there were the following assumptions:

1. Linear relationship between outcome and confounding variables
    a. Yes
2. No multicollinearity
    a. Yes, checked by examining Pearson correlations between variables and variance inflation factor (VIF) (Appendix VI)

3. Homoscedasticity
    a. No, checked by examining graph of regression standardized residuals against regression standardized predicted values (Appendix VI)
4. Normality
    a. No, checked by examining histogram. Data skewed to left (Appendix VI)

So my data would meet the assumptions for linear regression, the outcome variable (LOS) was transformed using natural log (Appendix VI). To analyse the results, a reverse log calculation was applied to the unstandardised coefficient and confidence intervals.

8.4.2 Associations between adherence with MQC and outcomes using multivariate analyses

8.4.2.1 Logistic regression to look for association between adherence with combined non-NNAP MQC and mortality

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	973.781[a]	.031	.140

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	group(1)	.202	.233	.750	1	.387	1.223
	GWk	-.409	.090	20.728	1	<.001	.664
	BBW	-.001	.000	5.274	1	.022	.999
	Gender(1)	-.025	.192	.016	1	.898	.976
	FetNum(1)	-.191	.234	.667	1	.414	.826
	IMD_Q			.474	4	.976	
	IMD_Q(1)	.008	.254	.001	1	.975	1.008
	IMD_Q(2)	.017	.282	.004	1	.951	1.017
	IMD_Q(3)	-.193	.315	.375	1	.540	.825
	IMD_Q(4)	-.025	.318	.006	1	.939	.976
	MetOfRes(1)	2.116	.274	59.624	1	<.001	8.295
	PoB			10.119	2	.006	
	PoB(1)	.418	.246	2.902	1	.088	1.520
	PoB(2)	.782	.246	10.114	1	.001	2.185
	Constant	8.880	2.330	14.524	1	<.001	7187.795

a. Variable(s) entered on step 1: group, GWk (gestational week), BBW (baby's birthweight), Gender, FetNum (foetal number), IMD_Q, MetOfRes (method of resuscitation), PoB (place of birth).

Variables in the Equation			
		95% CI for EXP(B)	
		Lower	Upper
Step 1[a]	group(1)	.775	1.931
	GWk	.557	.792
	BBW	.998	1.000
	Gender(1)	.670	1.422
	FetNum(1)	.522	1.307
	IMD_Q		
	IMD_Q(1)	.613	1.659
	IMD_Q(2)	.585	1.768
	IMD_Q(3)	.445	1.528
	IMD_Q(4)	.523	1.820
	MetOfRes(1)	4.848	14.191
	PoB		
	PoB(1)	.939	2.460
	PoB(2)	1.350	3.538
	Constant		

a. Variable(s) entered on step 1: group, GWk (gestational week), BBW (baby's birthweight), Gender, FetNum (foetal number), IMD_Q, MetOfRes (method of resuscitation), PoB (place of birth).

Babies in neonatal units in group 1 did not have a statically significant difference in adjusted odds of pre-discharge mortality compared to group 2 (aOR 1.22, 95% CI 0.78 – 1.93). Variables that were significantly associated with this outcome included gestational week, birthweight, requiring significant resuscitation at birth, and designation of unit of birth.

8.4.2.2 Linear regression to look for association between adherence with combined non-NNAP MQC and LOS

Model Summary						
					Change Statistics	
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change
1	.684[a]	.467	.466	18.838	.467	464.083

The variables inputted into the linear regression model explained 46.7% of the variation in LOS.

Following natural log transformation of LOS to meet the assumptions of the test:

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.734[a]	.538	.537	.30208
a. Predictors: (Constant), Group, GWk (gestational week), Gender, MetOfRes (method of resuscitation), FetNum (foetal number), IMD_Q, PoB (place of birth), BBW (baby's birthweight)				
b. Dependent Variable: ln_LOS				

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	450.239	8	56.280	616.771	.000[b]
	Residual	386.258	4233	.091		
	Total	836.498	4241			

a. Dependent Variable: ln_LOS

b. Predictors: (Constant), Group, GWk (gestational week), Gender, MetOfRes (method of resuscitation), FetNum (foetal number), IMD_Q, PoB (place of birth), BBW (baby's birthweight)

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	8.696	.118		73.844	.000
	GWk	-.145	.004	-.449	-33.223	<.001
	BBW	.000	.000	-.360	-26.505	<.001
	Gender	-.043	.009	-.048	-4.521	<.001
	FetNum	.059	.010	.060	5.666	<.001
	IMDQ	.004	.003	.012	1.173	.241
	MetOfRes	.098	.030	.034	3.224	.001
	PoB	.011	.006	.021	1.907	.057
	Group	.030	.011	.029	2.701	.007

a. Dependent Variable: ln_LOS

Coefficients[a]			
		95.0% Confidence Interval for B	
Model		Lower Bound	Upper Bound
1	(Constant)	8.465	8.927
	GWk	-.154	-.137
	BBW	-.001	.000
	Gender	-.061	-.024
	FetNum	.039	.080
	IMDQ	-.003	.011
	MetOfRes	.039	.158
	PoB	.000	.023
	Group	.008	.052

a. Dependent Variable: ln_LOS

Reversing the natural log on my output, the LOS for babies in neonatal units in group 1 was one day less than for group 2 and this was statistically significant (95% CI 1.008-1.053, p=0.007). Other variables that were also significantly associated with this outcome included gestational week, birthweight, gender, foetal number, and requiring significant resuscitation at birth.

8.4.2.3 Logistic regression to look for association between adherence with NNAP audit measures and mortality

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	900.562[a]	.032	.140

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	group(1)	-.280	.294	.905	1	.341	.756
	GWk	-.429	.094	20.895	1	<.001	.651
	BBW	-.001	.000	3.636	1	.057	.999
	Gender(1)	-.017	.199	.007	1	.932	.983
	FetNum(1)	-.269	.248	1.175	1	.278	.764
	IMD_Q			.599	4	.963	
	IMD_Q(1)	.086	.265	.104	1	.747	1.089
	IMD_Q(2)	.105	.292	.128	1	.720	1.110
	IMD_Q(3)	-.145	.327	.196	1	.658	.865
	IMD_Q(4)	.043	.328	.017	1	.896	1.044
	MetOfRes(1)	2.029	.286	50.395	1	<.001	7.609
	PoB			8.440	2	.015	
	PoB(1)	.467	.261	3.208	1	.073	1.595
	PoB(2)	.751	.258	8.438	1	.004	2.119
	Constant	9.332	2.425	14.806	1	<.001	11297.879

a. Variable(s) entered on step 1: group, GWk (gestational week), BBW (baby's birthweight), Gender, FetNum (foetal number), IMD_Q, MetOfRes (method of resuscitation), PoB (place of birth).

Variables in the Equation			
		95% C.I.for EXP(B)	
		Lower	Upper
Step 1[a]	group(1)	.425	1.346
	GWk	.542	.783
	BBW	.998	1.000
	Gender(1)	.665	1.453
	FetNum(1)	.470	1.243
	IMDQ		
	IMDQ(1)	.648	1.832
	IMDQ(2)	.626	1.968
	IMDQ(3)	.456	1.643
	IMDQ(4)	.549	1.984
	MetOfRes(1)	4.345	13.326
	PoB		
	PoB(1)	.957	2.658
	PoB(2)	1.277	3.516
	Constant		

a. Variable(s) entered on step 1: group, GWk (gestational week), BBW (baby's birthweight), Gender, FetNum (foetal number), IMD_Q, MetOfRes (method of resuscitation), PoB (place of birth).

Babies in neonatal units in group 1 did not have a statically significant difference in adjusted odds of pre-discharge mortality compared to group 2 (aOR 1.22, 95% CI 0.43 –

1.35). Variables that were significantly associated with this outcome included gestational week, requiring significant resuscitation at birth, and designation of unit of birth.

8.4.2.4 Linear regression to look for association between adherence with NNAP audit measures and LOS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics	
					R Square Change	F Change
1	.679[a]	.462	.460	19.170	.462	416.195

The variables inputted into the linear regression model explained 46.2% of the variation in LOS.

Following natural log transformation of LOS to meet the assumptions of the test:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.730[a]	.533	.532	.30636

a. Predictors: (Constant), Group, GWk (gestational week), Gender, MetOfRes (method of resuscitation), FetNum (foetal number), IMD_Q, PoB (place of birth), BBW (baby's birthweight)
b. Dependent Variable: ln_LOS

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	415.888	8	51.986	553.872	.000[b]
	Residual	364.643	3885	.094		
	Total	780.531	3893			

a. Dependent Variable: ln_LOS
b. Predictors: (Constant), Group, GWk (gestational week), Gender, MetOfRes (method of resuscitation), FetNum (foetal number), IMD_Q, PoB (place of birth), BBW (baby's birthweight)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.652	.125		69.201	.000
	GWk	-.145	.005	-.446	-31.478	<.001
	BBW	.000	.000	-.358	-25.196	<.001
	Gender	-.044	.010	-.048	-4.361	<.001
	FetNum	.059	.011	.059	5.292	<.001
	IMDQ	.003	.004	.009	.830	.407
	MetOfRes	.095	.032	.032	2.941	.003
	PoB	.008	.006	.014	1.245	.213
	group	.054	.012	.049	4.305	<.001

a. Dependent Variable: ln_LOS

Coefficients[a]			
Model		95.0% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	8.407	8.898
	GWk	-.154	-.136
	BBW	-.001	.000
	Gender	-.063	-.024
	FetNum	.037	.081
	IMDQ	-.004	.010
	MetOfRes	.032	.158
	PoB	-.004	.020
	group	.029	.078

a. Dependent Variable: ln_LOS

Reversing the natural log on my output, the LOS for babies in neonatal units in group 1 was one day less than for group 2 and this was statistically significant (95% CI 1.029-1.081, p=<0.001). Other variables that were also significantly associated with this outcome included gestational week, birthweight, gender, foetal number, and requiring significant resuscitation at birth.

9 Interpretation of results

In this chapter I present an in-depth analysis of my results regarding:

- Demographic profile and unit characteristics for the whole cohort, and when split by gestational age and my comparator groups.
- The findings of the univariate and multivariate analysis looking for associations between adherence to my non-NNAP MQC / NNAP audit measures and outcomes.

9.1 Modified Optiprem dataset

9.1.1 Describing demographics for entire cohort by gestational week of birth

Consistent with previous research (98), per unit, just over double the number of babies within my cohort were delivered in obstetric units with NICU as opposed to LNU. This was mainly accounted for by babies born at 27 and 28 weeks of gestation and explains why the birthweight of babies born in NICU was, on average, 75g less than for babies born in LNU. Otherwise, as expected, the proportion of babies increased by each gestational week of birth, as did their birthweight. The slight trend I found towards reduced multiplicity with lower gestational age at birth probably reflects the increased risk of complications and pre-admission mortality with more preterm birth in combination with increased multiplicity (275). The split by gender and the IMD_Q (index of multiple deprivation score) was relatively consistent across the gestational weeks, with a slight preponderance for male babies, and a reducing proportion from 1 (most deprived) to 5 (least deprived), consistent with research showing associations between socioeconomic deprivation and preterm birth (276, 277).

When I examined IMD_Q by designation of unit of birth, the most significant difference was a higher proportion of the most deprived quintile (1) in NICU compared to LNU. In general, NICU are found in hospitals with specialised obstetric units dealing with complex pregnancies, whereas LNU are found in hospitals with obstetric units that manage less high-risk pregnancies. This is so that pregnant women can receive their antenatal, and their baby's postnatal care within the same hospital. These perinatal centres incorporating NICU are generally situated in areas of highest population density, i.e.,

inner city, urban areas. They act as referral centres, providing advice and tertiary care for LNU and SCU within their network. A significant proportion of these lower-level units are in less urbanised areas (i.e., less heavily built up, with lower population densities) (278). This was a purposeful decision made by the UK Department of Health (DOH) when reorganising neonatal services (42); for level 2 units (subsequently renamed as 'local' neonatal units – LNU (45)) to remain in communities to prevent the significant travel and time burdens associated with major centralisation. This explains why it is more likely the population of LNU are less deprived than the population of NICU (279). I also found a significantly higher proportion of the top two quintiles (4 and 5) in NICU compared to LNU, but the numbers were much smaller. This is probably a reflection of the fact that a higher proportion of the wealthiest people in society live in cities as opposed to towns and villages (280, 281).

9.1.2 Adherence with non-NNAP MQC across entire cohort by gestational week of birth

The benefits of antenatal steroids for preterm babies are universally accepted and this has been part of standard practice for over 20 years (238). As such, it is unsurprising that adherence with Measure 1 (steroids) was uniformly >90% across the gestational age range. There will always be a small proportion of preterm babies who do not receive this treatment due to mothers presenting in labour and/or presence of contraindications. This was one of two of my measures of quality of care that were similar to audit measures used by the NNAP. The other was recording of normal temperature within one hour of admission (Measure 2). There are many factors relating to structure and process which contribute to this, including the temperature of delivery suite rooms/operating theatres, care taken to keep baby covered during stabilisation/resuscitation and use of plastic bags for babies <30 weeks of gestation, distance from delivery suite to the neonatal unit, mode of transport (e.g., transport incubator vs. resuscitaire), measuring of temperature prior to departure from delivery suite and use of a heated mattress. For my cohort of babies, units were only compliant with this measure for 73-75% of babies born at 30-31 weeks, falling to 67-68% for babies born at 27-29 weeks, perhaps reflecting the reduced ability of more preterm babies to maintain their body temperature and therefore the need for extra care, including the use of a plastic bag/wrap (282).

The nearly linear decrease in the proportion of babies requiring ventilatory support on day one of life supported using NIV (Measure 3) with decreasing gestational age at birth, perhaps reflects uncertainty and reluctance of clinicians to hold off intubating and ventilating a significant proportion of very preterm, and especially extremely preterm babies. This is despite growing evidence of the benefits of early PEEP, which can reduce the need for surfactant in these babies (262). Adherence with this measure is highly dependent on unit culture and this was born out with individual unit adherence ranging from 14-96%. For babies requiring intensive care on day one of life provided with 1:1 nursing care (Measure 4), overall adherence was very low despite this being a BAPM (British Association for Perinatal Medicine) standard (46). In part, this may be due to chronic shortages in nurse staffing numbers, due to acute and long-term sickness, and issues relating to training and retention of specialised staff (283-285). As might be expected, more of the babies at the lower end of the gestational age range requiring intensive care had 1:1 nursing, but even then, only 15.6%.

Missing data for Measures 1-4 was <1%. However, for Measures 5 and 6 (receipt of mother's milk on day one of life and delayed cord clamping), the amount of missing data was sufficiently high that I did not use them to categorise units (there would be no way to verify the accuracy of unit allocation, especially given that the missing data is highly unlikely to be random). However, despite this, it is interesting to note the relationship between adherence and gestational age at birth for this cohort of babies. Overall, delayed cord clamping seemed to have good adherence, reaching 81-83% for babies at 27-28 weeks. This could be because of the perceived increase in benefit of delayed cord clamping for more preterm babies, however, due to the significantly high amount of missing data for this measure (37.9%), the overall adherence and trend could just be due to selective recording of data, i.e., those babies who had delayed cord clamping were more likely to have it recorded. Based on this assumption, and the fact that the proportion of missing data is relatively constant across the gestational age range, the trend is probably true albeit the overall adherence is likely significantly lower.

For receipt of mother's milk on day one of life, overall adherence was very low, even if we assume that the 13.8% missing data was for babies who had received this measure. This is probably because this measure requires a truly multidisciplinary approach (which is not often achieved), including antenatal education on the benefits of breastmilk,

especially if mother's deliver preterm and even if they are not planning to breastfeed, reiteration of this message by the obstetric and neonatal team in the perinatal period with advice and support to start expressing as soon after delivery as possible, and neonatal protocols to use whatever colostrum is available as early as possible, if even just for mouth care (buccal/oropharyngeal colostrum). If we assume it to be accurate, the slight upward trend from 9.1% at 27 weeks to 15.0% at 31 weeks seems paradoxical (again, the proportion of missing data was relatively constant across the gestational age range). We might have expected there to be a greater drive to provide more preterm babies with their mother's milk on day one of life given the proven benefit in establishing a normal microbiome, protecting against necrotising enterocolitis (NEC), and helping to more quickly establish full enteral feeds and reduce need for parenteral nutrition (257-260). However, this could reflect a potentially greater degree of illness, and therefore inability to express, in mothers of more preterm babies (286). There is also a reluctance still found in many units, to feed their more preterm babies early, despite research indicating that this is the very cohort that benefits most (123, 287). In coming years, with better data completion, both of these measures will be potentially valuable markers of quality of care. It would be particularly interesting to investigate an association between early feeding and reduced length of stay (LOS).

9.1.3 Comparing patient populations when categorising units based on adherence with non-NNAP MQC

When using these measures of quality of care to categorise units and comparing their demographic profiles, several statistically significant differences were found. The proportion of LNU:NICU was significantly higher than expected for units within Group 1 for Measure 4 (nursing). We know that in general, LNU are smaller volume than NICU (98), and this is also true for my data for babies born between 27-31 weeks, as shown in Section 8.1.1. We also know that many neonatal units lack adequate nurse staffing (283-285). Therefore, we might expect that LNU would be more able to provide 1:1 nurse staffing for babies that require intensive care on day one of life, and therefore, disproportionately represented in the top quartile for this measure. A similar relationship with unit designation was found for the measures relating to normal temperature on admission and support with NIV for babies requiring respiratory support on day 1 of life, but it did not reach statistical significance for either. Therefore, it was unsurprising that

when categorising units by the combination of Measures, there was a significantly raised proportion of LNU:NICU in the units within the top quartile for two or more measures.

The proportion of babies with significant congenital anomalies was significantly higher in Group 1 than Group 2 for the combination of Measures (steroids, temperature, ventilation, nursing). These babies form a high-risk group, with increased care needs that can involve specialist input and transfer for surgery, and have increased risk of severe morbidity outcomes and mortality. If a large difference was seen in the proportion of these babies between units in Group 1 compared to Group 2, this would have been of clinical significance, especially when interpreting any differences in outcomes. However, as expected the total number of these babies was small (2.7%), and the difference between groups of only 1.1%. Therefore, this cannot be considered of clinical significance for my results. This is supported by the lack of finding a significant difference between the distribution of surgical NICU (who treat a majority of babies with significant congenital anomalies requiring surgical intervention) between groupings of neonatal units when categorising according to individual non-NNAP MQC or the combination of Measures.

The proportion of babies in each quintile for the IMD_Q, within units in the top quartile (Group 1) versus those not in the top quartile (Group 2) was significantly different from expected for Measures 1-4 and when combining measures. For Measures 1 (steroids) and 4 (nursing) the differences were clinically insignificant, but for Measures 2 (temperature) and 3 (ventilation), and the combination of Measures, there was a significant trend towards a less deprived population of patients in units that were in the top quartile, compared to units that were not in the top quartile. This relationship is unlikely to be direct and unifactorial, and more likely to be related to a systemic difference in the structure and processes of care employed by units and the wider healthcare system in more affluent, compared to more deprived areas (e.g., level of staffing and proficiency of healthcare workers, work environment and organisational culture, access to antenatal healthcare involving health promotion, etc.). Certainly, this relationship has previously been described (288), and in relation to other aspects of healthcare, e.g., smoking prevalence, availability of GP appointments, emergency admissions to hospitals, A&E waiting times, avoidable deaths (289, 290).

And finally, the proportion of babies requiring resuscitation involving cardiac massage or adrenaline was significantly lower in Group 1 compared to Group 2 for Measures 2 (temperature) and 3 (ventilation). Although it is just as important, if not arguably more important for babies requiring a significant degree of resuscitation at birth to be kept at a normal temperature, in practice this is more difficult. This is due to exposure of the baby as multiple healthcare staff are involved in simultaneously performing several different procedures, including ventilation, cardiac massage, monitoring of saturations and heart rate via pulse oximeter, and placement of an umbilical venous catheter and administration of fluid or drugs. While this could explain why units with a higher proportion of babies requiring significant resuscitation were not in the top quartile for this measure, the absolute number of babies requiring such resuscitation (3.1%), and the difference between the groups is small (1.5%). Similarly, it seems reasonable that babies who require significant resuscitation are less likely to be put on NIV and more likely to be intubated and invasively ventilated, and so units with less of these babies would be more likely to be within the top quartile for this measure. However, again, the absolute numbers of these babies are very small (3.1%), and furthermore, they were excluded from this measure for precisely this reason, since it would be inappropriate to support them using NIV after they have required such extensive resuscitation. Another possible reason for this difference, which was beyond the remit of this study to investigate, relates to the quality of obstetric care, especially in the perinatal period, and thereby the condition of the baby in the immediate postnatal period.

Therefore, when combining Measures 1-4, the only statistically and clinically significant differences detected between units and their populations that were in the top quartile for two or more measures (Group 1), compared to those in the top quartile for <2 measures (Group 2), was for unit designation and the IMD_Q. I.e., units in the top quartile for two or more measures were more likely to be LNU and have a more affluent and less deprived population. From the data, it also seems as if they were more likely to have a lower volume of larger birthweight babies (with p value of 0.08, and so nearing but not reaching statistical significance).

As previously discussed, my data shows that for this population of babies born between 27-31 weeks, on average per unit, more than double the number were born in NICU compared to LNU, and as a group, those born in LNU had a significantly higher

birthweight than their counterparts born in NICU. Furthermore, significantly more of the most deprived babies were born in NICU compared to LNU.

The Getting It Right First Time (GIRFT) programme published '*A snapshot of neonatal services and workforce in the UK*' (291), providing details of staffing levels in UK neonatal units for two days in September 2019. They found that only 60% of NICU met BAPM standards for nursing compared to 86% of LNU. Overall, 15% of units had nurse staffing levels below rostered, but NICU had three times as many gaps in Band 5+ rotas (40%), compared with LNU (13%). In comparison, only 5% of units had medical staffing levels below rostered, and NICU were more likely to meet BAPM standards (49-94% across weekday/weekend days/nights), compared to LNU (40-84%).

Putting this all together, the units that provided better quality of care (as per my evidence-based measures) were more likely to be LNU, and my data indicates that this was because they were smaller (lower volume), better staffed (higher nurse to patient ratio), and with a less deprived population. I cannot say this definitively, because I only had data for babies born between 27-31 weeks (not the whole gestational age range of neonatal admissions to these units), and no information on nurse staffing levels beyond what I used for Measure 4 (nursing).

9.1.4 Associations with outcomes

9.1.4.1 By gestational age at birth

The pre-discharge mortality rates for babies born between 27-31 weeks of gestation, discharged in 2018, was in keeping with international figures (Section 3.3.1, Table 5). As expected, mortality decreased with increasing gestational age at birth, and was approximately five-fold higher for babies born at 27 weeks than those born at 31 weeks. Similarly, LOS for babies born increased with decreasing gestational age at birth, being more than double for babies born at 27 weeks compared to 31 weeks and is in keeping with previous research (292).

However, in absolute terms, incidence of mortality for babies born between 27-31 weeks of gestation was low at 3.3%. Even for those born at 27 weeks (i.e., those with the highest mortality), this was only 7.7%. Therefore, this was not a sensitive outcome measure for this cohort of babies, especially when creating subgroups by gestational week of birth.

The results of any statistical analysis are less likely to be reliable due to the increased variability inherent in small sample sizes.

9.1.4.2 By gestational age and designation of unit of birth

Comparing pre-discharge mortality by designation of unit of birth (NICU vs. LNU), it was significantly higher for babies born between 27-31 weeks, for the cohort as a whole and for those born at 31 weeks of gestation. The trend was identical for babies born at 27, 28, 29 and 30 weeks but did not reach statistical significance. LOS was also consistently longer for babies born in NICU vs. LNU. The difference was significant for the whole cohort, and for babies born at 27, 30 and 31 weeks.

However, it would be wrong to conclude from this that for babies born between 27-31 weeks, being born in a NICU (i.e., receiving perinatal care) causes increased LOS and mortality compared to being born in an LNU. We know that high risk pregnancies are more likely to deliver in NICU compared to LNU, and that demographics such as the IMD_Q differ significantly between the two types of units. This difference in mortality by designation of unit of birth was also found by the statisticians of WS1 for the whole cohort of OptiPrem babies discharged between 2014-2018. However, when the babies in both groups were matched, and when an instrumental variables approach was used, this difference was no longer apparent. Therefore, it could be that this difference is seen due to NICU, in general, having an inherently more high-risk patient population compared to LNU, especially given that obstetric and fetal medicine units linked to NICU are often referral centres for higher risk pregnancies from smaller obstetric units that are linked to LNU.

An important group of babies that are looked after nearly exclusively in NICU are babies requiring significant surgical intervention, with around half of NICU being classified as ‘surgical’ NICU that would provide care for these babies pre- and post-surgery with input from the paediatric surgical team. Therefore, I compared outcomes for surgical NICU versus non-surgical NICU, and for non-surgical NICU versus LNU to get an indication of whether the differences I was finding when comparing NICU versus LNU were largely due to this high-risk subset of babies requiring surgical intervention.

I found a significantly higher pre-discharge mortality in surgical-NICU for babies born at 29 and 31 weeks, and the whole cohort compared to non-surgical NICU. This is in keeping with MBRRACE data (293), and expected since surgical NICU deal with a higher risk population than non-surgical NICU. However, when comparing non-surgical NICU with LNU, I also found a significantly higher pre-discharge mortality for the whole cohort. This could indicate that the increased mortality for babies born between 27-31 weeks in NICU compared to LNU is not wholly due to the 'surgical patient' cohort that is looked after only in NICU. Regarding LOS, I only found a statistically significant increased LOS between surgical and non-surgical NICU for babies born at 31 weeks, whereas when comparing non-surgical NICU to LNU the whole cohort had a significantly longer LOS. This indicates that the 'surgical patient' cohort contributes less towards the difference in LOS seen between NICU and LNU for babies born between 27-31 weeks of gestation, than it does for pre-discharge mortality.

While it is likely true that in general, all NICU (even non-surgical) have a higher risk population than LNU, it is still possible that systemic differences between NICU and LNU also contribute to babies born between 27-31 weeks of gestation in NICU having a higher mortality and longer LOS than those born in LNU. Foremost amongst these is that, in general, NICU are higher volume than LNU, and more poorly staffed with regards to nurses, both factors that as discussed in Section 9.1.3, can negatively impact outcomes such as mortality (73, 98, 168, 169).

9.1.4.3 By adherence with non-NNAP MQC

Comparing pre-discharge mortality by adherence with my measures of quality of care, there was only one result that reached statistical significance. This was for Measure 4 (nursing), for babies born at 27 weeks of gestation, which would seem to indicate that pre-discharge mortality for babies born at 27 weeks of gestation is twice as high in units that are more likely to provide 1:1 nursing care to babies born between 27-31 weeks of gestation requiring intensive care on day one of life, compared to units that are less likely to do this (13.5% vs. 6.6%, p=0.02) (Table 34). If we were to assume this reflects a true association, perhaps the simplest explanation for this would be that extremely preterm babies born at 27 weeks of gestation are more likely to receive 1:1 nursing on the day they are born and are also less likely to survive to discharge (i.e., the sickest babies are identified and targeted for 1:1 care). However, given that I did not find a significant

difference in pre-discharge mortality for babies born at any other gestational week (for babies born at 28 weeks the relationship seems to be reversed – Figure 52) or the whole cohort of babies, it is most likely an erroneous result due to the small numbers involved, and not of clinical significance.

Comparing LOS by adherence with my measures of quality of care, several significant results were found. Units in Group 1 for Measure 3 (ventilation) had significantly lower LOS for the whole cohort of babies born between 27-31 weeks, and those born at 27 weeks of gestation. A possible explanation for the difference in LOS could be that babies who are sicker at birth and more likely to require invasive ventilation are also more likely to have prematurity related complications and a longer LOS than babies that are less sick at birth and can be adequately supported using NIV, who are less likely to have prematurity related complications and a shorter LOS. The proportion of babies requiring significant resuscitation (chest compressions and/or adrenaline) in Group 2 (3.0%) was significantly higher than in Group 1 (1.8%). However, the absolute difference is minor (1.2%), and furthermore, these babies were excluded from analyses for Measure 3 for precisely this reason. Therefore, alternatively, as per my hypothesis, this result could indicate that there is an association between units that are more likely to appropriately provide babies born between 27-31 weeks of gestation with NIV instead of invasive ventilation, and a reduction in LOS, especially for babies born at 27 weeks of gestation. This would be in keeping with current evidence, which shows a reduction in incidence of CLD through use of prophylactic CPAP (262), and the significant impact on LOS that developing CLD has (294).

Regarding the combined Measures (steroids, temperature, ventilation, nursing), Group 1 had significantly reduced LOS for the whole cohort of babies, and also babies born at 27 and 31 weeks of gestation. Given that Group 1 contained a significantly higher proportion of LNU than Group 2, and LOS was significantly lower for babies born in LNU compared to NICU, this could provide an explanation. I have discussed the results of my multivariate analyses regarding adherence with non-NNAP MQC and outcomes in Section 9.3.

The overall low incidence of pre-discharge mortality rate in this cohort of babies is a possible explanation for not finding a significant association between adherence with my

non-NNAP MQC and this outcome measure. If we assume that units with a higher degree of adherence with these evidence-based processes of care have better outcomes, this could provide an explanation as to why an association was not found with pre-discharge mortality but was found with the much more sensitive outcome measure of LOS, which applies to every admitted baby. It is possible that if I was using a more preterm group of babies in which this outcome is significantly higher, it would have increased the likelihood of finding an association, if one was truly there.

9.2 NNAP audit measures

9.2.1 Comparing patient populations when categorising units based on adherence with, and missing data for NNAP audit measures

Unlike with my non-NNAP MQC, where the patient population was composed solely of babies born between 27-31 weeks of gestation, the patient population used by the NNAP differs by audit measure (e.g., antenatal magnesium sulphate applies to babies born before 30 weeks of gestation, bronchopulmonary dysplasia applies to babies born before 32 weeks, minimising inappropriate separation of mother and term baby applies to babies born at term). Because NDAU only provided OptiPrem with data for babies born between 27-31 weeks of gestation (as per the remit of the study), I did not have the data required to be able to make a meaningful comparison between groups of their demographic profiles (in the way I did for my non-NNAP MQC in Section 8.1.3). In other words, I cannot comment on whether there are any characteristics of the population of babies in Group 1 or 2 that make it more or less likely for units to be compliant with the NNAP audit measures. The only exception to this is in the context of looking for associations with outcomes because the outcome data relates specifically to babies born between 27-31 weeks of gestation.

9.2.2 Associations with outcomes

Regarding difference in pre-discharge mortality between Groups 1 and 2 for adherence with NNAP audit measures, a homogenous effect was seen, with a reduction ranging from 0.8-1.4% for each gestational week of birth. While this was not significant for individual gestational weeks, for the cohort as a whole (babies born between 27-31 weeks), the difference in mortality (a reduction of 1.4% – a relative reduction of 39%) achieved the level of significance with a p value of 0.04. Similarly, the LOS for Group 1 was significantly lower than Group 2 for the cohort as a whole (3.7-day difference in weighted mean LOS, 95% CI 0.6 – 6.8, p=0.02), and for babies born at 30 weeks of gestation (3.3-day difference in weighted mean LOS, 95% CI 0.0 – 6.6, p=0.04).

When comparing the demographic and unit profiles for Group 1 and 2, there was no statistically significant difference in the proportion of LNU/NICU, nor in the birthweight, gestational age, gender and multiplicity of the babies. There was also no statistically significant difference in the number of babies requiring resuscitation involving cardiac

massage or adrenaline, and although a statistical test was not conducted on the 5-minute Apgar scores due to >10% missing data, the weighted average of the medians for Group 1 and 2 were identical at 9. The only statistically significant difference found was for the IMD_Q, showing a trend towards a less deprived population in Group 1. Therefore, it is less likely that the differences in mortality or LOS for babies born between 27-31 weeks of gestation between Group 1 and 2 are due to differences in the populations. I have discussed the results of my multivariate analyses regarding adherence with NNAP audit measures and outcomes in Section 9.3.

In contrast to this, I failed to find a significant association between worse data completion for NNAP audit measures and pre-discharge mortality or LOS, either for the cohort as a whole or by individual gestational week. One interpretation of this finding could be that data completion is not a marker of care quality. However, as discussed in Sections 4.4 and 9.8, associations with outcomes are not required for validation of quality-of-care measures. Using a more preterm cohort of babies, in which there is increased sensitivity for all of the major neonatal outcomes, may help to reveal an association, if there is one there. It is also of note that I did not find a negative association between these outcomes and data completion, which may have indicated that units with sicker babies have less time to ensure adequate data entry due to being busier providing emergency clinical care.

9.3 Associations between adherence with NNAP audit measures and non-NNAP MQC and outcomes using multivariate analyses

I conducted multivariate analyses to look for associations between adherence with NNAP audit measures and my non-NNAP MQC and pre-discharge mortality and LOS, for the cohort as a whole. In the univariate analyses I had found a reduction in mortality and LOS for units in group 1 for adherence with NNAP audit measures, and a reduction in LOS for units in group 1 for my non-NNAP MQC. In the multivariate analyses, adjusting for the confounding variables with sufficient data completion, a smaller but still significant reduction in LOS was found for units in group 1 for adherence with NNAP audit measures and my non-NNAP MQC. The association between adherence with NNAP audit measures and pre-discharge mortality was no longer statistically significant in the multivariate analyses.

On face value, these results suggest that units that have better adherence with the NNAP audit measures, and/or practice more evidence-based medicine (at least according to the measures I chose), have a reduced length of stay by an average of one day, for babies born between 27-31 weeks of gestation. However, these results must be interpreted with caution. The multivariate analyses allowed for adjustment of important confounding variables, but as discussed in Section 8.4.2, less than half of the variance in outcomes was explained by the models. This was expected; we know several important, recognised confounders have been excluded due to lack of data (e.g., condition of baby at birth, mother's health status pre- and during pregnancy, ethnicity, etc.). And of course, there will be unknown confounding factors which can only be accounted for using complicated statistical methods such as instrumental variables. Therefore, it is still possible that this result does not reflect a true association.

Having said that, this is not an unexpected result and fits with my hypotheses as outlined in Section 6.3. Units that are striving to comply with national guidance in the form of NNAP audit measures and practice more evidence-based care would be expected to have better outcomes for their babies, and this could result in the small but significant difference in length of stay that I have demonstrated. This could also be because of differences in structure or provision of other processes of care that I have not measured, which might have an indirect (e.g., early implementation of breastmilk feeds, more

opportunities for parents to provide skin-to-skin care) or direct impact on length of stay (e.g., discharging on nasogastric tube feeding, and/or availability of community neonatal nurse follow-up).

In my univariate analysis, I found that units in group 1 (i.e., those more likely to delivery good quality of care, as defined by my evidence-based MQC), were more likely to be LNU. This could have accounted for the reduction in LOS I found, since babies in LNU had a lower average LOS than babies in NICU. However, it is interesting that in my multivariate analysis, where I adjusted for designation of unit, I still found an association between units in group 1 and a reduction in LOS.

What are the implications of these results, i.e., a potential reduction in length of stay of, on average, one day, for all babies born between 27-31 weeks of gestation? This day is going to be a day of special care (with carer present), which neonatal units are reimbursed £535 by NHS England to provide (295). 5638 babies were born between 27-31 weeks of gestation, discharged in 2018. Therefore, this gives a maximal potential cost saving to the NHS in England of just over £3 million per year.

To work out a more conservative estimate, we could use the average patient number for 40 units (based on 41 units being in the top quartile for zero of my non-NNAP MQC, correlating with 40 units meeting threshold/being in the top quartile for <3 NNAP audit measures). The 119 neonatal units (NICU and LNU) had an average of 47 patients each, born between 27-31 weeks of gestation, per year. This gives a total of 1895 patients (nearly exactly 1/3rd of the total 5638). Using this more conservative estimate still gives a potential annual saving of just over £1 million, which is still substantial.

Furthermore, being able to discharge these babies one day earlier will have an effect on cot capacity, and therefore, movement within and between neonatal units. It is difficult to quantify the beneficial effect this would have. It is very likely that there would also be a reduction in LOS for babies of other gestational age ranges which may even be greater than for babies born between 27-31 weeks. Therefore, in theory, this is an underestimate of the true potential effect of units striving to comply with NNAP audit measures and implement evidence-based care.

9.4 Comparing groupings of units across different categorisations

One of the aims of analysing adherence and missing data for my NNAP audit measures, was to perform comparisons investigating whether the same units were found in Group 1 for adherence and Group 2 for missing data (i.e., do the same units comply with national guidance and evidence-based practice, as have good levels of data completion).

Comparing adherence with missing data for the NNAP audit measures did reveal a significant overlap of units. Ignoring the three excluded units, all apart from two units found in Group 1 for adherence were found in Group 2 for missing data, and found in Group 1 for missing data were found in Group 2 for adherence. I.e., in general, units that complied with NNAP audit measures also had good data completion, and units that had poor data completion did not comply with the audit measures. This supports my hypothesis that both adherence with the NNAP audit measures and data completion for the same are surrogate markers for the organisational culture of neonatal units and tells me something about their efforts to provide good quality of care.

Interestingly, a relationship was not found when comparing Group 1 for adherence with NNAP audit measures with Group 1 for adherence with my non-NNAP MQC (overlap of 24%-48%). This was despite two of my non-NNAP MQC (Measure 1 - steroids and 2 - temperature) being similar to NNAP audit measures (although applying to different patient cohorts). In this may lie the explanation, that the NNAP data is fundamentally different from the data I used for my non-NNAP MQC since it applies to a much larger patient cohort, which hampers any comparison between the two.

The NNAP audit measures are well established national guidance for practice for neonatal units. Publicly available annual reports allow units and networks to compare their adherence and data completion levels with each other. Furthermore, units that are outliers face the potential of investigation by the CQC. Therefore, there is strong incentive to comply with the NNAP audit measures. So, it is possible units that comply with NNAP audit measures may not comply with non-NNAP measures of quality of care, even ones relating to evidence based processes of care. Alternatively, those units that comply with NNAP audit measures may also practice more evidence-based medicine. However, for this to be a robust comparison, the entire population that each non-NNAP MQC applied

to would need to be included, rather than just the OptiPrem patient cohort (babies born between 27-31 weeks). In spite of this, if a significant overlap of units had been found between NNAP and non-NNAP measures, a conclusion could be drawn from this, but the converse is not true.

10 Discussion

In this chapter I put this PhD into context by looking at several other studies investigating quality of care in neonatal medicine and highlight what is novel about my work, as well as its strengths and weaknesses. In the conclusion I summarise my most pertinent findings and end with a discussion of planned and possible future work.

10.1 Comparing my work to other studies measuring quality of neonatal healthcare

Given the difficulty in defining and measuring quality of care, and of assessing its impact on outcomes, it is interesting to compare my approach with that of other researchers who have focussed on neonatal patients.

10.1.1 A national longitudinal study

Lee et al. (296) measured the impact of engagement with the national Evidence-based Practice for Improving Quality (EPIQ) program launched in the Canadian Neonatal Network in 2003, on survival (to discharge) without major morbidity. Their cohort included babies born between 23-32 weeks of gestation (grouped into those born between 23-25, 26-28, 29-30, and 31-32 weeks), from 2004-2017 (split into three epochs: 2004-2008, 2009-2012, 2013-2017). Major morbidities included late-onset sepsis, NEC, BPD, severe ROP (stage III or worse or requiring treatment), and severe neurological injury (grade III/IV IVH or periventricular echogenicity). Like the UK, Canada has a national electronic patient data entry system for neonatal units, allowing them to analyse data for >80% of eligible infants (n=50,831), of which 96% had complete data for outcomes and demographics. During the time period of the study many quality improvement care bundles were introduced targeting nosocomial infection (e.g., strategic placement of cleanser dispensers, restricting number of skin breaks per patient, early cessation of antibiotics in culture negative patients), BPD (e.g. prophylactic surfactant for babies born <28 weeks of gestation, restricting hand ventilation, targeting oxygen saturations of 88-92%), neurologic injury (e.g., delayed cord clamping, antenatal magnesium sulphate, minimising use of inotropes), ROP, and NEC (e.g., feeding guidelines, early feeding, use of donor milk, holding enteral feeds during red blood cell transfusions). They found that survival without major morbidity improved for all gestational age groups, over each

epoch. This correlated with an increase in the number of infants normothermic on admission, and who received antenatal steroids. From this, they concluded that engagement with the EPIQ program was associated with a 25% increase in survival without major morbidity in babies born <32 weeks of gestation.

However, unlike my work, this study was not comparing outcomes of units that engaged to differing degrees with the quality improvement initiatives (i.e., there was no 'control group'). Instead, as a retrospective longitudinal study, it measured changes in outcomes over time, finding, as expected, a significant improvement in neonatal outcomes over 13 years. Therefore, even the finding of a correlation is difficult to interpret. Having said that, during this period, Canada's outcomes for preterm babies showed greater improvement compared to other developed nations (297), which they attribute to the nearly national engagement with quality improvement. One method of obtaining a control group could have been to categorise units based on engagement with the quality improvement care bundles (which they did not measure), to see if those units which engaged more, had better outcomes (as per my analysis).

Lee et al. chose to group their babies, rather than look at each gestational week of birth. This had the advantage of larger patient groups, giving higher statistical power. Indeed, one of the issues with the cohort of babies used for WS2 was that when splitting by gestational week of birth, the groups became relatively small. This was further compounded by the scarcity of mortality outcome in babies born between 27-31 weeks of gestation.

Another reason for the large population was the extended duration of the study. I opted against this, due to changes in audit measures and other aspects of neonatal unit organisation and structure that change from year to year and can impact the delivered quality of care. Instead, I focussed on 2018, the final year of the OptiPrem study. However, compared to the entire OptiPrem cohort (discharged between 2014-2018), this resulted in a significantly smaller patient group.

Therefore, future consideration may be given to using NNAP audit measures that remain constant over several years, as well as non-NNAP MQC which remain evidence based. Units could be categorised based on consistent adherence/good data completion, or

improvement over time, to analyse for associations with outcomes. Alternatively, or in conjunction with this, alongside splitting the babies by each gestational week of birth, slightly larger groupings could be used (e.g., 27-29 weeks, 30-31 weeks).

10.1.2 A questionnaire study interrogating unit culture

Kaempf et al. (298) wanted to investigate persistent variation in morbidity outcomes among NICU for VLBW infants. From 2000-2014, 39 NICU belonging to the Vermont Oxford Network (VON) were scored according to the 'Benefit Metric' – a predesigned risk adjusted, composite mortality and morbidity score (incorporating CLD, grade III/IV IVH/PVL, stage III/IV ROP, late onset sepsis, NEC/focal intestinal perforation, $<10^{th}$ centile discharge weight), for their combined 58,272 VLBW infants. The weighted mean difference of the annual score was compared against the group mean to categorise the NICU into three groups: green group A (score significantly above the mean), yellow group B (score insignificantly different from mean), and red group C (score significantly below the mean). To investigate these groups, a multidisciplinary team consisting of a neonatologist, nurse, advanced nurse practitioner and respiratory therapist from each NICU completed a 103-question survey (designed using a modified Delphi method), targeting quality improvement methodology, medical therapies, staffing, unit structure, and organisational culture.

The investigators found that the Benefit Metric for the entire group (of which 14 NICU were in group A, 16 in group B, and 9 in group C), increased by 40% over the 14-year study period, with each group showing an improvement. Analysing for differences in questionnaire components between group A and groups B and C, several descriptors relating to better team working, higher morale, more learning opportunities, and better staffing were identified. However, they also unexpectedly found an inverse correlation with other 'positive' descriptors (having a formal palliative care team, paediatric trainees regularly performing ward rounds on premature infants, increased nursing continuity of care for VLBW infants, increased outdoor facing windows, and regular, formal, staff-celebrating, positive-feedback events).

If we would not conclude from this that, e.g., celebrating staff members or having better continuity of care for babies and their families, is associated with worse outcomes, does this invalidate all results from this study? In Section 9.8 I discuss why measures of quality

of care are not validated using outcomes but can be used to look for associations with outcomes to try and help explain the variation we see between units of the same designation/type, as this study set out to do. Indeed, this provides an alternate way to answer my research question, i.e., instead of categorising units by adherence/missing data for measures of quality of care and looking for associations with outcomes, to instead categorise units based on outcomes and look for associations with adherence/missing data for measures of quality of care. But just as with my analysis we cannot necessarily expect to find associations with outcomes, using this method can result in finding associations which are counter intuitive.

Similar to this study, I had initially planned, designed and piloted a questionnaire to gather data on the structure and process of care employed by units, including engagement with quality improvement (Section 5). The main reason for not going ahead with this was the poor response rate, despite truncating its length to only include 28 questions, 27 of which were multiple choice, and the need for it to only be completed by one respondent. In contrast to that, in this study they had a 100% response rate, despite a 103-question survey requiring 4-5 respondents of the multidisciplinary healthcare team. This is likely due to the research team and all the NICU belonging to the VON, and actively involved in continuous quality improvement initiatives asked about within the questionnaire. A similar UK national NHS initiative, funded by the Department of Health, was 'Getting It Right First Time' (GIRFT). It recently published the results of its questionnaire exploring variation in care provided within different specialities, including neonatology (291). As part of an NHS Improvement programme, it also had a 100% response rate. Therefore, in future work, it would be interesting to group units according to the GIRFT findings and investigate for associations with outcomes.

10.1.3 An international study linking individual patient care to outcomes

Zeitlin et al. (299) also wanted to investigate the variation in outcomes for very preterm babies between countries and even units. They did this looking at whether babies born between 24-31 weeks of gestation in 19 regions of 11 European countries, born between 2011-2012 (n=7336), received evidence-based care. This included four components, chosen based on a high level of evidence for effects on pre-discharge mortality and severe morbidity outcomes, and the ability to reliably ascertain application from the medical records. The four components were: 1) delivery in a maternity unit with appropriate

neonatal services, 2) any administration of antenatal steroids before delivery, 3) effective prevention of hypothermia (<36^{O}C), 4) early nCPAP use, or surfactant administration within two hours of birth for babies born <28 weeks of gestation. Severe morbidity included IVH grade III/IV, cystic PVL, ROP stage III/IV, severe NEC (assessed by surgery or peritoneal drainage, since Bells stages were not used in all regions). BPD was not included due to large regional variation in definitions and management. Covariates for use in their multivariate model included gestational age, sex, multiple pregnancy, pregnancy complications, SGA, type of delivery, and Apgar score at 5 minutes.

The investigators found that 58.3% of infants received all four components, increasing to 91.4% receiving at least two of them. The likelihood for receiving this care varied by patient type (reduced for infants born <26 weeks of gestation, singletons, SGA, Apgar score <7 at 5 minutes of age, and born on the day of maternal admission), and by region (32.0% to 75.5%). In their propensity score weighted models, mortality was 28% lower, and combined mortality or severe morbidity 18% lower, for those infants who received all four components of care. Through their modelling they predicted a reduction in mortality of 17.9% if all infants had received all four components (reducing to 11.8% if 90% of infants met this threshold). Therefore, they concluded that striving to deliver evidence-based care would have a significant effect in reducing mortality and severe morbidity outcomes for very preterm infants.

Similar to this study, the non-NNAP MQC I chose were evidence based. However, while their analysis was at a patient level, I was looking at units since OptiPrem was interested in investigating the best place of birth and care for my cohort of babies. Therefore, my analysis was of a more 'indirect' nature. The patient cohort used by this study was also much larger (in terms of gestational age range, time period, and geographical location). These factors may explain why, in contrast to my analyses, they found a relatively large and statistically significant difference between babies receiving all four of their components of care, compared to those that did not, Another reason may have been the use of an 'all-or-none' approach. The equivalent of this in my work was grouping units according to the number of non-NNAP MQC or NNAP audit measures for which they were in the top quartile/meeting the threshold. However, as you get further away from the direct recipient of the specified care process, the harder it becomes to use an all-or-none approach since it reduces the size of your positive comparator group significantly. Indeed,

for my non-NNAP MQC only one unit was in the top quartile for all four measures, and a further six units in the top quartile for three measures. Similarly, of the nine NNAP audit measures I was using, no units met the thresholds for all nine, or eight of the measures, with two units meeting the threshold for seven measures, a further one unit meeting the threshold for six measures, and four units meeting the threshold for five measures. Therefore, using a strict all-or-none approach would not have been feasible, and to conduct my analyses I had to use more pragmatic cut-offs as described in Sections 7.1.3 and 7.2.3.

OptiPrem has the relevant data to be able to conduct a similar analysis on a patient level using my non-NNAP MQC. The purpose of this would be to assess whether babies born between 27-31 weeks of gestation who received the specific evidence-based care had significantly different mortality and morbidity outcomes. If we did, it would be interesting to note if by adjusting for these care practices, it helps remove some of the heterogeneity in outcomes seen between the units that care for these babies, and if so, to what extent.

10.2 What is novel about this PhD work?

- Systematic review looking at outcome of babies born between 27-31 weeks of gestation, or with birthweight between 1000-1500g, by level of neonatal unit of birth or that they receive care in.
    - My review revealed a lack of evidence to answer the clinical question whether preterm babies born between 27-31 weeks of gestation have different outcomes if born in / cared for in neonatal units that provide secondary or tertiary level care.
    - This is especially of relevance in the UK given the structure of neonatal healthcare, where this cohort of preterm babies can receive care in LNU and NICU (formally level 2 and 3 units, respectively).
    - Narrative review published in BMJ Paediatrics Open (1).
- Exploring heterogeneity of outcomes for babies born between 27-31 weeks and understanding these in the context of fetal biology.
    - Babies born between 27-31 weeks of gestation are an understudied group despite accounting for ~12% of all preterm babies born in England. In part, this is because they fall between those born <27 weeks (i.e., extremely preterm babies), and the more mature 'moderately preterm' or 'late preterm' babies. In studies that do include them, they are often grouped together.
    - Reviewing the literature and from personal communication with national statistical bodies, I presented gestational week specific outcome data for mortality and major neonatal morbidities. This highlighted the degree of heterogeneity present within this cohort of babies, especially when comparing babies born at 27 weeks with those born at 31 weeks of gestation.
    - Furthermore, these findings were presented in the context of developmental changes that occur in utero during this five-week period, and the disruption caused by preterm birth and postnatal care.
    - Review article published in Journal of Neonatal Nursing (300).

- Assessing quality of care delivered by units to babies born between 27-31 weeks of gestation.
    - To my knowledge, there are no previous studies assessing quality of care delivered specifically to this cohort of preterm babies within the UK.
    - To do this, I categorised neonatal units (LNU and NICU) according to the proportion of their babies born between 27-31 weeks of gestation that they provided specific evidence-based processes of care to, relating to the peripartum period.
    - I compared the groups of units resulting from these categorisations with regards to differences in unit characteristics and patient demographics to analyse what this told me about provision of evidence-based care.
    - I also used adherence with, and data completion for the NNAP audit measures to categorise units.
    - I looked for associations between my categorisation of units according to the quality-of-care measures, and mortality and length of stay.
    - Especially with regards to the NNAP audit measures, they have not been used in this way before.
    - The multivariate analysis I conducted indicated that babies born between 27-31 weeks of gestation in units which provide good quality of care (as defined by the quality-of-care measures I have used), have, on average, a lower length of stay by one day, compared to similar babies born in other units.
    - Review article published in BMJ Open Quality (301) describing methodology of using compliance with and data completion for NNAP audit measures to look for associations with clinical neonatal outcomes.

10.3 Strengths and weaknesses

10.3.1 Data source

Both the greatest strength and weakness of my work was the data source I used. As described in Section 6.1, NDAU collates data entered by frontline healthcare staff (doctors and nurses) into the electronic patient record (BadgerNet) used in every neonatal unit within England, Scotland and Wales. It is a database containing national information on approximately 1 million babies and 10 million days of care. It is the system used for reimbursement of NHS Trusts for the neonatal care they provide, thereby providing a strong incentive for complete, accurate data entry, as can be seen by the very low levels of missing data for the non-NNAP MQC, NNAP audit measures, and outcomes I used for my analyses.

However, there is still a great degree of variation in this, and as described in Sections 7.1.2, 7.2.1, and 7.3, setting a threshold for missing data at 10%, there were certain demographic details and MQC I was not able to use. Better quality and completion of BadgerNet data would have improved my analyses, both univariate and multivariate. This would have allowed me to use Measures 5 (milk) and 6 (cord) of my non-NNAP MQC, so I could better categorise units using the combination of six evidence-based measures rather than four. It could also have allowed me to use bloodstream infection, an intermediate outcome which is an important measure of quality of neonatal healthcare due to evidence showing that introduction of relevant care bundles can reduce rates significantly, and the impact it has on mortality and length of stay (302, 303).

10.3.2 Adjustment for confounding factors

The multivariate analyses I conducted was limited to the cohort as a whole, and not for babies born at each gestational week. Regarding my non-NNAP MQC, the multivariate analysis was for the combination of Measures, and not for each individual measure. I have detailed in Section 7.4.2 the reasons why the multivariate analyses was limited to the most relevant groupings and outcomes, and did not replicate the entirety of the univariate analyses.

The regression models I used show that less than half of the variance in outcomes was explained by the demographic variables being used. This related to the previous point on

data completion and was expected, since we know many important variables were not able to be used due to high proportion of missing data. For example, maternal health (38.8%), and parameters related to the condition of the baby in the immediate postnatal period (5-minute Apgar scores - 10.7%, umbilical cord blood gas results - 60.5%, worst base excess in first 24 hours -100%). However, even if the data had been more complete and I was able to adjust for more confounders, this would not have helped against unknown confounders, or known confounders for which the level of missing data was still too high. For this, more complicated statistical analyses, e.g., case matching or instrumental variables, as used by the statisticians in WS1, would be required. Therefore, the results of my multivariate analyses must be interpreted with caution and cannot be taken at face-value.

10.3.3 Timeframe of study

Due to using national data from NDAU, I had a very large dataset to work with, enabling me to do comparisons and look for associations by each gestational week of birth within the OptiPrem cohort of babies. However, for certain variables the patient number was still low, e.g., babies requiring significant resuscitation, pre-discharge mortality, and this is in part a reflection of the patient cohort itself. In the UK, babies born between 27-31 weeks generally survive and especially for the more mature babies within this cohort, do not require a significant degree of intensive care, if any (see Section 3.3). To tackle this, I could have used data from 2014 to 2018 rather than just 2018. However, as described in Section 6.2.1, this was an active decision because I was interrogating care practices and adherence with national guidelines, as surrogate markers of quality of care. Not only can and do guidelines change year by year, so too do neonatal units, in terms of which areas of care they are focussed on improving and the methods being used. However, having conducted these analyses using data from 2018, future work could use a longer time span.

10.3.4 Use of quantiles

The aim of my analyses was to find a way to categorise units based on the quality of care they provided babies born between 27-31 weeks of gestation. As surrogates, or markers of quality of care, I chose my non-NNAP MQC and used the NNAP audit measures. Based on adherence with, and missing data for these measures, I sorted the list of neonatal units into hierarchical order. For many of the NNAP audit measures there are targets set

for adherence. Therefore, I was able to use these to categorise units into those that met the audit threshold versus those that did not. For benchmarking audit measures which do not set targets, I used quartiles to categorise the units into those that were in the top quartile versus the rest. I chose not to use quintiles because this would have further reduced the number of units (and therefore the number of babies) in the top grouping. Following categorisation of my units, I compared unit characteristics and demographic profiles of the babies between these two groupings and looked for associations with outcomes. Commonly, the top quartile is compared to the bottom quartile. I chose not to do this because, when categorising units according to data completion, for several measures there were significantly more than a quartile of units that had 100% complete data. Furthermore, comparing units in the top quartile versus the rest, was more in keeping with my categorisation of units according to adherence with NNAP audit measures, into those that met the set target and those that did not. Therefore, this method was used also for adherence with, and data completion for my non-NNAP MQC.

As with whatever system of unit categorisation I would have used, there are negatives associated with the one I chose. When comparing the top and bottom quantile they are at either end of the measuring scale and so if there is a true difference in the populations regarding demographic profiles or outcomes, that will be easier to uncover. By comparing units in the top quantile versus the rest (therefore, including those on either side of the arbitrary threshold or set target), this can reduce the likelihood of finding any such differences between the groups. To investigate this further, I carried out a sensitivity analysis, comparing outcomes for units within the top and bottom quartiles for my non-NNAP MQC and for analyses involving a combination of measures (e.g., NNAP audit measures), using more extreme definitions to separate out my groupings of units. For example, for adherence with NNAP audit measures I compared units meeting threshold/in top quartile for 5 or more measures with those meeting threshold/in top quartile for <3 measures. My previous analysis involved comparing units meeting threshold/in top quartile for 4 or more measures with those meeting threshold/in top quartile for <4 measures. Therefore, whereas previously I was comparing 21 units (group 1) with 70 units (group 2), now I was comparing 7 units (group 1) with 40 units (group 2). My sensitivity analysis did not reveal any additional significant results but some of the previous associations were no longer found, probably as a result of the reduction in patient numbers within the bottom quartile group.

This discussion touches upon another problem inherent in using quantiles to group continuous data. Each quantile covers a range of values, but the assumption is being made that within each quantile the values are homogenous (304). If the data is not symmetrically distributed, the breadth of this range can also vary quite significantly between quantiles. This can mean comparing and summating quantile related data is not straightforward. This was especially evident when looking to categorise units according to data completion for multiple NNAP audit measures. Some of the measures had very low overall missing data, and so missing data for a single patient or two would affect whether a unit was in the top quartile or not. Units within the top quartile for these measures were not comparable to those in the top quartile for other measures which high levels of missing data. Therefore, I excluded measures where the overall level of missing data was <10%, and for included measures, did not include units that were otherwise falling within the top quartile if their level of missing data was <10%. An alternative to using quantiles would have been a multivariate regression model in which the units were not categorised into two groups, but with data for adherence/data completion inputted for each unit, alongside demographic factors and outcomes.

10.3.5 Choosing quality of care measures

As described in Section 6.2.2, my choice of NDAU-data based, non-NNAP MQC was largely determined by the presence of strong supporting evidence. In part, this was because the PhD did not include the scope to use a consensus approach in, for example, a modified Delphi process. If I had been able to do this, it could have resulted in additional MQC to categorise units. However, this process would still have been limited by what constitutes the Neonatal dataset (NDS), which is by its very nature designed to capture information regarding process. For this purpose, to interrogate practice in UK neonatal units, it is the best source of data available, but for investigating aspects of structure, especially organisational culture, its data cannot be used as anything other than a surrogate (as I did using adherence with, and data completion for the NNAP audit measures).

To give a specific example, within the three-tier regionalised neonatal healthcare system in the UK, it is possible to argue that LNU could be split into two levels. This is because, largely depending on volume of patients and activity level, there are two models of

staffing employed. Some LNU are staffed from the same pool of doctors and nurses as work on the paediatric unit in the same hospital (as is the staffing model used in SCU), while other LNU are staffed from a separate pool of doctors and nurses that work specifically on the neonatal unit (as is the staffing model used in NICU). It would have been interesting to separate these two types of LNU, to see if their adherence with, and data completion for my non-NNAP MQC and the NNAP audit measures differ, and if they have a difference in outcomes. However, this data is not entered onto BadgerNet and so not collected by NDAU.

To collect this sort of data requires something like a questionnaire, and the work done by GIRFT shows what is possible when there are mechanisms in place to ensure a very high response rate. Obtaining a sense of how engaged units are with quality improvement work, patient centred care, effective teamworking, leadership, and problem solving, are likely as important, if not more so, than which mode of invasive ventilation is preferred, whether CPAP or high flow is used post-extubation, and whether diuretics are used to treat babies with CLD, for example. However, a questionnaire is not best suited for collecting data on clinical practice since the answers may reflect guideline recommendations and not what is actually done.

Therefore, by using NDAU data, and a 10% threshold for acceptable degree of missing data, I was able to categorise all units based on an accurate portrayal of their practice but were limited with regards to what aspects of process and structure I could use (see Appendix II for proposed questions used in pilot questionnaire). Future work on interrogating quality of neonatal healthcare should probably use a combination of the two methods to combine their strengths and compensate for their individual weaknesses.

10.3.6 Parental and healthcare staff perceptions of quality of care

Another way I could have measured quality of neonatal healthcare was through interrogating the subjective experience of those receiving and delivering care, i.e., parents and healthcare staff (nurses and doctors).

Such surveys are used ubiquitously throughout healthcare (305), including neonatal services (306, 307). The questions could be based on previous such questionnaires or designed anew using a Delphi process. The most significant difficulty, as with any

questionnaire study, would be to obtain an adequate number of respondents. To assess perceptions of quality of care received/delivered would require replies from multiple parents/healthcare staff at a significant majority of units caring for my patient cohort (i.e., LNU and NICU). Otherwise, there is increased risk of self-selection bias. In 2014, the Picker institute, in consultation with Bliss and representatives from neonatal networks in England, conducted a national survey of parent's experience of neonatal care (307). They mailed a paper-based self-completion questionnaire to 15944 parents from 88 English neonatal unit, receiving only 6000 replies (37.6% - ranging from 9%-59% from individual, participating units). In their reports to units, they felt the need to exclude results for questions with less than 20 replies. Therefore, it would not have been feasible to include this work within this PhD, but it could form the basis of possible future work, perhaps in conjunction with WS4 (the qualitative branch of OptiPrem which looked at parents' and clinicians' perspectives regarding place of care, place of care decisions and transfers) (5).

The results from such work could be used to categorise units and as per my analyses, look for differences in demographics and unit structure between groupings, and for associations with outcomes. However, in contrast to more objective measures of quality of care involving process and structure, we might have less of an expectation of finding associations between outcomes and perceptions of quality of care (308). Parents of babies admitted to the neonatal unit but are generally well can still experience care they are dissatisfied with, just as parents of babies who die or suffer significant morbidity outcomes might be very satisfied with the care they receive. This is because there are many factors other than care quality involved in determining outcomes. Again, this supports the argument that quality-of-care measures are not validated by associations with outcomes. Just because an association was not found between parental satisfaction with the care their baby has received, or staff perception of the quality of care they deliver, does not mean we should not strive to provide good quality of care.

Of more value would be to try and identify differences within the organisation culture and structure of units leading to the range of scores. This could be via further qualitative research interviewing parents and healthcare staff from units within the top versus the bottom quantile for the respective parameters. It would be interesting to assess the degree

of overlap in unit scores for parent satisfaction and healthcare staff perception of the quality of care delivered.

Furthermore, I could look for associations between units with better parent/staff scores and adherence/data completion for my non-NNAP MQC and NNAP audit measures. We might expect to find that units with better adherence/data completion for evidence-based care and national guidance in the form of the NNAP audit measures also have better parent/staff scores (309). However, the relationship might be more complex, since patient satisfaction and healthcare staff perception of quality of care is partly dependent on factors such as social/cultural/religious background which forms their expectations of what 'good quality' healthcare entails (310). In other words, two units could provide the same care (regarding structure and process), but it is perceived as good quality by parents and healthcare staff in one unit, and poor in the other.

This can go on to impact the actual quality of care delivered so that it improves in the one unit and worsens in the other. If parents believe their babies are receiving good quality of care this will impact on their behaviour, resulting in friendlier interactions with healthcare staff and an overall more positive mood in the unit. Similarly, doctors and nurses who believe the unit they are working in provides good quality of care will take greater pride in their work and strive to maintain and even improve on the care they provide (193). The opposite is also true. Parents who believe their baby is not receiving good quality of care will be more anxious and defensive, and their interactions with healthcare staff (who might label them as 'difficult' parents) will be more strained. This can result in an overall more negative mood in the unit. Healthcare staff that feel undervalued and demotivated, overworked and unsupported, can go on to provide worse quality of care (311-313). This stresses the importance of not only focussing on aspects of structure or process when measuring quality of care but creating a more holistic picture by integrating with this the perception of those delivering the care and those receiving it.

10.3.7 Data sorting for length of stay (LOS)

As described in section 7.4.1, to analyse LOS data, babies who died pre-discharge were excluded, since units which had a higher proportion of these babies would have an artificially reduced mean LOS.

Units with reduced mortality (e.g., due to providing more effective emergency/intensive care and keeping more babies alive, or because they had a culture where it was less likely reorientation of care to palliation/comfort care would take place), could have increased LOS. Units with increased mortality (e.g., due to being less proficient in providing life-saving emergency/intensive care or because they were more proactive about reorientation of care), could have reduced LOS. However, this would be of relevance if I was comparing a combined outcome of pre-discharge mortality and LOS for units.

This is also dependent on obstetric care (which was not under the remit of the OptiPrem study and so for which I did not have any data), since obstetric units that have a reduced intrauterine death (IUD) rate, may be delivering babies that die in the neonatal period or have severe hypoxic-ischaemic encephalopathy (HIE) and prolonged LOS. Possible surrogate data I had from the NNRD that could tell me about the condition of a baby at birth (e.g., five-minute Apgar scores, cord blood gas result, worst base excess in first 24hrs) was too poor to be included in my multivariate analyses. In theory, if I could have adjusted for these factors, I would not have needed to exclude babies who died pre-discharge from the LOS analyses. It was pointed out in the PhD viva voce examination that the effect of excluding babies who died pre-discharge on LOS analyses could have been tested by a sensitivity analysis in which I did not exclude those babies.

10.3.8 Choice of statistical tests

Feedback from the PhD viva voce examination highlighted that it would have been more appropriate to use Chi squared test for trend rather than the standard Chi squared test whenever analysing differences between IMD_Q scores between groupings of units, due to the IMD_Q categories being ordered. However, this was very unlikely to yield significantly different results.

It was also pointed out that for length of stay (LOS) data, median should have been used instead of mean, since the data was skewed and not normally distributed, and therefore a non-parametric test used to compare LOS between groupings of units. However, this was only an issue for the univariate analyses since to meet the requirements of the multivariate analyses, LOS data was transformed using natural log which resulted in a normal distribution.

10.4 Conclusion

What is the take-home message from this PhD? On the face of it, it would seem to be the positive association found in the multivariate analyses, between units that comply with my evidence based non-NNAP MQC and/or NNAP audit measures, and a reduction in length of stay. However, as previously discussed in Section 9.7.2, further statistical analyses need to be undertaken to account for important known confounding factors (currently excluded due to high proportion of missing data), and potentially important, unknown confounding factors, to determine whether these associations are real.

In the absence of results from such analyses, let us consider both possible outcomes. If the association is true, greater adherence with national guidance (in the form of the NNAP audit measures), and practice of evidence-based care (in the form of my non-NNAP MQC), are associated with an average reduction of one day in length of stay for babies born between 27-31 weeks of gestation. It is likely that an association with LOS is also present for preterm babies of other gestational ages (i.e., <27 weeks and >31 weeks). This has significant implications for cot capacity in neonatal units and transfer of babies, and overall cost-saving for the NHS (as discussed in Section 9.3).

Now, let us consider the possibility that after such analyses, the associations between adherence with NNAP audit measures and/or non-NNAP MQC and length of stay are not statistically significant. Does this mean that adherence with my non-NNAP MQC and the NNAP audit measures does not constitute good quality of care, and so neonatal units can stop wasting time and effort on them? In the multivariate analyses, I did not find an association between data completion and outcome, so similarly, does this mean good quality data entry does not constitute good quality of care?

We often assume that the gold standard in assessing validity of MQC is to find associations with outcomes. However, if there is strong evidence supporting the structure or process we are using as a quality of care measure, does it matter if we can find an association between adherence and outcomes? Even if we cannot, does that mean it does not constitute good quality of care and does not need to be practised? Or would we have to assume there was a problem in our analysis (e.g., unaccounted for confounding factors)?

Similarly, for measures relating to aspects of structure and processes of care that do not directly impact outcomes, yet by consensus of expert opinion constitute good quality of care, how would we assess their validity, since no analysis is likely to find a positive association with outcomes? Would this mean we stop using such measures as markers of quality of care? Therefore, even though we might expect individuals/units/hospitals that strive to provide good quality of care to have better outcomes for their patients, quality of care measures are not validated or dependent on finding associations with outcomes.

So why carry out such analyses in the first place? Finding a positive association between adherence with national guidance, evidence-based practice, good data completion and better outcomes can serve as a stimulus for neonatal units to better engage with such processes. For the NNAP audit measures, this is the first time such an analysis has been carried out. Finding such associations can also be helpful when investigating differences in outcomes between seemingly similar healthcare providers, even after adjusting for known confounders. Prior to the disruption caused to the OptiPrem study due to the Coronavirus pandemic, it was envisaged that WS1 would use my categorisation of units according to adherence with and missing data for my non-NNAP MQC and the NNAP audit measures in this way. This is now not going to happen as part of the original study, but as follow-on work.

As discussed in Section 6.2.2 and Appendix III, my non-NNAP MQC are all evidence-based processes of care which ensure preterm babies get the best start to life to maximise their chances of better outcomes. This is reflected in the subsequently published BAPM *'Antenatal Optimisation for Preterm Infants less than 34 weeks'* quality improvement toolkit (314), which includes antenatal steroids, optimal cord management, normothermia, and optimising early maternal breastmilk. They also mention that *'additional perinatal optimisation interventions such as respiratory management are recognised, but are not yet included in this pathway'*.

Therefore, the results of the multivariate analyses are important in serving as an encouragement for neonatal units to strive to provide good quality of care by complying with NNAP audit measures and practicing evidence-based medicine (by showing that this can result in a reduction in length of stay for preterm babies born between 27-31 weeks

of gestation). However, even if the multivariate analyses had not found any significant associations, units should still strive to provide this level of care.

So, perhaps the more important outcome from my work is what we learn about differences between units that practice more evidence-based care and have an organisational culture that means they strive to comply with national guidance, and those that do not. When categorising units according to the combination of my non-NNAP MQC, units in the top quartile for two or more measures were more likely to be LNU. As discussed in Section 9.1.3, GIRFT has recently published data showing that nurse staffing in LNU is in general, less inadequate than in NICU. And my own data showed that just over double the number of preterm babies were born, on average, per NICU compared to LNU, at least for babies born between 27-31 weeks of gestation. This would indicate that an important factor in the ability of a neonatal unit to deliver evidence-based, good quality of care, is a more adequate nurse-to-baby ratio. Certainly, this is well supported by the evidence, as described in section 6.2.2. While LNU may be doing a better job of this than NICU, that is relative, as exemplified by the data for adherence with Measure 4 (nursing). For babies born between 27-31 weeks of gestation that required intensive care on day 1 of life, only 10.8% of them had the BAPM recommended, 1:1 nursing care (Section 8.1.2.4).

Furthermore, units in the top quartile for two or more measures were also more likely to have a more affluent and less deprived population. In fact, this was the most consistent relationship found between groupings of units when categorising them according to my non-NNAP MQC and NNAP audit measures. As previously discussed in Section 9.1.3, this relationship is unlikely to be direct and unifactorial, and more likely to be related to a systemic difference in the structure and processes of care employed by units and the wider healthcare system in more affluent, compared to more deprived areas. It was also true that significantly more babies in the most deprived IMD_Q quintile were born in NICU compared to LNU, and as discussed in Section 9.1.3, this is likely related to NICU being more likely to be situated in inner-city, more deprived populations than LNU.

These problems are not new. Weightman et al. (288) conducted a systematic review and meta-analyses of longitudinal and record-linkage studies published between 1994 – 2011 that included socioeconomic data and health outcomes for UK infants (e.g., preterm birth,

birthweight, morbidity and mortality, use of primary or secondary healthcare, development and growth). They excluded studies investigating children with congenital anomalies or pre-existing illness, or children in social care or adopted. This was so the review population was representative of the general population. Their search strategy found 5173 citations, of which 88 papers were examined in full. Of these 36 were selected for inclusion (13 prospective cohort or case-control, 23 retrospective cohort or routinely collected data, i.e., record linkage). To be included, a study needed to compare (using a statistical test), risk of a specific health outcome based on a measure of socioeconomic status of an individual or geographical area. Studies that used similar outcomes and deprivation measures were included in the meta-analysis. While there was a high degree of heterogeneity in the included studies, the effect direction was generally uniform, resulting in significant combined odds ratios of unadjusted data. In seven studies that looked at birthweight outcome, the combined OR when comparing most to least deprived areas was 1.81 (95% CI 1.71-1.92). In six studies that used preterm birth as an outcome, the combined OR was 1.72 (95% CI 1.59-1.86). In three studies looking at neonatal mortality, the combined OR was 1.42 (95% 1.33-1.51). In three studies looking at post-neonatal infant mortality, the combined OR was 2.31 (95% CI 2.03-2.64). These findings were in keeping with systematic reviews of international studies (315, 316). The 2021 MBRRACE report on deaths in 2019 found that compared to women living in the least deprived areas of the UK, women living in the most deprived areas were twice as likely to have a stillbirth and the risk of neonatal death was 73% higher (293). This excess risk had increased from 2015 to 2019 despite overall rates of stillbirth and neonatal mortality decreasing.

Of note, there is a recognised link between ethnicity and deprivation, ethnicity and quality of care (including neonatal healthcare), and ethnicity and outcomes. However, I was not able to interrogate this due to significant amounts of missing data regarding maternal ethnicity. In the 2001 UK census, as a proportion, twice as many people from ethnic minorities lived in deprived neighbourhoods compared to other, less deprived parts of the city (317). In their 2015 report, the ONS found that babies with the highest infant mortality rates were those of Pakistani, black Caribbean and black African parents, and were also more likely to come from areas of high deprivation (318). The 2021 MBRRACE report found that stillbirth rates and neonatal mortality for Asian and Asian British, and Black and Black British ethnicity women was 43-60% higher than for White

women (293). In a systematic review by Sigurdson et al. they found evidence that ethnic minority infants in the US experienced worse quality of care in terms of structure (nurse staffing levels, delivery in hospital with better outcomes) and process (kangaroo care, breastfeeding support and breastfeeding support at discharge, follow-up post discharge) (319).

Regarding nurse staffing, in the 2015 report by the Bliss baby charity (320), they reported on the state of neonatal healthcare within England. They collected their data via a survey, for which they received responses from 101 neonatal units (63%) and 14 neonatal transport services (100%). Using the latest national guidance (from the DOH, NICE and BAPM), they created a snapshot of staffing levels based on reports for a single day. Of the 81 units that provided data on nurse staffing, 52 (64%) were understaffed, but this was most noticeable for NICU (26 out of the 30 – 87%). Regarding specialist nurses, 59 out of 91 units did not have required numbers to meet standards, and the proportion of specialist nurses to all nurses had actually fallen by 19% since their last report in 2010. Nor do these problems regarding healthcare inequality and understaffing only affect neonatal medicine (289, 290, 321-323). It was beyond the scope of this PhD to explore causes and possible solutions.

Therefore, my findings indicate that inadequate nurse staffing and healthcare inequality will be roadblocks in the ability of many neonatal units improving the care they provide, including implementation of the BAPM guidance around optimising perinatal care. It would be interesting to explore whether similar relationships are found for other gestational age ranges. It is likely this will be the case, which further stresses the importance of tackling these two intransigent problems, the benefits of which will not only been seen within neonatal medicine, but the wider healthcare system.

10.5 Future work: planned and possible

10.5.1 Adjusting for confounding factors

In conjunction with a statistician, further analyses of my data will involve finding an appropriate method to adjust for confounding factors beyond the multivariate analyses I have conducted. This will have to take into account that several important confounders have a significant amount of missing data, and so more complex methods such as matching or instrumental variables might need to be considered.

When further analyses are conducted, statistical advice can also be sought on the need for a Bonferroni correction. As can be seen in Sections 8.1.4.1.3 and 8.1.4.2.2, the high number of statistical comparisons increases the risk of a Type I error (i.e., of a false positive), since even with the standard p value of 0.05, that still means 1 in 20 tests will give a statistically significant result by chance. To counter this, the Bonferroni Correction involves dividing the p value by the number of analyses, thereby reducing its value and the risk of finding an association by chance. The downside of using the Bonferroni Correction is that it increases the risk of Type II errors, i.e., false negatives.

10.5.2 Looking for associations between quality of care and major morbidity outcomes

Future statistical work will also involve looking for associations between units categorised according to my quality-of-care measures and significant neonatal morbidity outcomes, as was done in WS1. This includes necrotising enterocolitis requiring surgery, retinopathy of prematurity ≥grade III or requiring treatment, intraventricular haemorrhage ≥grade III, periventricular leukomalacia, porencephalic cysts, hydrocephalus, and bronchopulmonary dysplasia.

10.5.3 Incorporating results of WS2 into WS1

The categorisation of units based on my non-NNAP MQC and NNAP audit measures will be incorporated into analyses carried out by WS1 as confounding factors, to be adjusted or matched for. This will allow us to explore to what degree the quality-of-care units provide (insofar as is measured by these parameters) accounts for heterogeneity in outcomes seen between units of the same designation.

10.5.4 Expanding the patient cohort

I could also consider conducting similar analyses for babies born at other gestational age ranges, e.g., 23-26 weeks. It would be interesting to note similarities and differences within results when categorising units according to my non-NNAP MQC and NNAP audit measures as applied to this cohort of babies, and comparing demographic profiles / unit characteristics, and looking for associations with outcomes.

If expanding the patient cohort to all neonatal unit admissions, it would be possible to conduct the analyses just using publicly available NNAP data, since the audit measures already include some outcomes (i.e., BPD, NEC), which are to be expanded in future to include pre-discharge mortality (324) and neonatal preterm brain injury (325). Alternatively, NNAP data (in terms of adherence and missing data for non-outcome audit measures) could be linked to mortality outcome data from MBRRACE, which has the advantage of being stabilised and adjusted (as described in Section 3.4.7).

Another way of expanding the patient cohort would be to use the same gestational age range, but include babies discharged between 2014-2018 rather than just 2018. This would increase the sensitivity of analyses looking for associations with outcomes, especially when separating by gestational week of birth, due to the relatively low frequency mortality occurs in babies born between 27-31 weeks.

10.5.5 Using results from Getting It Right First Time (GIRFT): Neonatology, to categorise units

As part of a national programme involving many medical and surgical specialties, GIRFT: Neonatology used a questionnaire method to collect a large amount of data regarding structure and process withing UK neonatal units, with the aim of assessing the current situation on the ground, and how it can be improved going forward. As discussed in Section 9.7.5, a questionnaire can collect relevant data for assessing quality of care that may not be possible or available from other data sources, e.g., NNRD. Future work could look at using data from 'GIRFT: Neonatology' to categorise units according to aspects of structure and process felt relevant to quality of care provided and look for differences in demographic profiles / unit characteristics that could help explain those differences, and for associations with outcomes.

GIRFT data could also be used to analyse my results. Data for units in Group 1 (for adherence with NNAP audit measures and my non-NNAP MQC) could be compared with data for units in Group 2, to try and identify any differences in process and structure which might provide further explanation as to why babies in Group 1 had a reduction in LOS compared to babies in Group 2.

10.5.6 Measuring adherence with new BAPM perinatal optimisation care pathway

With the relatively recent release of the BAPM toolkit for perinatal optimisation (235), future work could include measuring adherence with its different components (both obstetric and neonatal) as a marker of the quality of care provided to babies born under 34 weeks of gestation, in the perinatal context.

10.5.7 Working with WS4 to interrogate perceptions of quality of care for parents and healthcare staff

In conjunction with WS4, a questionnaire study could be planned to explore parent and healthcare staff perceptions of the care they received/delivered (as discussed in Section 9.7.6). The results could be used in conjunction with my data to obtain a more holistic picture of the quality of neonatal healthcare provided in LNU and NICU.

11 Appendix I: Systematic review

11.1 Recurring words and phrases in titles of read articles

Articles comparing neonatal outcomes by level of neonatal unit	
Phrases containing 4 words	**Occurrences**
very low birth weight	10
low birth weight infants	10
of very low birth	4
extremely low birth weight	3
low birth weight infant	3
weight infants born in	2
regionalization of perinatal care	2
and very low birth	2
of neonatal intensive care	2
the regionalization of perinatal	2
preterm and very low	2
on mortality of very	2
by level of hospital	2
and level of care	2
for very low birth	2
birth weight infants in	2
mortality of very low	2
birth weight infants born	2
and neonatal mortality in	2
of hospital of birth	2
to place of birth	2
Phrases containing 3 words	**Occurrences**
low birth weight	15
very low birth	10
birth weight infants	10
perinatal care in	5
the effect of	4
of very low	4
by level of	3
level of care	3
level of hospital	3
regionalization of perinatal	3
extremely low birth	3
hospital of birth	3
place of birth	3
and neonatal mortality	3
neonatal mortality in	3
birth weight infant	3
neonatal intensive care	2
and very low	2
mortality of very	2
mortality in very	2
of birth on	2
infants born in	2
evaluation of a	2
preterm and very	2
of hospital of	2
of perinatal care	2
to place of	2
on mortality of	2
normal birth weight	2
on neonatal mortality	2
weight infants born	2
weight infants in	2
and level of	2
the regionalization of	2
for very low	2
of a national	2
of neonatal intensive	2
Phrases containing 2 words	**Occurrences**
birth weight	19
low birth	15
very low	11
weight infants	10
level of	8
of birth	8
perinatal care	8
neonatal mortality	7

of hospital	6
care in	6
mortality in	5
of perinatal	4
on mortality	4
the effect	4
place of	4
and neonatal	4
of very	4
care and	4
effect of	4
of a	3
extremely low	3
of care	3
intensive care	3
in finland	3
regionalization of	3
infants in	3
mortality rates	3
by level	3
weight infant	3
high risk	3
hospital of	3
and very	3
a national	3
perinatal mortality	3
impact on	3
outcome for	2
study of	2
extremely preterm	2
in sweden	2
perinatal regionalization	2
evaluation of	2
of extremely	2
for extremely	2
born in	2
mortality the	2
preterm and	2
transport in	2
south carolina	2
hospital level	2
volume and	2
infants born	2
study group	2
the regionalization	2
on neonatal	2
of neonatal	2
birth on	2
infants of	2
to place	2
for very	2
in very	2
level and	2
mortality of	2
of delivery	2
and level	2
normal birth	2
outcome of	2
survival of	2
neonatal intensive	2
regionalization and	2
washington state	2
in washington	2
in the	2
of the	2
preterm infants	2
hospitals in	2

Articles comparing neonatal outcomes by in-utero vs. ex-utero transfer	
Phrases containing 4 words	**Occurrences**
very low birth weight	5
low birth weight infants	5
survival and place of	2
and very low birth	2
of preterm and very	2
transport to a regional	2
preterm and very low	2
to a regional perinatal	2
maternal and neonatal transport	2
birth weight infants in	2

a regional perinatal center	2
Phrases containing 3 words	**Occurrences**
low birth weight	6
very low birth	5
birth weight infants	5
and very low	2
comparison between maternal	2
tertiary care centers	2
survival and place	2
mortality and morbidity	2
to a regional	2
and neonatal transport	2
preterm and very	2
outcome of preterm	2
maternal and neonatal	2
birth weight 500	2
transport to a	2
and place of	2
regional perinatal center	2
a regional perinatal	2
of preterm and	2
weight infants in	2
a comparison between	2
versus neonatal transport	2
Phrases containing 2 words	**Occurrences**
neonatal transport	8
birth weight	8
low birth	6
weight infants	5
very low	5
in the	4
of preterm	4
and neonatal	3
a comparison	3
versus neonatal	3
outcome of	3
a regional	3
maternal and	3
comparison between	3
effect of	3
to a	2

perinatal center	2
and place	2
mortality and	2
preterm and	2
between maternal	2
transport to	2
care centers	2
place of	2
survival and	2
of neonatal	2
infants in	2
of inborn	2
maternal transport	2
weight 500	2
in utero	2
survey of	2
neonatal survival	2
infants with	2
regional perinatal	2
preterm birth	2
and very	2
and morbidity	2
tertiary care	2

Articles comparing neonatal outcomes by volume of patients	
Phrases containing 4 words	**Occurrences**
very low birth weight	4
low birth weight infants	4
on mortality of very	2
of very low birth	2
mortality of very low	2
Phrases containing 3 words	**Occurrences**
very low birth	4
birth weight infants	4
low birth weight	4
neonatal intensive care	2
the effect of	2
mortality of very	2
population based study	2
size of delivery	2

on mortality of	2
in low risk	2
of very low	2
Phrases containing 2 words	**Occurrences**
low birth	4
weight infants	4
birth weight	4
very low	4
volume and	3
population based	3
the outcome	2
intensive care	2
on mortality	2
the effect	2
and neonatal	2
size of	2
risk adjusted	2
hospital volume	2
patient volume	2
low risk	2
of very	2
mortality of	2
of delivery	2
based study	2
of small	2
neonatal intensive	2
in low	2
neonatal mortality	2
of the	2
effect of	2

11.2 Quality assessment of studies included in systematic review using modified QUIPS tool

Type of study	Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Study confounding
In-utero vs. ex-utero transfer to level 3 or regional perinatal centre	Lamont et al.	• Single network study • Defined exclusion criteria (lethal congenital anomalies, infants transferred for surgical correction of congenital anomaly) • Comparison of baseline characteristics (GA, BW)	• Unclear whether retrospective or prospective • Completeness of data on demographic/confounding factors 100% • Outcome analysis for all babies 28-31 weeks meeting inclusion criteria	• Undefined birth location for transferred babies (from all referring hospitals to the single University Hospital)	• Survival to discharge	• Unadjusted for confounding factors
	Truffert et al.	• Population based study • Defined exclusion criteria (GA<25 and >33 weeks) • Comparison of baseline characteristics (private hospital, >1200 births/year, staff present at delivery, spontaneous labour, mode of delivery, presentation, multiple pregnancy, gender, Apgar score, temperature on transfer from delivery room)	• Retrospective • Completeness of data on demographic/confounding factors 99.5% (comparison only conducted for babies born between 31-32 weeks) • Outcome analysis for all babies 27-30 weeks meeting inclusion criteria	• Undefined birth location for transferred babies (included random selection of all hospitals with maternity units)	• Mortality, disability (CP, deafness, Brunet Lezine developmental score <80), survival without disability at 2 years	• Unadjusted for confounding factors
	Hauspy et al.	• Single network study • Defined exclusion criteria (GA<24 and >35 weeks) • Comparison of baseline characteristics (preterm labour, preterm PROM, IUGR/SGA, preeclampsia, placental abruption, placenta praevia, vaginal bleeding, gender, GA, BW)	• Retrospective • Completeness of data on demographic/confounding factors 100% • Outcome analysis for all babies 28-31 weeks meeting inclusion criteria	• Undefined birth location for transferred babies (from all referring hospitals to the single University Hospital)	• Neonatal mortality, RDS	• Unadjusted for confounding factors
	Lee et al.	• Population based study • Defined exclusion criteria (GA≥32 weeks, moribund, admission to NICU >4 days) • Comparison of baseline characteristics (GA, BW, Apgar score, SNAP-II score, SGA, multiple gestation, maternal hypertension, antenatal care, mode of delivery, antenatal corticosteroids, presentation)	• Prospective • Completeness of data on demographic/confounding factors 100% • Outcome analysis for all babies 27-31 weeks meeting inclusion criteria • Cranial US (and therefore analysis for IVH) available for 82% of infants	• Defined birth location for transferred babies (including level 1 hospitals where care can be provided by 'family physicians')	• Pre-discharge mortality, IVH (≥grade 3), ROP (≥stage 3), RDS, CLD, NEC, survival without major morbidity (IVH, CLD, NEC, ROP)	• Adjusted for confounding factors (GA, Apgar score, SGA, mode of delivery, multiple gestation, maternal hypertension, presentation, antenatal corticosteroids, antenatal care, SNAP-II score)
	Boland et al.	• Population based study • Defined exclusion criteria (GA<22 and >32 weeks, lethal congenital anomalies) • Comparison of baseline characteristics (maternal age, multigravida, multiple pregnancy, APH, hypertensive disorders of pregnancy, prelabour ROM, spontaneous preterm labour, mode of delivery, gender, GA, BW)	• Retrospective • Completeness of data on demographic/confounding factors 99.9-100% • Outcome analysis for all babies 28-31 weeks meeting inclusion criteria	• Defined birth location for transferred babies (including hospitals without special care or obstetric units and births before arrival at hospital)	• Infant mortality	• Adjusted for confounding factors (GA, BW, gender)

Level of unit of birth (level 3 or perinatal regional centre vs. lower level or local unit)	Holmgren et al.	• Population based study • Defined exclusion criteria (GA>37 weeks) • No comparison of baseline characteristics of population by level of unit	• Retrospective • Completeness of data on demographic/confounding factors 100% • Outcome analysis not carried out on entire population of babies meeting inclusion criteria (at 28-31 weeks possible lack of data on 13.4% of babies for neonatal and infant mortality)	• Comparing outcomes of tertiary and secondary level units • Explanation given of facilities available (e.g., level 2 units provide IC or have neonatal 'care units')	• Perinatal (up to 1 week), neonatal, and infant mortality, severe asphyxia (Apgar score <5 at 10 minutes of age)	• Unadjusted for confounding factors
	Johansson et al.	• Population based study • Defined exclusion criteria (GA<24 and >32 weeks, births at units without paediatric services) • Comparison of baseline characteristics (maternal BMI, smoking status, cohabitation with father, age, country of birth, placenta praevia, abruptio-placenta, APH, preeclampsia, hypertension, GDM, chronic diseases, GA, BW for GA, mode of delivery, gender, presentation, SGA, LGA, major congenital anomalies)	• Retrospective • Completeness of data on demographic/confounding factors 97.1-100% • Outcome analysis for 99.6% of babies 28-31 weeks meeting inclusion criteria (hospital of birth unknown for 9 babies)	• Comparing outcomes of university and general hospitals • Explanation of facilities available (e.g., general hospitals have 1000-5000 deliveries/year, similar obstetric and anaesthetic facilities to university hospitals, provide neonatal IC before transfer)	• Infant mortality	• Adjusted for confounding factors (mode of delivery, GA, BW for GA, gender, presentation, placental complications, maternal hypertension)
Level of care (level 3 or perinatal regional centre vs. lower level or local unit)	Field et al.	• Population based study • No exclusion criteria • Comparison of baseline characteristics (GA, BW, RDS, presentation, Apgar score, multiple pregnancy)	• Prospective • Unable to determine completeness of data on demographic/confounding factors • Unable to determine proportion of babies for which outcome analysed	• Comparing outcomes of 'large'/IC units and 'small'/SC units • Explanation of facilities available (e.g., some SC units provided IC to their own babies, others transfer out, none had out of hours middle-grade paediatricians on site, 5-420 ventilation days annually)	• Mortality period unspecified	• Unadjusted for confounding factors
	Jonas et al.	• Population based study • Defined exclusion criteria (GA <20 and >32 weeks, known BW) • No comparison of baseline characteristics of population by level of unit	• Retrospective • Completeness of data on demographic/confounding factors 93.3-100% • Outcome analysis for all babies 28-31 weeks meeting inclusion criteria	• Comparing outcomes of level 3 and non-level 3 units • No explanation of facilities available in different levels of units	• Neonatal mortality	• Adjusted for confounding factors (gender, mode of delivery, intubation, year of birth, multiple gestation, GA, BW, presentation, maternal age, parity, marital status)

Table 47 Quality assessment of studies characterising neonates by gestational age using modified QUIPS tool
Reproduced with permission from Ismail et al. (1)
GA (gestational age), BW (birthweight), NICU (neonatal intensive care unit), IC (intensive care), SC (special care), ROM (rupture of membranes), SGA (small for gestational

age), IUGR (intrauterine growth retardation), LGA (large for gestational age), SNAP-II (Score for Neonatal Acute Physiology (271)), US (ultrasound), CP (cerebral palsy), RDS (respiratory distress syndrome), IVH (intraventricular haemorrhage), CLD (chronic lung disease), NEC (necrotising enterocolitis), ROP (retinopathy of prematurity), APH (antepartum haemorrhage), BMI (body mass index), GDM (gestational diabetes mellitus

Type of study	Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Study confounding
In-utero vs. ex-utero transfer to level 3 or regional perinatal centre	Miller et al.	• Single network study • Defined exclusion criteria (BW <1000g and >1500g, lethal congenital anomalies) • Comparison of baseline characteristics (GA, presentation, premature ROM, vaginal bleeding, cervix >3cm, premature labour, mode of delivery, admission-delivery time, SGA)	• Retrospective • Completeness of data on demographic/confounding factors 76-100% • Outcome analysis for all babies meeting inclusion criteria	• Undefined birth location for transferred babies (from all referring hospitals to perinatal tertiary centre)	• Pre-discharge mortality	• Unadjusted for confounding factors
	Watkinson et al.	• Single network study • Defined exclusion criteria (BW>2000g, lethal congenital anomalies) • Comparison of baseline characteristics (preterm labour, pre-eclampsia, APH, PROM, abnormal CTG, IUGR)	• Retrospective • Completeness of data on demographic/confounding factors 100% • Outcome analysis for all babies with BW 1000g-1499g meeting inclusion criteria	• Undefined birth location for transferred babies (from all referring hospitals to perinatal tertiary centre)	• Neonatal mortality	• Unadjusted for confounding factors
	Obladen et al.	• Single network study • Defined exclusion criteria (BW>1500g) • Comparison of baseline characteristics (maternal age, parity, nationality, marital status, social index, anternatal steroids, time and mode of delivery, Apgar scores, umbilical artery pH, plurality, male gender, BW, GA, person providing primary care, endotracheal intubation, admission age, systolic BP, temperature, pH, BE, blood glucose)	• Prospective • Completeness of data on demographic/confounding factors 100% • Number of babies for which IVH outcomes given does not match total number of VLBW infants - data missing for 30% (maybe due to babies who did not have cranial US) • Figures for survival correspond to singleton births only, therefore multiple births (27% of VLBW population) excluded	• Undefined birth location for transferred babies (from all referring hospitals to perinatal tertiary centre, outborn infants may not have paediatrician present at birth and transported using in-house staff)	• Survival to discharge, IVH (grade III or IV)	• Adjusted for confounding factors (RDS, BW, IVH, pH at admission, GA, gender)
	Mohamed et al.	• Population based study • Defined exclusion criteria (BW>1500g, missing data for transport, transport >48 hours of age, congenital anomalies which can contribute to IVH or outcomes) • Comparison of baseline characteristics (ELBW, gender, ethnicity, RDS, sepsis, NEC, PDA, pulmonary haemorrhage, apnoea of prematurity, perinatal asphyxia, pneumothorax, PPHN, maternal hypertension, chorioamnionitis, breech delivery)	• Retrospective • Completeness of data on demographic/confounding factors 99.9% • From data provided not able to assess study attrition • ICD-9 diagnostic codes for grade of IVH not available for all patients, no details provided for how many patients had cranial US	• Undefined birth location for transferred babies (inter-hospital transfers, direction of transfer not defined)	• All IVH, severe IVH (grade III or IV)	• Adjusted for confounding factors (gender, ethnicity, ELBW, birth asphyxia, fetal acidaemia, apnoea of prematurity, RDS, PPHN, pneumothorax, pulmonary haemorrhage, PDA, sepsis, NEC, maternal hypertension, chorioamnionitis, APH, cord prolapse, breech presentation, instrumental delivery)

Level of unit of birth (level 3 or perinatal regional centre vs. lower level or local unit)	Gortmaker et al.	• Population based study • Defined exclusion criteria (BW>1501g) • No comparison of baseline characteristics	• Retrospective • Data from linked birth and death certificates • Completeness of data on demographic/confounding factors 99% • From data provided not able to assess study attrition • Figures for 750g-1500g BW do not correspond to total infants meeting inclusion criteria (by 33%) - could be due to numbers of infants with BW<750g	• Comparing outcome of level 3 and rural/urban units (grouping level 1 and 2) • No explanation of facilities available in different levels of units	• Early neonatal (0-4 days), neonatal, and infant mortality	• Adjusted for confounding factors (GA, plurality)
	Powell et al.	• Population based • Defined exclusion criteria (BW<501g and >2000g) • No comparison of baseline characteristics of population by level of unit	• Retrospective patient • identification with prospective follow-up • Completeness of data on demographic/confounding factors 53-100% • Outcome analysis for 97.7% of babies meeting inclusion criteria (32 lost to follow up) • Mortality figures only available for infants with BW<1500g, therefore 70 infants unaccounted for - probable deaths in 1501-2000g BW category	• Comparing outcome of regional and district hospitals • No explanation of facilities available in different levels of units	• Survival to 2 years of age	• Unadjusted for confounding factors
	Powell et al.	• Population based • Defined exclusion criteria (BW<500g and >2499g, hospitals without obstetric services, lethal congenital anomalies) • Comparison of baseline characteristics (plurality, maternal age, ethnicity, marital status, residence, smoking status, antenatal care, parity)	• Retrospective • Completeness of data on demographic/confounding factors 3.1-81.5% • Outcome analysis for 27.7% of babies meeting inclusion criteria	• Comparing outcomes of level 3 and level 2 units • Explanation of facilities available in different level units (e.g., level 2 units have ≥500 births/year, obstetricians and paediatricians, 1:4 maximum nurse:patient ratio)	• Infant survival	• Unadjusted for confounding factors
	Yeast et al.	• Population based study • Defined exclusion criteria (BW<500g, lethal congenital anomalies) • No comparison of baseline characteristics	• Retrospective • Outcome analysis for potentially all VLBW births meeting inclusion criteria	• Comparing outcomes of level 3 and level 2 units • Explanation of facilities available in different level units (e.g., level 2 units have >1000 births/year, anaesthetics available at all times)	• Neonatal mortality	• Adjusted for confounding factors (BW, ethnicity, plurality)
	Sanderson et al.	• Population based study • Defined exclusion criteria (BW<500g and >1499g, lethal congenital anomalies, births outside a delivery hospital, missing data on GA or birth hospital) • Comparison of baseline characteristics (maternal transfer, infant transfer, ethnicity, marital status, age, residence, education, antenatal care, year of birth, multiple birth, gender, BW, GA)	• Retrospective • Completeness of data on demographic/confounding factors 96.3-100% • Outcome analysis for all babies meeting inclusion criteria	• Comparing outcomes of level 3 and level 2 units • Explanation of facilities available in different level units (e.g., level 2 units have >500 births/year, care for infants >1500g BW and >32 weeks GA, can provide resuscitation, short term ventilation, exchange transfusion)	• Neonatal mortality	• Adjusted for confounding factors (ethnicity)

	Gould et al.	• Population based study • Defined exclusion criteria (BW<500g, non-hospital births, missing data on BW) • No comparison of baseline characteristics	• Retrospective • Multiple births and deaths due to congenital anomalies excluded (numbers undefined), therefore not possible to assess study attrition	• Comparing outcomes of regional NICUs (level 3) and intermediate NICUs (level 2) • Explanation of facilities available in different level units (e.g., level 2 units care for infants >1500g BW not requiring assisted ventilation)	• Neonatal mortality	• Unadjusted for confounding factors
	Warner et al.	• Population based • Defined exclusion criteria (BW<499g and >1499g, lethal congenital anomalies) • Comparison of baseline characteristics (BW, GA, ethnicity, sex, SGA, multiple gestation, Apgar score, maternal hypertension or preeclampsia, CRIB score, antenatal steroids, ante/intrapartum antibiotics)	• Retrospective • Completeness of data on demographic/confounding factors 100% • Outcome analysis for all babies with BW 1000g-1499g meeting inclusion criteria	• Comparing outcome of perinatal centres vs referring hospitals • Explanation of facilities available in different levels of units (e.g., non-perinatal centres do not have 24-hour on site physician for newborn care, some provide CPAP, mechanical ventilation only to stabilise for transport)	• Pre-discharge mortality or <120 days, BPD or death, severe IVH (grade III or IV) or death, ROP (requiring laser or cryotherapy) or death, NEC (Bell stage II or III) or death, mortality or major morbidity (BPD, severe IVH, severe NEC, severe ROP)	• Adjusted for confounding factors (GA, BW, gender, ethnicity, SGA, Apgar score, plurality, maternal hypertension/pre-eclampsia, antenatal antibiotics, glucocorticoids, CRIB score)

Table 48 Quality assessment of studies characterising neonates by birthweight using modified QUIPS tool
Reproduced with permission from Ismail et al. (1)
GA (gestational age), BW (birthweight), NICU (neonatal intensive care unit), CTG (cardiotocograph), ROM (rupture of membranes), BP (blood pressure), BE (base excess), SGA (small for gestational age), IUGR (intrauterine growth retardation), VLBW (very low birthweight), ELBW (extremely low birthweight), APH (antepartum haemorrhage), PPHN (persistent pulmonary hypertension of the newborn), RDS (respiratory distress syndrome), NEC (necrotising enterocolitis), PDA (patent ductus arteriosus), IVH (intraventricular haemorrhage), ICD-9 (International Classification of Diseases, Ninth Revision), CRIB (clinical risk index for babies) (326), BPD (bronchopulmonary dysplasia), ROP (retinopathy of prematurity)

11.3 Prisma checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	3.2
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	N/A
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	3.1.2.3
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	3.2
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	3.2.1.1
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	3.2.1.1-2
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	3.2.1.1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	3.2.1.1
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	3.2.1.1
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	3.2.1.1
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	3.2.1.1
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the	3.2.1.3

Section and Topic	Item #	Checklist item	Location where item is reported
		process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	3.2.2.2
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	3.2.2.2
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	N/A
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	3.2.3.2
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	3.2.1.2
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	3.2.3.3
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 8
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	3.2.2.1
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	3.2.2.1
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	3.2.2.2

Section and Topic	Item #	Checklist item	Location where item is reported
syntheses	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	3.2.3
	23b	Discuss any limitations of the evidence included in the review.	3.2.3.1-2
	23c	Discuss any limitations of the review processes used.	3.2.3.1-2
	23d	Discuss implications of the results for practice, policy, and future research.	3.2.4
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	N/A
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	1.1
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

12 Appendix II: clinical questionnaire data source

12.1 Clinical questionnaire (version 1)

Unit properties

1. **What is the level of your neonatal unit?**
    a. Level 2
    b. Level 3

2. **Does your unit differ from the current BAPM standards for level 2 / 3 units with regards to the gestational age of babies you admit, or the type of care you provide (e.g. cooling)?**
    a. No
    b. Yes:_____

3. **Does your hospital have a foetal medicine unit?**
    a. No
    b. Yes

4. **How far are you from the nearest NICU?**
    a. N/A (we are an NICU)
    b. ____ miles

5. **How far are you from the nearest hospital offering neonatal surgery?**
    a. On-site
    b. Off-site: ____ miles

Volume of patients

6. **Total number of births per annum?**
    a. _____

7. **Total number of neonatal admissions to NNU per annum?**
    a. _____

8. **Number of babies born between 23-27 weeks of gestation per annum (admitted or transferred out ex-utero)?**
    a. _____

9. **Total number of admissions of babies born between 27-31 weeks gestation per annum?**
    a. _____

10. Total number of admissions of babies born between 27-31 weeks gestation, who are extremely low birthweight (ELBW), per annum?

a. _____

11. Total number of admissions of babies born between 27-31 weeks gestation, who are small of gestational age (SGA), per annum?

a. _____

Staffing

12. Average number of nurses on shift per 24 hours?

a. _____

13. Average number of nurses with QIS (neonatology) on shift per 24 hours?

a. _____

14. Average number of consultants on the unit per 24 hours?

a. _____

15. Average number of consultants with CCT (neonatology) on shift per 24 hours?

a. _____

16. Average number of registrars on shift per 24 hours?

a. _____

17. Average number of registrars (neonatal GRID trainees) on shift per 24 hours?

a. _____

18. Is consultant on-site 24 hours a day, 7 days a week?

a. No

b. Yes

Obstetric

19. Does your maternity unit have a policy advocating corticosteroids for threatened preterm labour at 27-31 weeks gestation?

a. No
b. Only for threatened preterm labour <____ weeks gestation
c. Yes, for all threatened preterm labour at 27-31 weeks gestation
d. Other:_____

_____

20. Does your maternity unit have a policy advocating magnesium sulphate for threatened preterm labour at 27-31 weeks gestation?

a. No
b. Only for threatened preterm labour <____ weeks gestation
c. Yes, for all threatened preterm labour at 27-31 weeks gestation
d. Other:_____

_____

21. Does your maternity unit have a policy advocating delayed cord clamping for babies born between 27-31 weeks of gestation?

a. No
b. Only for >____ weeks gestational age
c. Yes, for all babies born between 27-31 weeks of gestation
d. Other:_____

_____

Respiratory

22. Does your NNU practice mandatory intubation and ventilation of babies born between 27-31 weeks of gestation?

a. No
b. Yes, for <____ weeks gestational age
c. Yes, for all babies born between 27-31 weeks of gestation
d. Other:_____

_____

23. Is surfactant given prophylactically to babies born between 27-31 weeks of gestation requiring intubation and ventilation?

a. No, unless there is evidence of surfactant deficient lung disease (SDLD/RDS)
b. Yes, for all <____ weeks gestational age
c. Yes, for all babies born between 27-31 weeks of gestation
d. Other:_____

_____

24. Is the ETT position checked by Xray prior to administration of surfactant?

a. No
b. Yes, in all cases
c. Yes, if <____ weeks gestational age

25. Is surfactant given prophylactically to babies born between 27-31 weeks of gestation requiring non-invasive ventilation (NIV)?

a. No, unless there is evidence of surfactant deficient lung disease (SDLD/RDS)
b. Yes, for all <____ weeks gestational age
c. Yes, for all babies born between 27-31 weeks of gestation
d. Other:_____

_____

26. Is a second dose of surfactant considered?

a. No, a second dose is never given
b. Yes, a second dose is given in babies with continuing evidence of surfactant deficient lung disease (SDLD/RDS)
c. Other:_____

_____

27. Does your NNU use minimally invasive methods for surfactant administration in non-ventilated babies born between 27-31 weeks of gestation (e.g. LISA)?

a. No
b. Yes

28. Does your NNU use high frequency oscillatory ventilation (HFOV) for babies born between 27-31 weeks of gestation?

a. No, HFOV is never used for babies born between 27-31 weeks of gestation
b. Yes, HFOV is used only as step up from conventional ventilation for babies born between 27-31 weeks of gestation
c. Yes, HFOV is used prophylactically (i.e. from birth for infants requiring ventilation) for all <____ weeks gestational age
d. Yes, HFOV is used prophylactically (i.e. from birth for infants requiring ventilation) for all babies born between 27-31 weeks of gestation
e. Other:_____

_____

29. What is your NNU's default mode of conventional ventilation (i.e. what ventilated neonates are routinely commenced on)?

a. Volume guaranteed
b. Pressure limited
c. Other:_____

_____

30. Does your NNU use inhaled nitric oxide (iNO) for babies born between 27-31 weeks of gestation?

a. No, never in babies born between 27-31 weeks of gestation
b. Yes, used for babies born between 27-31 weeks of gestation that are difficult to oxygenate (e.g. due to PPHN)
c. Other:_____

_____

31. Does your NNU use BiPAP for babies born between 27-31 weeks of gestation?

a. No
b. Yes

32. Does your NNU use both CPAP and/or High Flow for babies born between 27-31 weeks of gestation?

a. Only CPAP
b. Mainly CPAP, sometimes High Flow
c. Both CPAP and High Flow equally
d. Mainly High Flow, sometimes CPAP
e. Only High Flow

33. What oxygen saturations does your NNU target for babies born between 27-31 weeks of gestation?

a. ____% to ____%

34. Are ventilated babies born between 27-31 weeks of gestation routinely started on morphine?

a. No
b. Yes

35. Are ventilated babies born between 27-31 weeks of gestation routinely paralysed?

a. No
b. Yes

36. Does your NNU use diuretics for ventilated babies born between 27-31 weeks of gestation who are developing CLD?

a. No
b. Yes

37. Does your NNU use postnatal systemic corticosteroids in babies born between 27-31 weeks of gestation who are ventilator dependent?

a. No
b. Yes

Cardiovascular

38. What is your units first line treatment for babies born between 27-31 weeks of gestation with hypotension?

a. Dopamine
b. Dobutamine
c. Noradrenaline
d. Adrenaline
e. Hydrocortisone
f. Fluid bolus (____ml/kg)

39. What is your second line treatment (e.g. increasing dose, or starting another agent)?

a. _____

40. What is your third line treatment (e.g. increasing dose, or starting another agent)?

a. _____

41. What does your NNU use for medical management of PDA in babies born between 27-31 weeks of gestation?

a. N/A, we do not treat PDAs in any circumstance, whether suspected or confirmed
b. Ibuprofen
c. Paracetamol

42. Do your babies born between 27-31 weeks of gestation who fail medical management of their PDA get referred for surgical closure?

a. No, not in any circumstance
b. Yes, if clinically required

Gastrointestinal

43. How does your NNU treat NEC, Bell's stage 1 (mild clinical signs, e.g. increase in desaturations and bradycardias, abdominal distension, bilious aspirate; non-diagnostic AXR, e.g. thickened bowel loops; no laboratory changes) in babies born between 27-31 weeks of gestation?

a. Close monitoring, but continue current management (i.e. don't stop feeds or start ABs)
b. Antibiotics, for a minimum of ____ days
c. Feeds stopped for a minimum of ____ days

44. How does your NNU treat NEC, Bell's stage 2 (moderate clinical signs, e.g. abdominal tenderness; abnormal AXR, e.g. pneumotosis intestinalis; laboratory changes, e.g. metabolic acidosis, thrombocytopenia) in babies born between 27-31 weeks of gestation?

a. Close monitoring, but continue current management (i.e. don't stop feeds or start ABs)
b. Antibiotics, for a minimum of ____ days
c. Feeds stopped for a minimum of ____ days

45. Does your NNU use maternal EBM for mouth cares prior to starting trophic feeds in babies born between 27-31 weeks of gestation?

a. No
b. Yes, for all <____ GA neonates
c. Yes, for all babies born between 27-31 weeks of gestation

46. Does your NNU use donor EBM (DEBM) or preterm formula, in the absence of maternal EBM for babies born between 27-31 weeks of gestation?

a. Preterm formula for all babies born between 27-31 weeks of gestation
b. DEBM for all <____ GA neonates
c. DEBM for all babies born between 27-31 weeks of gestation

47. Does your NNU have a policy on increasing feeds (e.g. low/medium/high risk, fast/medium/slow)?

a. No
b. Yes

48. Does your NNU have a nutrition team (e.g. allocated consultant, dieticians, specialist nurses)?

a. No
b. Yes

49. How often do they do a ward round?

a. Daily
b. Every other day
c. Twice weekly
d. Weekly
e. Fortnightly
f. Monthly

50. Is parenteral nutrition (PN) used for babies born between 27-31 weeks of gestation while enteral feeds are established?

a. No, we use IV fluids for all babies born between 27-31 weeks of gestation
b. Yes, we use PN for all <____ weeks GA neonates (above this gestation, IV fluids are used)
c. Yes, we use PN for all babies born between 27-31 weeks of gestation

51. If PN is used, on average, how soon after birth is it commenced?

a. Day of birth
b. After 24 hours
c. After 48 hours

52. Does your NNU have any of the following for mothers of babies born between 27-31 weeks of gestation:

- a breastfeeding support team?

- an educational programme (verbal or written) regarding benefits of breastfeeding?

- a peer-to-peer support programme?

a. No

b. Yes

53. Does your NNU use prebiotics?

a. No

b. Yes, if <____g birthweight

c. Yes, if growth <____ centile

d. Yes, if <____ weeks GA (regardless of weight or growth)

e. Yes, for all babies born between 27-31 weeks of gestation (regardless of weight or growth)

54. Does your NNU use probiotics?

a. No

b. Yes, if <____g birthweight

c. Yes, if growth <____ centile

d. Yes, if <____ weeks GA (regardless of weight or growth)

e. Yes, for all babies born between 27-31 weeks of gestation (regardless of weight or growth)

Infection control

55. What infection control procedures are in place for central line insertion?

a. Sterile gloves and field

b. Sterile gown, gloves, and field

c. Matching Michigan

d. Aseptic non-touch technique

e. Other:_____

_____

56. What are the minimum number of operators involved in line insertion?

a. 1

b. 2

57. Is an insertion checklist used (e.g. WHO)?

a. No
b. Yes

58. What infection control procedures are in place for accessing the central line?

a. Non-sterile gloves with/without gown
b. Sterile gloves and gown
c. Matching Michigan
d. Aseptic non-touch technique
e. Other:_____

_____

59. What are the minimum number of operators involved in accessing the line?

a. 1
b. 2

60. On average, how soon after a central line is no longer required is it removed (e.g. following establishment of full enteral feeds)?

a. Immediately
b. Within 24 hours
c. Within 48 hours

61. What does your NNU use as first line antibiotics?

a. _____

62. What does your NNU use as second line antibiotics?

a. _____

63. What does your NNU use as third line antibiotics?

a. _____

Neurodevelopment

64. Does your NNU have trained developmental care nurses?

a. No
b. Yes

65. How often do they see the babies?

a. Daily
b. Every other day
c. Twice weekly
d. Weekly
e. Fortnightly
f. Monthly

66. Are parents able to provide skin-to-skin / Kangaroo care for babies born between 27-31 weeks of gestation who are ventilated, but stable?

a. No
b. Yes, if >____ weeks GA
c. Yes, for all babies born between 27-31 weeks of gestation

67. Are blood tests and other interventions timed to occur with cares?

a. Less than a quarter of the time (<25%)
b. A quarter to half of the time (25-50%)
c. Half to three quarters of the time (50-75%)
d. Three quarters to all of the time (75-100%)

Blood products

68. Does your NNU give albumin transfusions?

a. No
b. Yes

Metabolic

69. Does your NNU treat hyperglycaemia with insulin?

a. No
b. Yes

Discharge from hospital

70. Does your NNU discharge babies on tube feeds?

a. No
b. Yes

71. Does your NNU discharge babies on home oxygen?

a. No
b. Yes

72. Does your NNU have a post-discharge support service (e.g. specialist nurses who do routine home visits)?

a. No
b. Yes

73. Is there provision for post-discharge physio care?

a. No
b. Arranged for all ex-27-31 week babies pre-discharge
c. Arranged for babies who are at high risk of developing CP (e.g. lower GA, ELBW, IVH/PVL) pre-discharge
d. Arranged for babies who require it when seen in OP clinic
e. Other:_____
_____

74. Is there provision for post-discharge cognitive development?

a. No
b. Arranged for all ex-27-31 week babies pre-discharge
c. Arranged for babies who are at high risk of developing cognitive impairment (e.g. lower GA, ELBW, IVH/PVL) pre-discharge
d. Arranged for babies who require it when seen in OP clinic
e. Other:_____
_____

12.2 Questions removed or refined during development of questionnaire

Questions simplified or removed due to delving too deep into minutiae of clinical practice	**Questions removed due to interrogating aspects of structure or process not linked to outcomes**	**Questions removed due to not being suitable for a questionnaire (i.e. respondent would not be expected to be able to provide the information requested)**	**Questions removed due to relating to neurodevelopmental outcomes**	**Questions removed due to data availability from the Neonatal Data Analysis Unit (NDAU)**
• Does your unit differ from the current BAPM standards for level 2 / 3 units …? • What oxygen saturations does your NNU target …? • What is your units first line treatment for babies … with hypotension? • What is the second line treatment (for hypotension) …? • What is the third line treatment (for hypotension) …? • How does your NNU treat NEC, Bell's stage 1 (…) in babies born between 27-31 weeks of gestation? • How does your NNU treat NEC, Bell's stage 2 (…) in babies born between 27-31 weeks of gestation? • What are the minimum number of operators involved in (central) line insertion? • Is an (central line) insertion checklist used (…)?	• Does your hospital have a foetal medicine unit? • How far are you from the nearest NICU? • How far are you from the nearest hospital offering neonatal surgery? • Is consultant on-site 24 hours a day, 7 days a week? • Is the ETT position checked by Xray prior to administration of surfactant? • Is surfactant given prophylactically to babies … requiring non-invasive ventilation (NIV)? • What does your NNU use for medical management of PDA …? • Do your babies … who fail medical management of their PDA get referred for surgical closure? • Does your NNU use maternal EBM for mouth cares prior to starting trophic feeds …? • Does your NNU use donor EBM (DEBM) or preterm	• Total number of births per annum? Total number of neonatal admissions to NNU per annum? • Number of babies born between 23-27 weeks of gestation per annum (admitted or transferred out ex-utero)? • Total number of admissions of babies born between 27-31 weeks gestation per annum? • Total number of admissions of babies born between 27-31 weeks gestation, who are extremely low birthweight (ELBW), per annum? • Total number of admissions of babies born between 27-31 weeks gestation, who are small of gestational age (SGA), per annum? • Average number of nurses on shift per 24 hours? • Average number of nurses with QIS (neonatology) on shift per 24 hours?	• Does your NNU have trained developmental care nurses? • How often do they see the babies? • Are blood tests and other interventions timed to occur with cares? • Is there provision for post-discharge physio care? • Is there provision for post-discharge cognitive development?	• Does your maternity unit have a policy advocating corticosteroids for threatened preterm labour at 27-31 weeks gestation? • Does your maternity unit have a policy advocating magnesium sulphate for threatened preterm labour at 27-31 weeks gestation? • Is a second dose of surfactant considered? • Does your NNU use BiPAP …? • Does your NNU use inhaled nitric oxide (iNO) …? • Does your NNU discharge babies on tube feeds? • Does your NNU discharge babies on home oxygen?

• What are the minimum number of operators involved in accessing the (central) line? • On average, how soon after a central line is no longer required is it removed (…)? • What does your NNU use as first line antibiotics? • What does your NNU use as second line antibiotics? • What does your NNU use as third line antibiotics?	formula, in the absence of maternal EBM …? • How often do they (the nutrition team) do a ward round? • Is parenteral nutrition (PN) used for babies … while enteral feeds are established? • If PN is used, on average, how soon after birth is it commenced? • Does your NNU use prebiotics? • Does your NNU give albumin transfusions? • Does your NNU treat hyperglycaemia with insulin?	• Average number of consultants on the unit per 24 hours? • Average number of consultants with CCT (neonatology) on shift per 24 hours? • Average number of registrars on shift per 24 hours? • Average number of registrars (neonatal GRID trainees) on shift per 24 hours?		

Table 49 Questions removed or refined during development of questionnaire

12.3 Clinical questionnaire (version used for pilot)

Unit properties

1. **What is the designation of your neonatal unit?**
    a. Local neonatal unit (LNU)
    b. Neonatal intensive care unit (NICU)

2. **[For LNUs] What is your criteria for transfer based on:**
    a. Gestational age:
        i. We transfer out all neonates < ____ weeks GA
    b. Duration of intensive care:
        i. We transfer out all neonates requiring >____ hours of intensive care

3. **What is your NNUs method of transferring a neonate to another unit?**
    a. Dedicated transport service
    b. Arranged using in-house staff

Obstetrics

4. **Does your maternity unit have a policy advocating delayed cord clamping for 27-31 week gestational age (GA) neonates?**
    a. No
    b. Only for >____ weeks gestational age
    c. Yes, for all babies born between 27-31 weeks of gestation

Respiratory

5. **Does your NNU practice mandatory intubation and ventilation of babies born between 27-31 weeks of gestation?**
    a. No
    b. Yes, for <____ weeks gestational age
    c. Yes, for all babies born between 27-31 weeks of gestation

6. **Does your NNU practice early non-invasive respiratory support (e.g. CPAP from birth) for babies born between 27-31 weeks of gestation who do not require intubation as part of their resuscitation?**
    a. No
    b. Yes, for <____ weeks gestational age
    c. Yes, for all babies born between 27-31 weeks of gestation

7. **Is surfactant given prophylactically to babies born between 27-31 weeks of gestation requiring intubation and ventilation?**
    a. No, unless there is evidence of surfactant deficient lung disease (SDLD/RDS)
    b. Yes, for all <____ weeks gestational age
    c. Yes, for all babies born between 27-31 weeks of gestation

8. **Does your NNU use minimally invasive methods of surfactant administration (i.e. LISA) in babies born between 27-31 weeks of gestation who require it (compared to intubation and ventilation, i.e. INSURE)?**
    a. No
    b. Yes

9. **What is your NNU's default mode of conventional ventilation for babies born between 27-31 weeks of gestation (i.e. what ventilated neonates are routinely commenced on)?**
    a. Volume guaranteed
    b. Pressure limited

10. **Does your NNU use high frequency oscillatory ventilation (HFOV) for babies born between 27-31 weeks of gestation?**
    a. No, HFOV is never used for babies born between 27-31 weeks of gestation
    b. Yes, HFOV is used only as step up from conventional ventilation for babies born between 27-31 weeks of gestation
    c. Yes, HFOV is used prophylactically (i.e. from birth for infants requiring ventilation) for all <____ weeks gestational age
    d. Yes, HFOV is used prophylactically (i.e. from birth for infants requiring ventilation) for all babies born between 27-31 weeks of gestation

11. **Post-extubation, does your NNU use CPAP and/or High Flow for babies born between 27-31 weeks of gestation?**
    a. Generally, only CPAP
    b. Both CPAP and High Flow
    c. Generally, only High Flow

12. **Are ventilated babies born between 27-31 weeks of gestation routinely sedated (e.g. morphine)?**
    a. No
    b. Yes

13. **Are ventilated babies born between 27-31 weeks of gestation routinely commenced on therapeutic muscle relaxation (e.g. atracurium, rocuronium)?**
    a. No
    b. Yes

14. **Does your NNU use diuretics for ventilated babies born between 27-31 weeks of gestation who are developing, or have established CLD?**
    a. Infrequently / never
    b. Frequently / always

15. **Does your NNU use postnatal systemic corticosteroids in babies born between 27-31 weeks of gestation who are ventilator dependent?**
    a. Infrequently / never
    b. Frequently / always

Cardiovascular

16. **Does your NNU use >1 10ml/kg fluid bolus in babies born between 27-31 weeks of gestation for treatment of hypotension?**
    a. Infrequently / never
    b. Frequently / always

17. **Does your NNU use sodium bicarbonate in babies born between 27-31 weeks of gestation for correction of metabolic acidosis?**
    a. Infrequently / never
    b. Frequently / always

Gastrointestinal

18. **Are feeds for babies born between 27-31 weeks of gestation stopped during blood transfusions?**
    a. No
    a. Yes, for all <____ GA neonates
    b. Yes, for all babies born between 27-31 weeks of gestation

19. **Does your NNU have a guideline on increasing feeds (e.g. low/medium/high-risk; fast/medium/slow)?**
    a. No
    b. Yes

20. Does your NNU have a nutrition team (e.g. allocated consultant, dieticians, specialist nurses, use of paediatric services)?

a. No
b. Yes

21. Does your NNU have any of the following for mothers of babies born between 27-31 weeks of gestation:

- a breastfeeding support team?

- an educational programme (verbal or written) regarding benefits of breastfeeding?

- a peer-to-peer support programme?

a. No
b. Yes

22. Does your NNU use probiotics for babies born between 27-31 weeks of gestation?

a. No
b. Yes, if <____ weeks GA
c. Yes, for all babies born between 27-31 weeks of gestation

Infection control

23. What infection control procedures are in place for central line insertion?

a. Sterile gown, gloves, and field
b. Maximal standard barriers (i.e. cap, mask, sterile gloves, gown and field; e.g. Matching Michigan, aseptic non-touch technique)

24. What infection control procedures are in place for accessing the central line?

a. Sterile gloves and gown
b. Maximal standard barriers (i.e. cap, mask, sterile gloves, gown and field; e.g. Matching Michigan, aseptic non-touch technique)

25. Is there a daily assessment (e.g. on the ward round) of whether a central line in situ is still required?

a. Infrequently / never
b. Frequently / always

General

26. Are parents able to provide skin-to-skin / Kangaroo care for babies born between 27-31 weeks of gestation who are ventilated, but stable?

a. No
b. Yes, if >____ weeks GA
c. Yes, for all babies born between 27-31 weeks of gestation

Discharge from hospital

27. Does your NNU have a post-discharge support service (e.g. specialist nurses who do routine home visits)?

a. No
b. Yes

Unit ethos

28. What is the 'philosophy' / 'ethos' of your unit (i.e. what are the priorities for your NNU, what do you pride yourselves on, what do you believe you do better than other NNUs)? This may include:

a. Identification of areas of clinical practice or patient outcomes which require improvement
b. Use of care bundles / packages of care / combination of measures to tackle specific issues (e.g. golden hour care, CLABSI, NEC)
c. Encouragement of healthcare professionals (other than consultant body, e.g. junior doctors, nurses, pharmacists, dieticians, physiotherapists, etc.) to investigate and introduce change (e.g. quality improvement projects)
d. Risk assessment strategies (e.g. regular meetings between obstetricians and neonatologists regarding high-risk deliveries, or neonatal medical and nursing teams regarding staffing issues, parental concerns or to raise concerns regarding clinical management)
e. Regular teaching sessions using high/low-fidelity simulation

12.4 Email invitation to NNU for participating in questionnaire pilot

Dear colleague,

My name is Dr Abdul Qader Ismail. I am a clinical fellow working on my PhD, on workstream 2 of the OptiPrem study (http://www.royalwolverhampton.nhs.uk/research-and-development/OptiPrem-improving-neonatal-service-delivery/). My co-supervisors are Dr Tilly Pillay (Chief investigator) and Dr Elaine Boyle.

The aim of the study is to improve care for preterm babies born between 27-31 weeks of gestation. Part of this involves determining if the designation of unit they are cared for in makes a difference.

However, even though both LNUs and NICUs care for these babies, this 'divide' may be artificial as there is huge variation in how units are set up and how they manage their patients which is independent of designation. It may be that differences in the care provided by neonatal units affect outcomes for these babies. Therefore, work stream 2 in the study investigates whether differences in care not just between LNUs and NICUs, but also within LNUs, and NICUs (i.e. between all neonatal units which care for these babies), matter. This will possibly allow a statistical analysis on outcomes based on clustering of types of care provided, rather than by designation.

As part of my PhD I need to determine whether it is possible to tease out differences in clinical care (or care packages) provided, before I can investigate whether these impact on outcomes for this group of babies. In exploring how best to achieve this, I have begun by devising a clinical questionnaire to assess care provided in key areas where variation may exist. The questionnaire is attached.

I understand that one of the chief limitations of the questionnaire is whether units will have time to complete it, and so have embarked on a pilot of ten neonatal units to assess feasibility. May I kindly ask permission to use your team to pilot this concept please.

If yes, all that is needed is a contact email and telephone number for a relatively senior individual on your unit with whom I will be able to engage, to complete the questionnaire, and who could provide me with relevant copies of your unit guidelines/protocols.

If the pilot proves unsuccessful, I will move on to further methods of evaluation. I will share the results of the pilot with you, after they have been reviewed and validated for conclusion by my supervisors, the OptiPrem team, and the Study Steering Committee. For all questions regarding the full OptiPrem Project please contact chief investigator, and my co-supervisor Dr Pillay on 07791936718 or tilly.pillay@nhs.net

Kind regards,
Dr Abdul Qader Ismail

13 Appendix III: Assessment of variables using criteria for evaluating MQC

MQC judging criteria	Non-NNAP MQC						NNAP audit measures
	Receipt of antenatal steroids	**Delayed cord clamping**	**Normal admission temp**	**Receipt of breastmilk on day 1 of life**	**Ratio of babies given NIV on day 1 out of all requiring ventilatory support**	**Ratio of babies requiring IC provided with 1:1 nursing care on day 1**	
Evidence of effect on outcomes	Affects mortality, incidence of IVH, NEC (238)	Affects mortality (244)	Affects mortality (254, 255)	Affects NEC (123)	Affects CLD, CLD and mortality (262)	Affects mortality and infection (168, 169)	There is evidence for some audit measures (e.g., antenatal steroids, admission temperature), but not others (e.g., receipt of breastmilk on discharge, timing of ROP screening). However, I was not planning on using them individually, but to create a composite measure for quality of care based on meeting of audit standards and data completion. This would form a surrogate marker for the organisational culture of units, which can affect mortality (172)
Relevance for patients/healthcare system	Neonatal care is of importance to families of individual babies, NHS in the short term due to high cost of providing neonatal care, and long-term regarding costs associated with neurodisability (47, 327-330)						
Care providers can influence area of care	Some of variables are nearly identical to NNAP audit measures, the rest are similar processes of care. Serial annual NNAP report details how units have worked to improve their adherence with audit measures (175).						Serial annual NNAP report details how units have worked to improve their adherence with audit measures (175)
Reliability and validity	Assessment of data completion and accuracy of data submitted to NDAU in recent study (231) by comparing with data from Office of National Statistics (ONS) and the PiPs study (232). By 2012 100% of neonatal units in England were contributing data to the NNRD, and for >98% of their babies born between 25-33 weeks of gestation. For babies born <32 weeks of gestation (2012-2015), percentage missing data for gestational age, sex, birthweight, antenatal steroids, mode of delivery, multiple birth and survival to discharge, was <8%. Using the PiPs data as the gold standard (i.e., assuming 100% accuracy), the degree of discordancy was assessed, with major discordancy defined as difference in binary items (e.g., whether gastrointestinal perforation occurred), or +/- ≥5 days difference for a continuous variable (e.g., antibiotic course duration). This was found for expected date of delivery, 5-minute Apgar, maternal ethnicity and LSOA (Lower Layer Super Output Area – small geographic areas linked to the Index of Multiple Deprivations 2010), mode of delivery, duration of high dependency care, central venous line in situ, antibiotic therapy, type of milk given on first day of milk feed, and receipt of supplementary oxygen on reaching 36 weeks postmenstrual age. Therefore, discordancy rates were <5% for 13/16 patient characteristics, 9/16 processes of care, and 10/11 outcomes. Considering outcomes, sensitivity of NNRD data was 50-100% and specificity was 86-100%.						

	NDAU data variable used is similar to NNAP audit measure (any dose of steroid given before birth). This does not differentiate between units giving a single dose within hours of birth, or those managing to administer a complete course (evidence indicates benefit if given 24 hours pre-birth).	Some units may practice stripping of blood from cord, especially for more preterm babies (instead of delayed cord clamping), which also has some evidence supporting it.	NDAU data variable used is similar to NNAP audit measure (temperature measured within 1 hour of birth). Therefore, possible for babies to be hypothermic on admission, warmed up and subsequent normal temperature taken within the first hour to be the one recorded.				
	Degree of data completion would be assessed from data provided by NDAU, if any variable had above a prespecified level of missing data (10%), it would not be used						Using degree of data completion as variable for measuring quality of care
Appropriate risk adjustment	Not using outcome measures so risk adjustment is of less relevance with processes of care and aspects of structure. However, some degree of risk adjustment would be required to account for variation in outcomes due to differences in socioeconomic background and illness severity of patient populations between units.						
Unintended consequences/risk	none	none	none	none	none	Units may provide sick babies with 1:1 nursing care with the rest of the staff being overstretched providing sub-optimal care to babies not requiring intensive care	No previous analysis for associations between adherence with NNAP audit measures and neonatal outcomes. If no associations with outcome are found, this could dissuade units from complying with NNAP audit measures.
Feasibility	Data collection already in place, incorporated into recording of daily patient care. Data is collected nationally by NDAU, cleaned and anonymised for use in research projects.						

Table 50 Assessment of variables using criteria for evaluating MQC

14 Appendix IV: list of major congenital anomalies

atresia and stenosis of small intestine

Atresia of bile ducts

Atresia of oesophagus with tracheo-oesophageal fistula

Atresia of oesophagus without fistula

Atresia of urethra

Atrioventricular septal defect (AVSD)

Coarctation of aorta

Coarctation of the aorta

Congenital absence, atresia, and stricture of auditory canal (external)

Congenital absence, atresia/stenosis of anus with/without fistula

Congenital absence, atresia/stenosis of rectum with/without fistula

Congenital cardiac disease - acyanotic

Congenital cardiac disease - non-cyanotic

Congenital malformations of aortic and mitral valves

Congenital malformations of cardiac chambers and connections

Congenital malformations of pulmonary and tricuspid valves

Down syndrome (translocation)

Down syndrome (Trisomy 21)

Down's syndrome

Edwards syndrome (Trisomy 18)

Encephalocele

Encephalocele (unknown or unspecified cause)

Eventration of diaphragmatic hernia

Eventration of the diaphragm

Exomphalos

Exomphalos (major)

Exomphalos (minor)

Exomphalos Malrotation

Exotrophy of urinary bladder

Gastroschisis

Hypoplasia of aortic arch

Malformation of aorta

Oesophageal atresia

Oesophageal atresia with distal tracheal fistula

Oesophageal atresia with tracheoesophageal fistula

Oesophageal atresia without distal fistula

Other congenital malformations of aortic arch

Polycystic kidney

Polycystic kidney (unknown or unspecified cause)

Potter's syndrome

Spina bifida

Spina bifida (unspecified)

Stenosis of aorta (AS)

Stenosis of pulmonary artery (PS)

Tetralogy of Fallot

Total anomalous pulmonary venous connection

Total anomalous pulmonary venous drainage

Transposition great arteries (TGA)

Trisomy 18

Trisomy 21

15 Appendix V: graphs comparing patient populations when categorising units based on non-NNAP measures of quality of care

15.1 Measure 1 - receipt of any dose of antenatal steroids

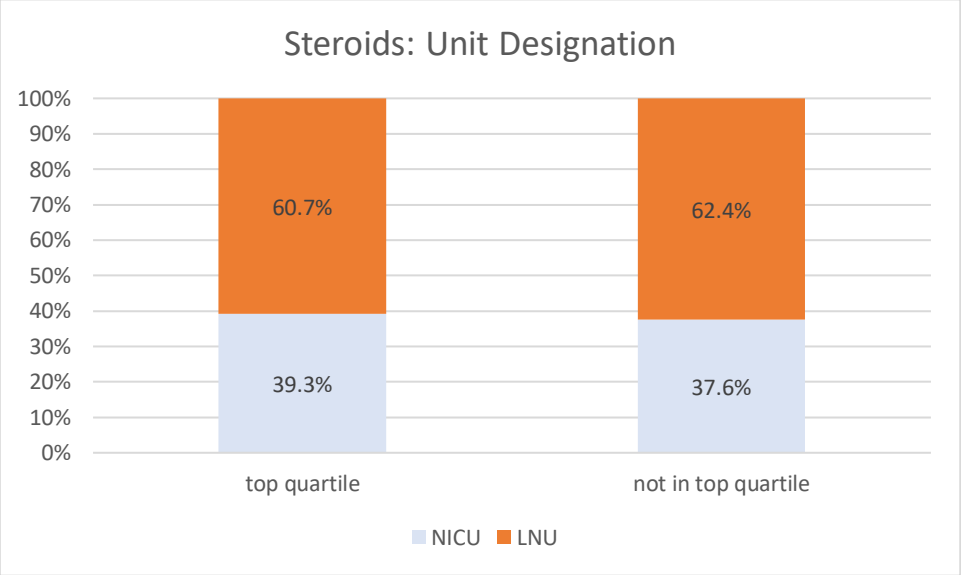


Figure 67 Proportion of babies born in NICU versus LNU when categorising units according to Measure 1 of my non-NNAP MQC

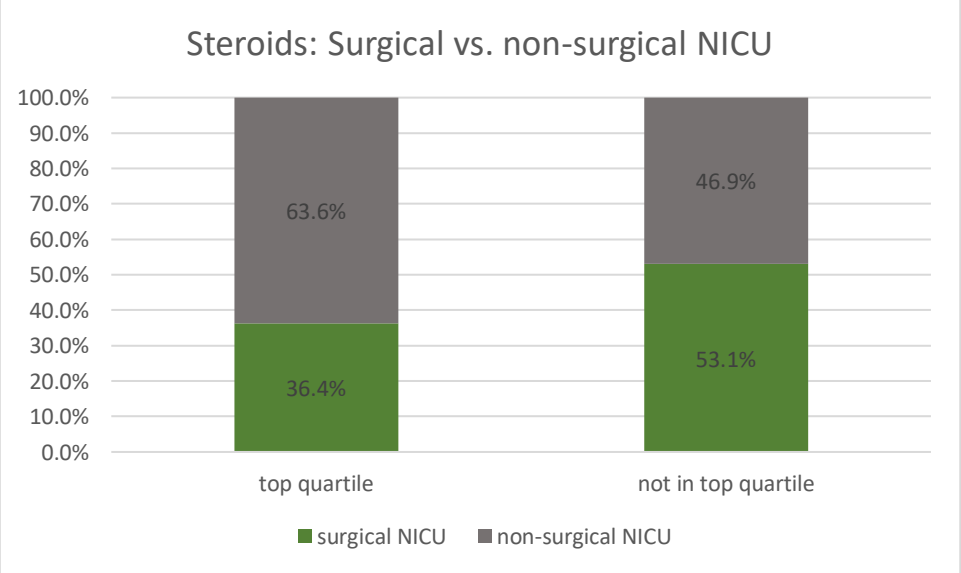


Figure 68 Proportion of babies born in surgical versus non-surgical NICU when categorising units according to Measure 1 of my non-NNAP MQC

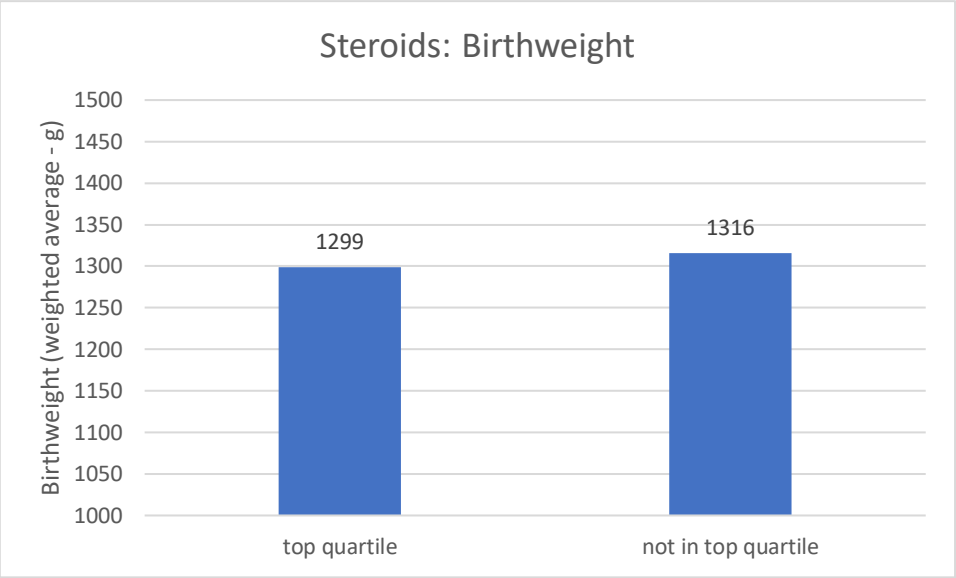


Figure 69 Mean birthweight (weighted average) of babies born in units categorised according to Measure 1 of my non-NNAP MQC

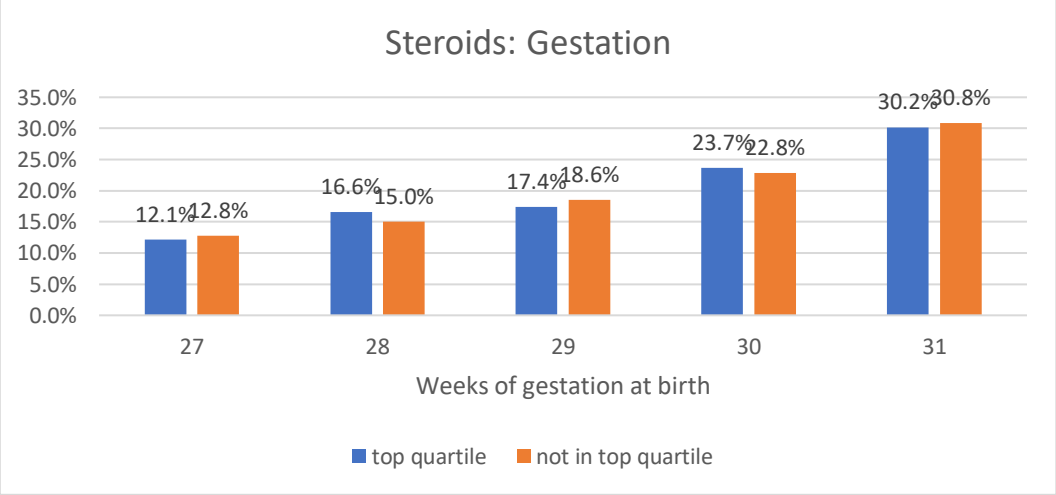


Figure 70 Proportion of babies born at each gestational week when categorising units according to Measure 1 of my non-NNAP MQC

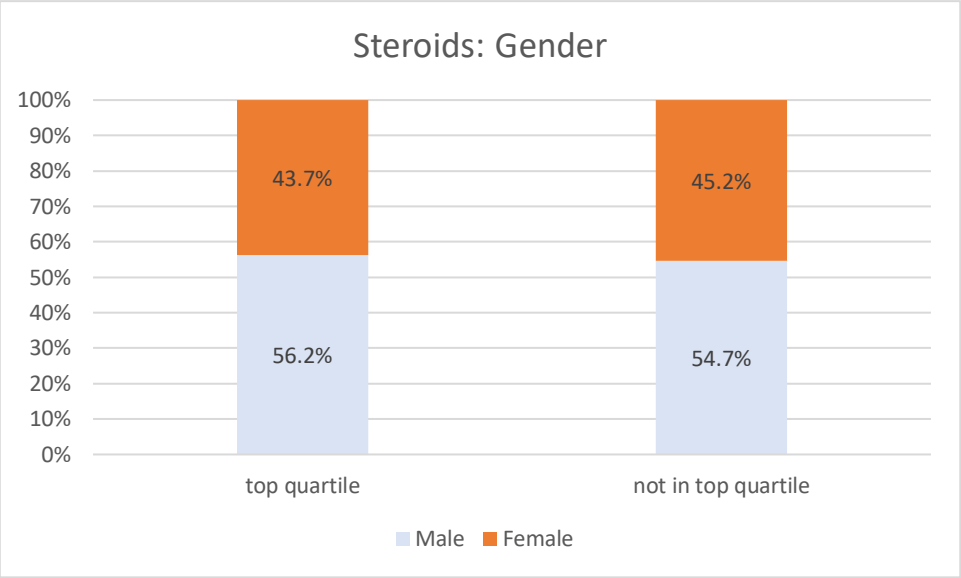


Figure 71 Proportion of male versus female babies when categorising units according to Measure 1 of my non-NNAP MQC

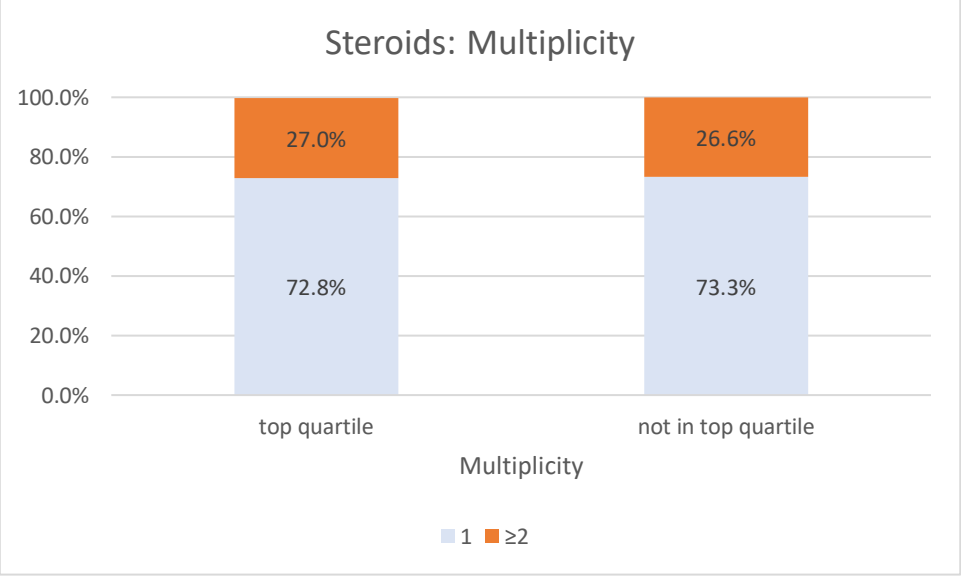


Figure 72 Proportion of singletons versus pregnancies with multiple foetuses when categorising units according to Measure 1 of my non-NNAP MQC

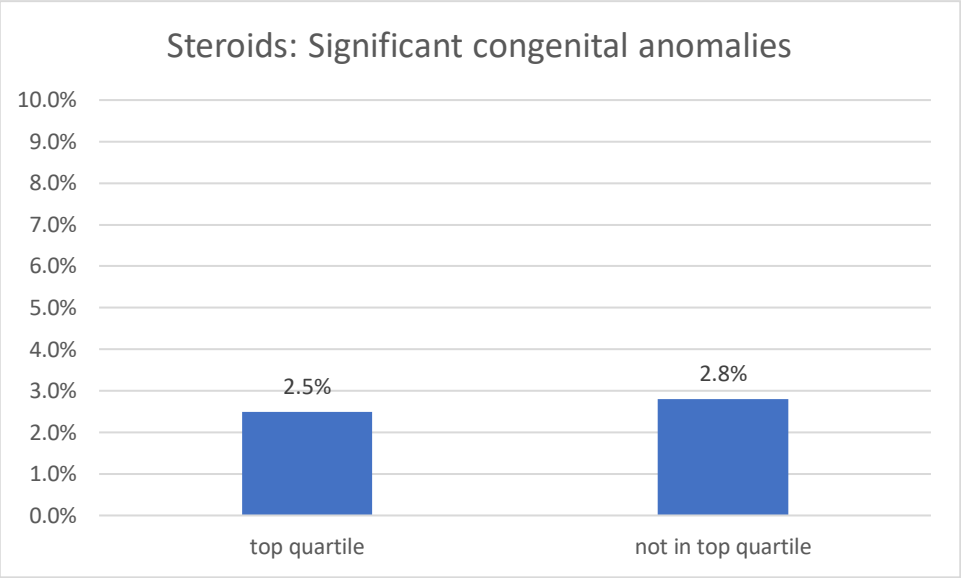


Figure 73 Proportion of babies with significant congenital anomalies when categorising units according to Measure 1 of my non-NNAP MQC

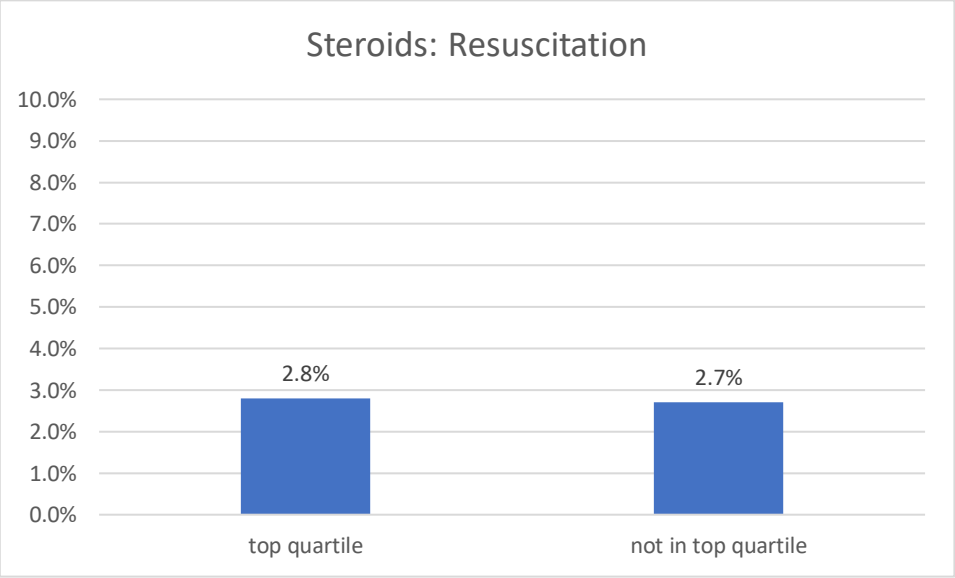


Figure 74 Proportion of babies requiring significant resuscitation when categorising units according to Measure 1 of my non-NNAP MQC

15.2 Measure 2 - normal temperature recorded within one hour of admission

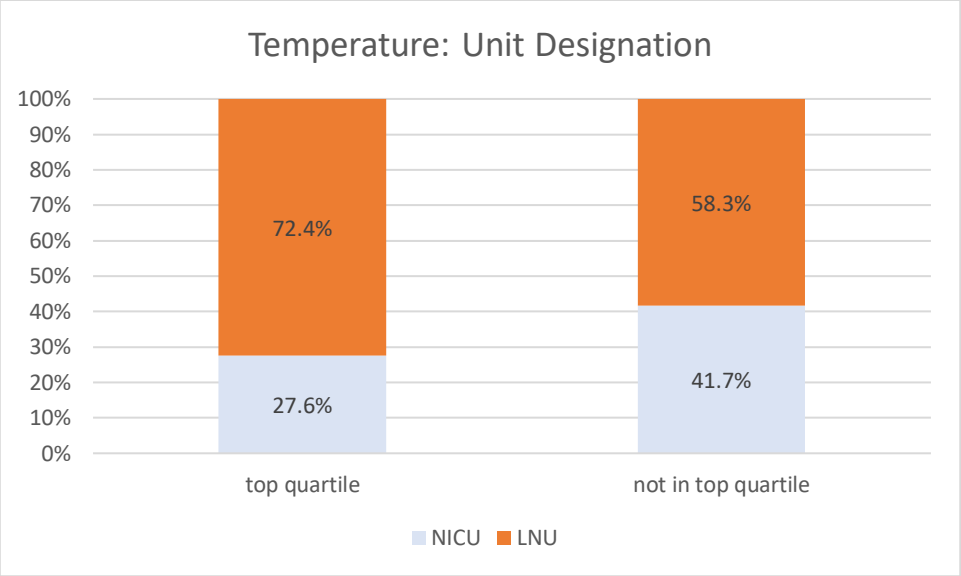


Figure 75 Proportion of babies born in NICU versus LNU when categorising units according to Measure 2 of my non-NNAP MQC

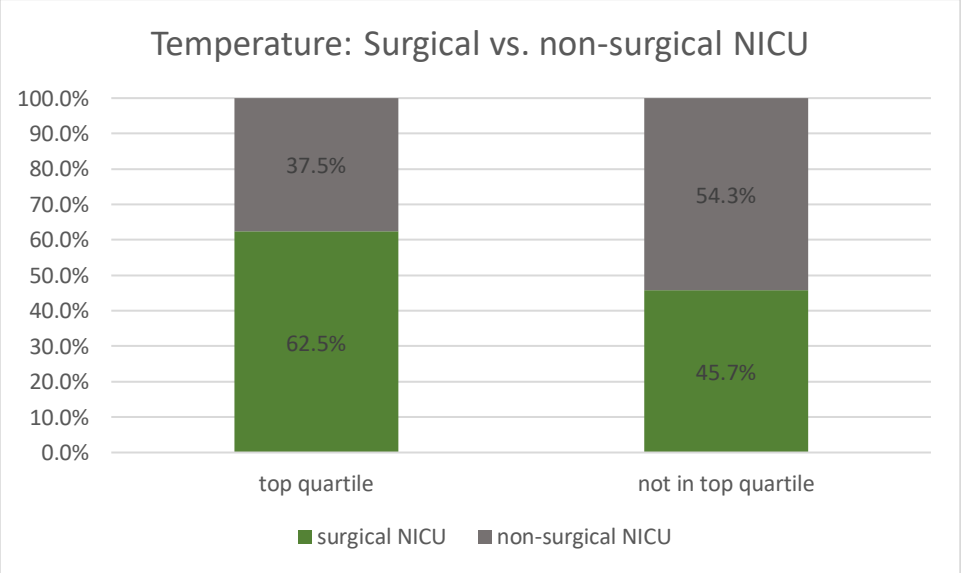


Figure 76 Proportion of babies born in surgical versus non-surgical NICU when categorising units according to Measure 2 of my non-NNAP MQC

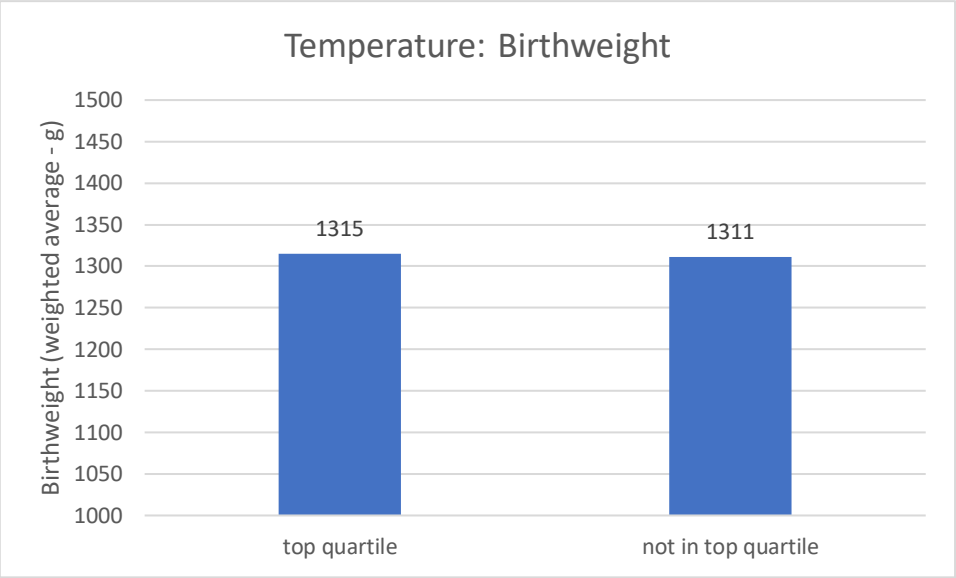


Figure 77 Mean birthweight (weighted average) of babies born in units categorised according to Measure 2 of my non-NNAP MQC

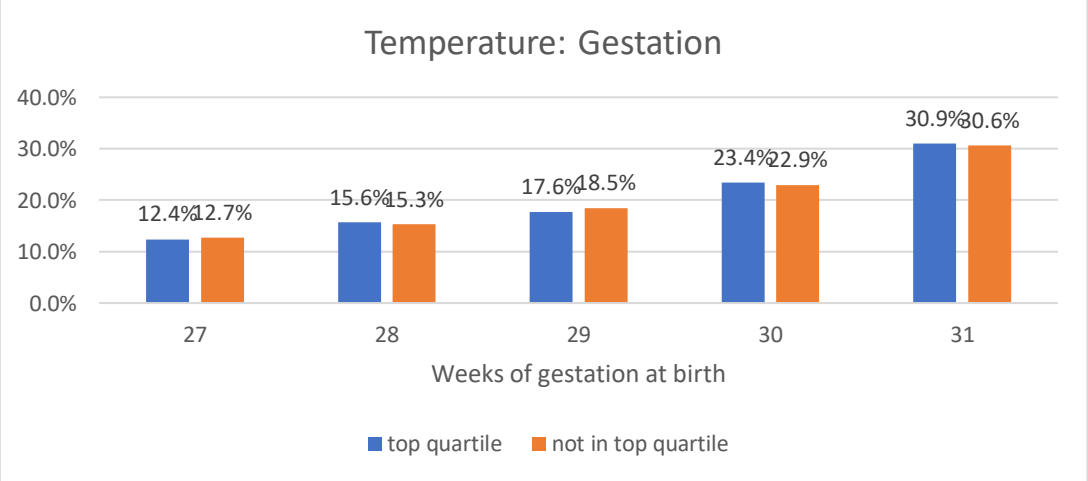


Figure 78 Proportion of babies born at each gestational week when categorising units according to Measure 2 of my non-NNAP MQC

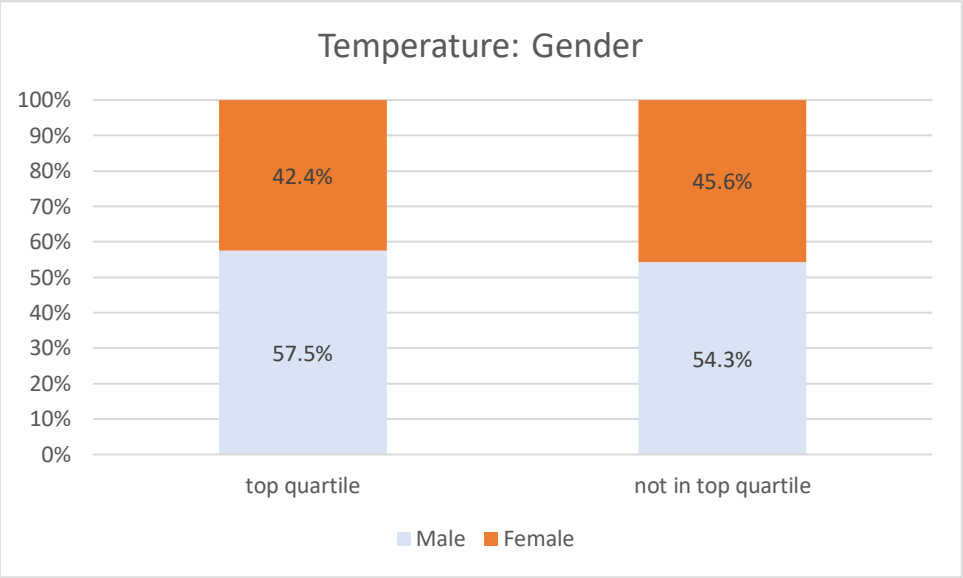


Figure 79 Proportion of male versus female babies when categorising units according to Measure 2 of my non-NNAP MQC

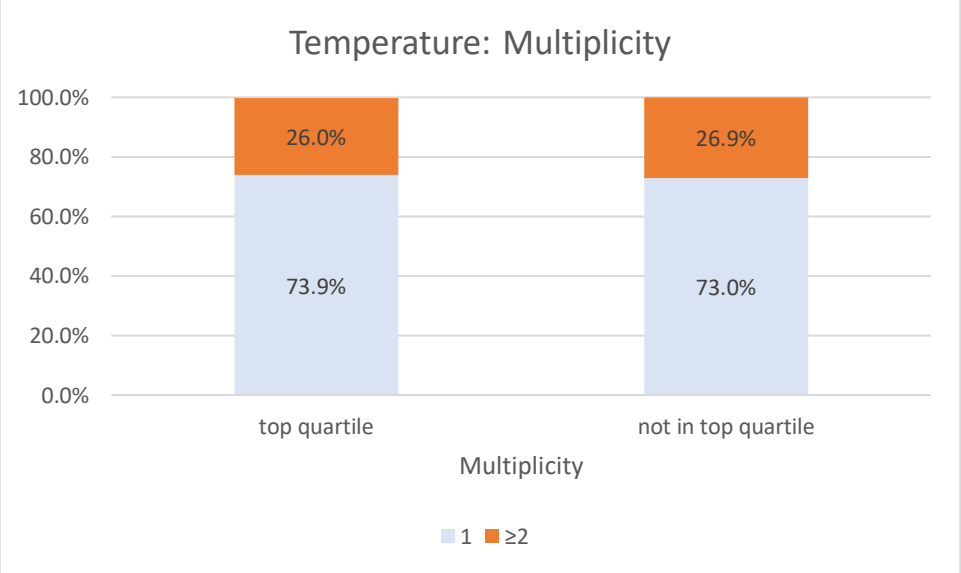


Figure 80 Proportion of singletons versus pregnancies with multiple foetuses when categorising units according to Measure 2 of my non-NNAP MQC

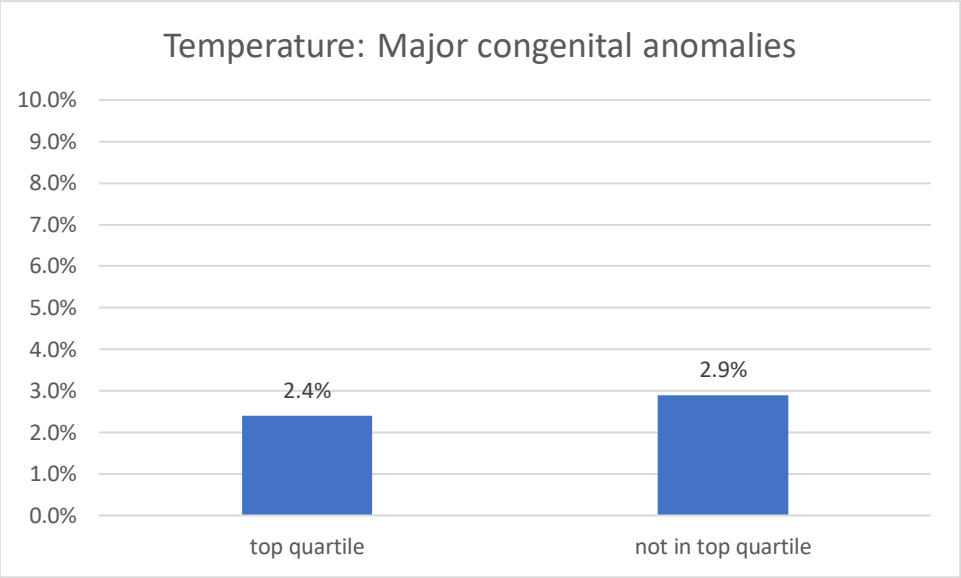


Figure 81 Proportion of babies with significant congenital anomalies when categorising units according to Measure 2 of my non-NNAP MQC

15.3 Measure 3 - babies requiring ventilatory support on day one of life supported with non-invasive ventilation (NIV)

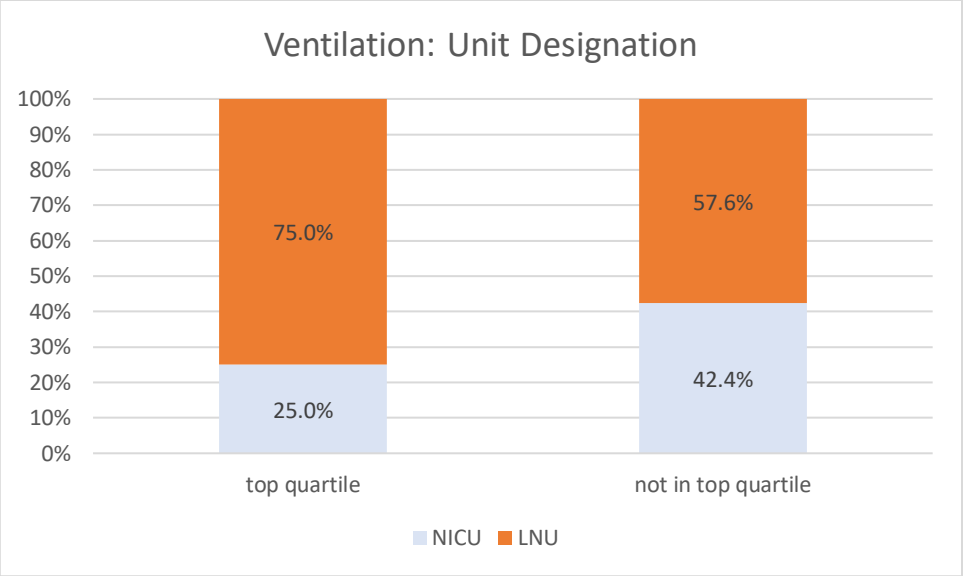


Figure 82 Proportion of babies born in NICU versus LNU when categorising units according to Measure 3 of my non-NNAP MQC

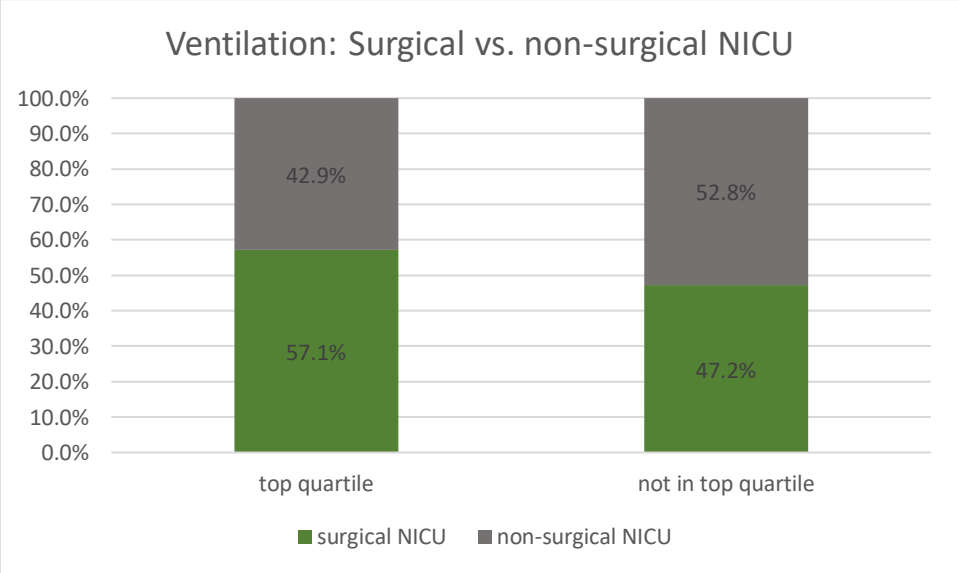


Figure 83 Proportion of babies born in surgical versus non-surgical NICU when categorising units according to Measure 3 of my non-NNAP MQC

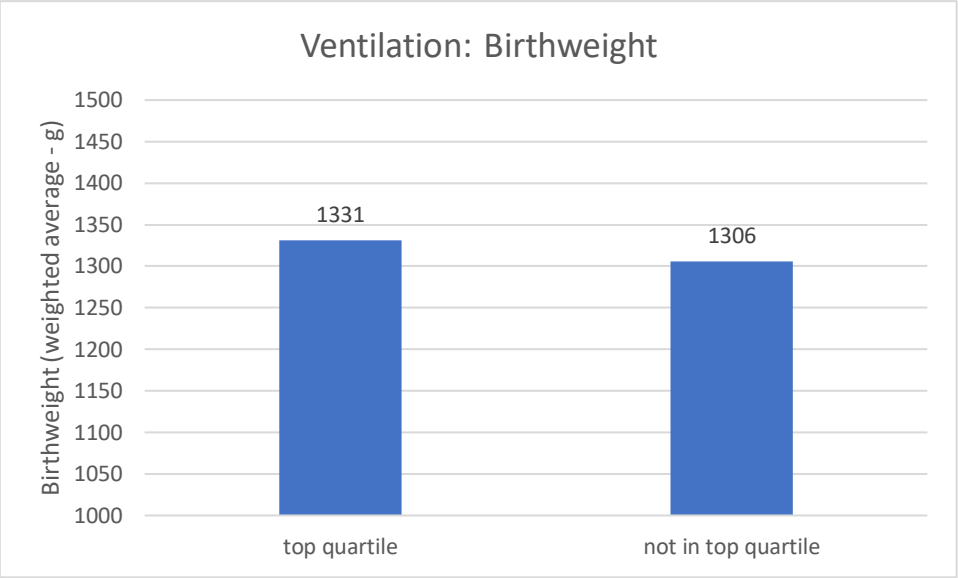


Figure 84 Mean birthweight (weighted average) of babies born in units categorised according to Measure 3 of my non-NNAP MQC

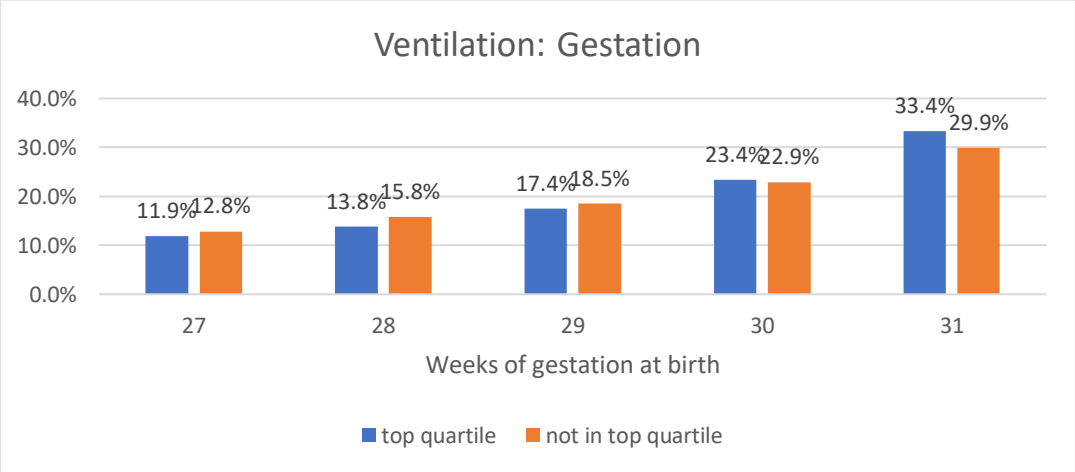


Figure 85 Proportion of babies born at each gestational week when categorising units according to Measure 3 of my non-NNAP MQC

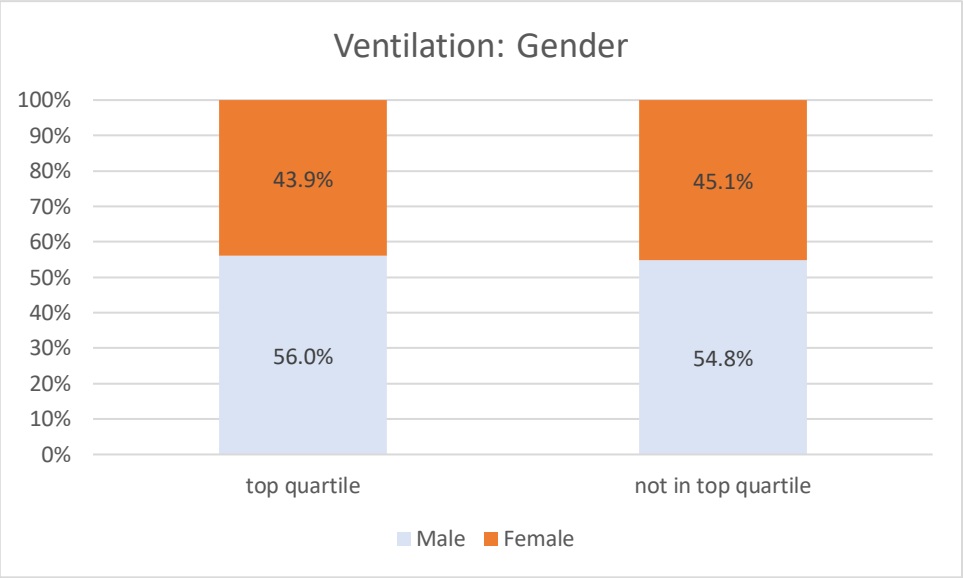


Figure 86 Proportion of male versus female babies when categorising units according to Measure 3 of my non-NNAP MQC

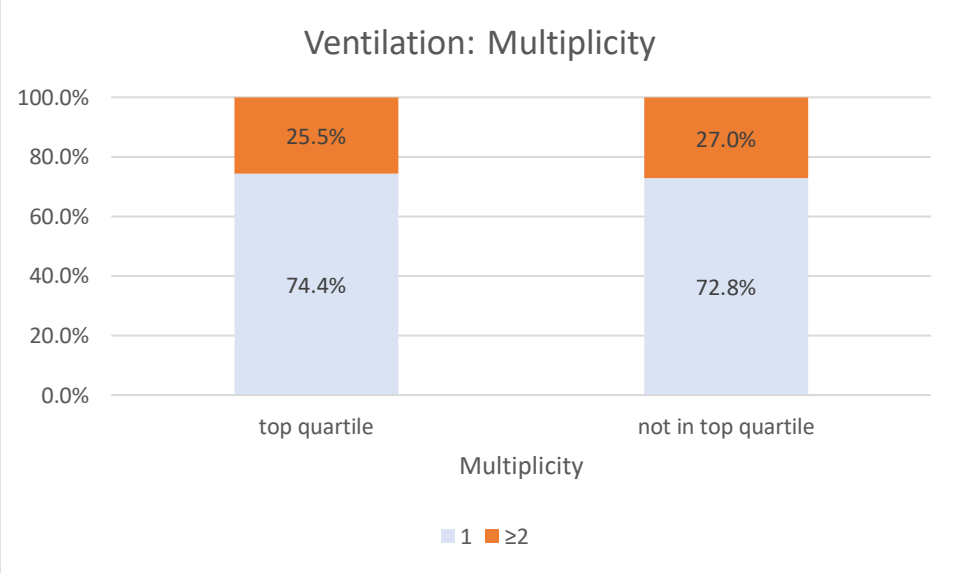


Figure 87 Proportion of singletons versus pregnancies with multiple foetuses when categorising units according to Measure 3 of my non-NNAP MQC

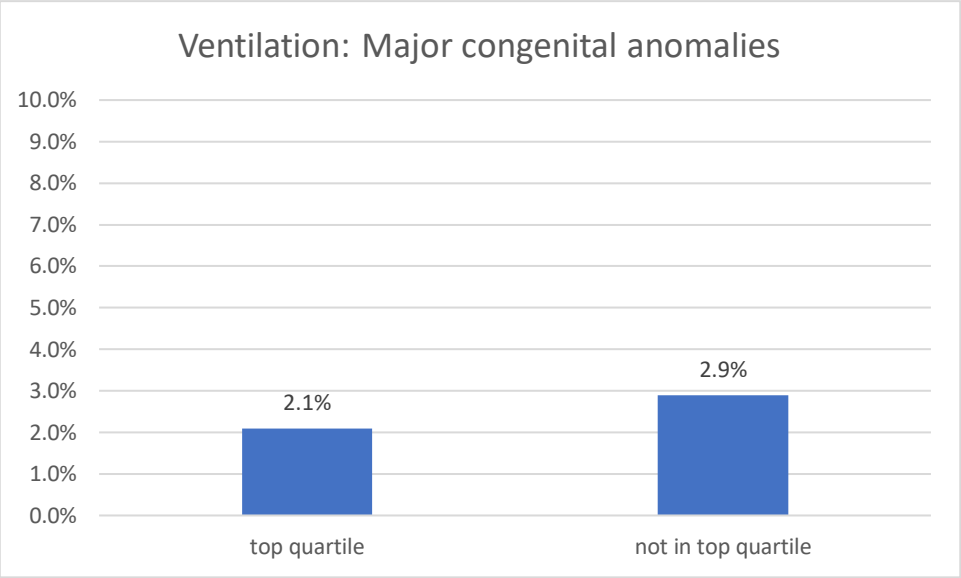


Figure 88 Proportion of babies with significant congenital anomalies when categorising units according to Measure 3 of my non-NNAP MQC

15.4 Measure 4 - babies requiring intensive care on day one of life provided with 1:1 nursing care

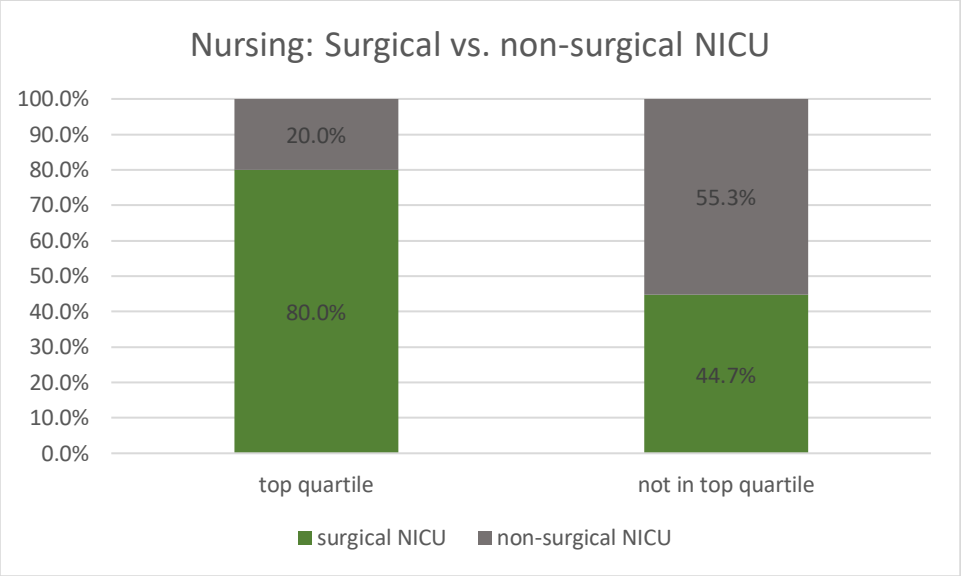


Figure 89 Proportion of babies born in surgical versus non-surgical NICU when categorising units according to Measure 4 of my non-NNAP MQC

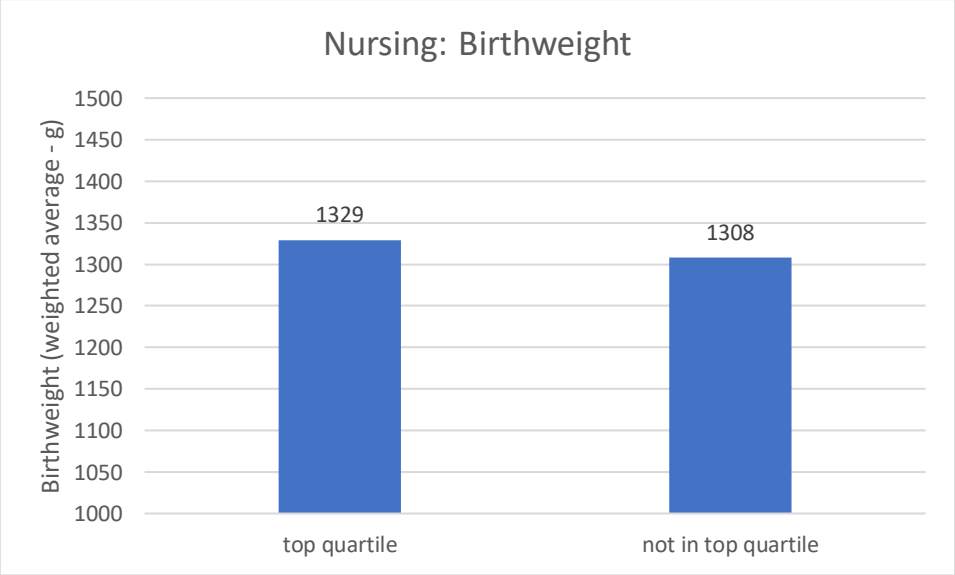


Figure 90 Mean birthweight (weighted average) of babies born in units categorised according to Measure 4 of my non-NNAP MQC

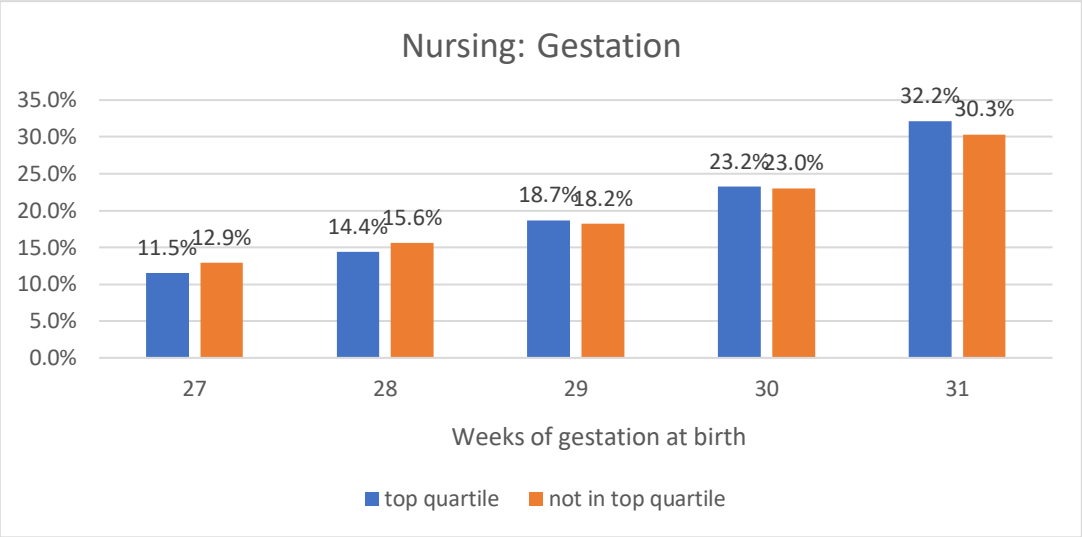


Figure 91 Proportion of babies born at each gestational week when categorising units according to Measure 4 of my non-NNAP MQC

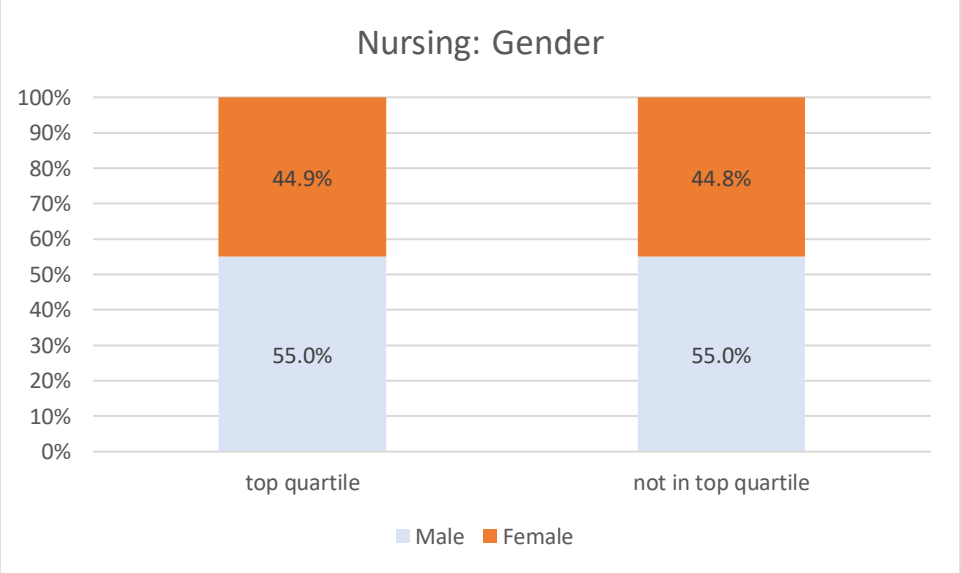


Figure 92 Proportion of male versus female babies when categorising units according to Measure 4 of my non-NNAP MQC

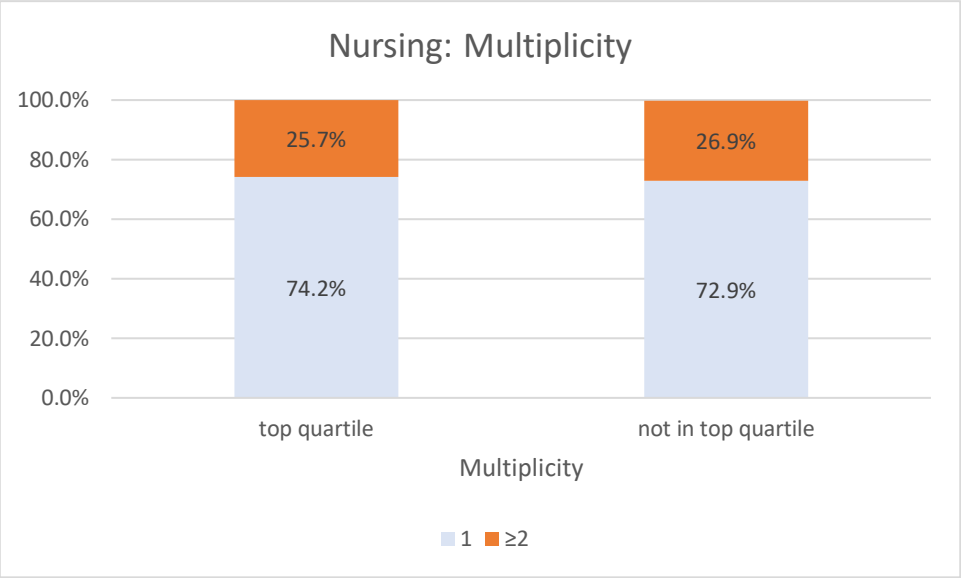


Figure 93 Proportion of singletons versus pregnancies with multiple foetuses when categorising units according to Measure 4 of my non-NNAP MQC

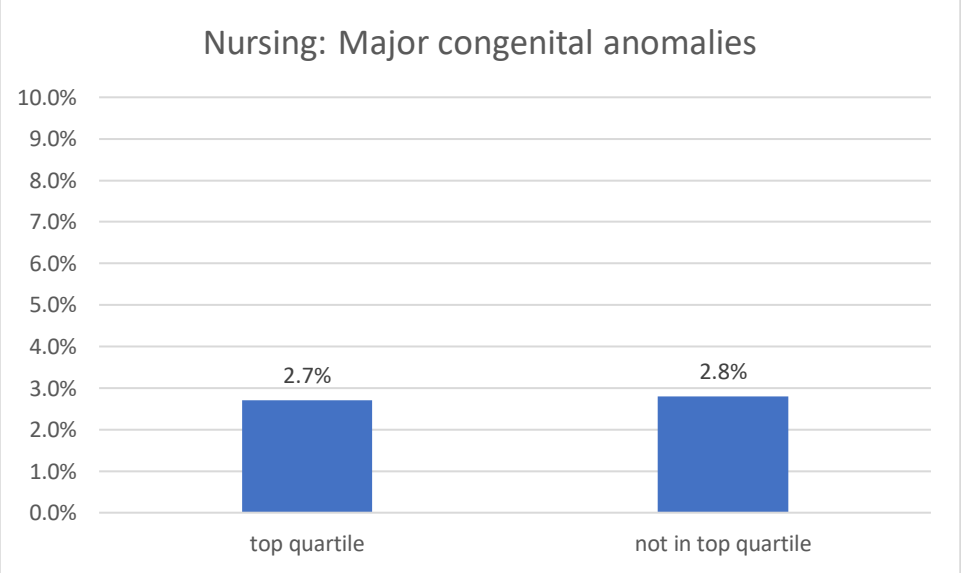


Figure 94 Proportion of babies with major congenital anomalies when categorising units according to Measure 4 of my non-NNAP MQC

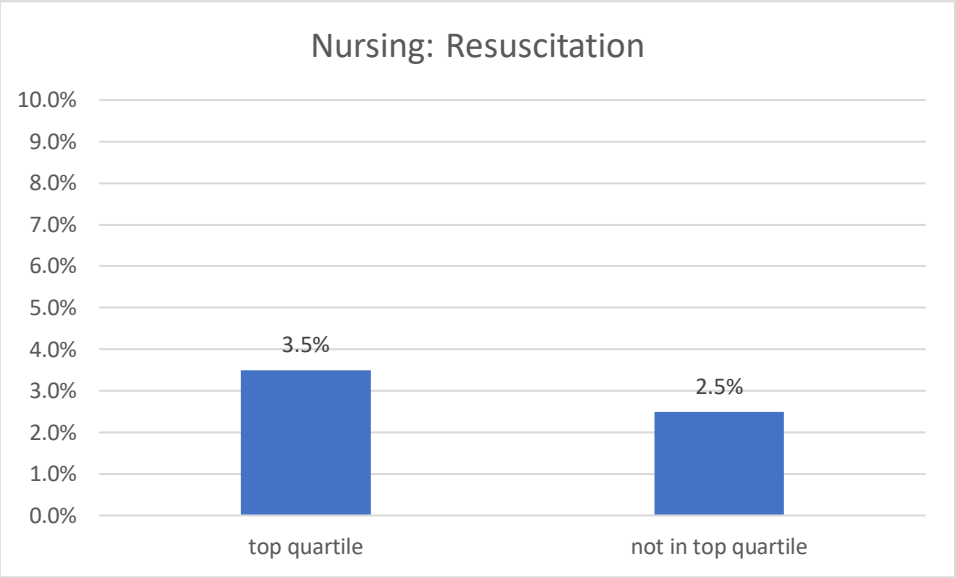


Figure 95 Proportion of babies requiring significant resuscitation when categorising units according to Measure 4 of my non-NNAP MQC

15.5 Combined measures of quality of care (non-NNAP)

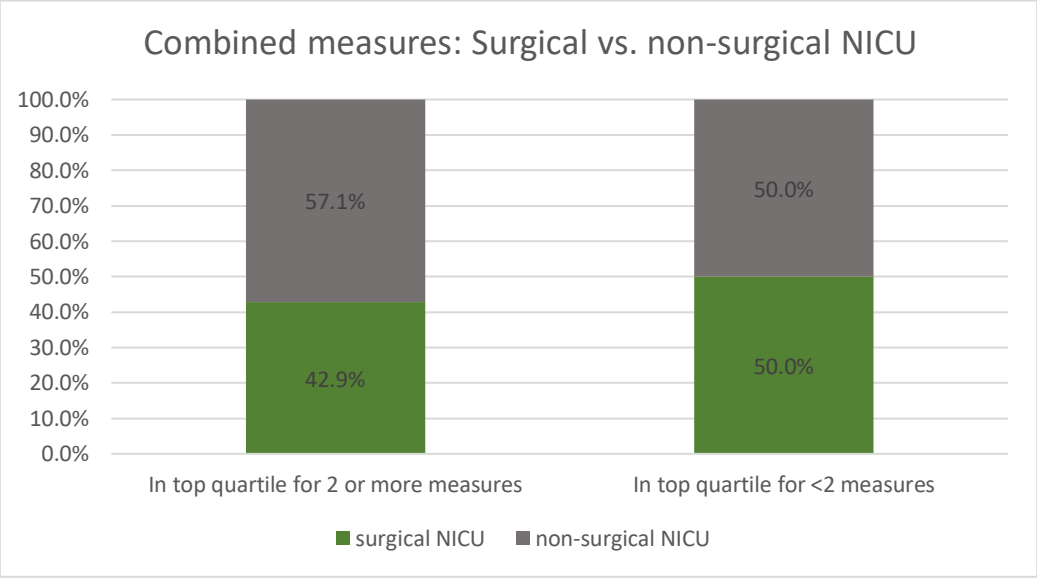


Figure 96 Proportion of babies born in surgical versus non-surgical NICU when categorising units according to the combination of my non-NNAP MQC Measures

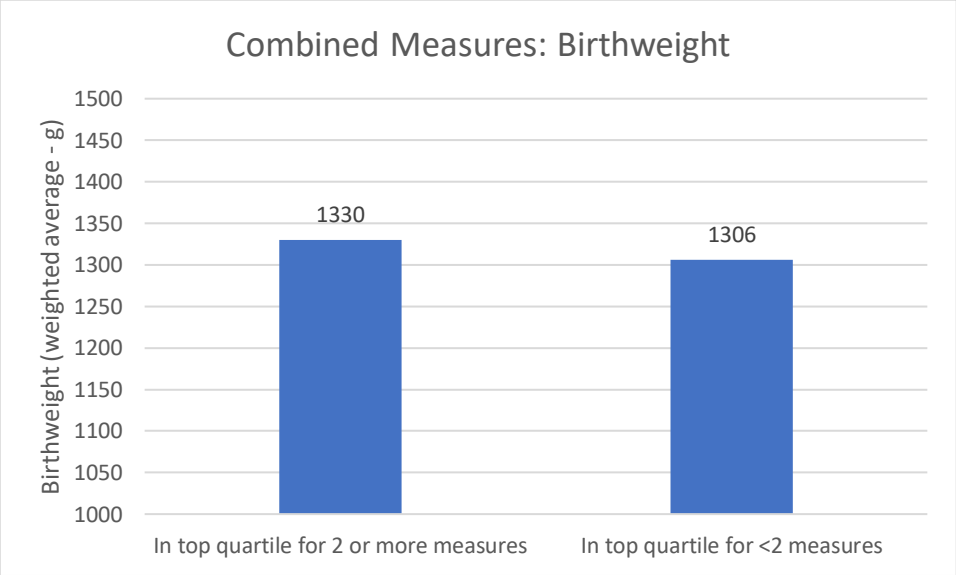


Figure 97 Mean birthweight (weighted average) of babies born in units categorised according to the combination of my non-NNAP MQC Measures

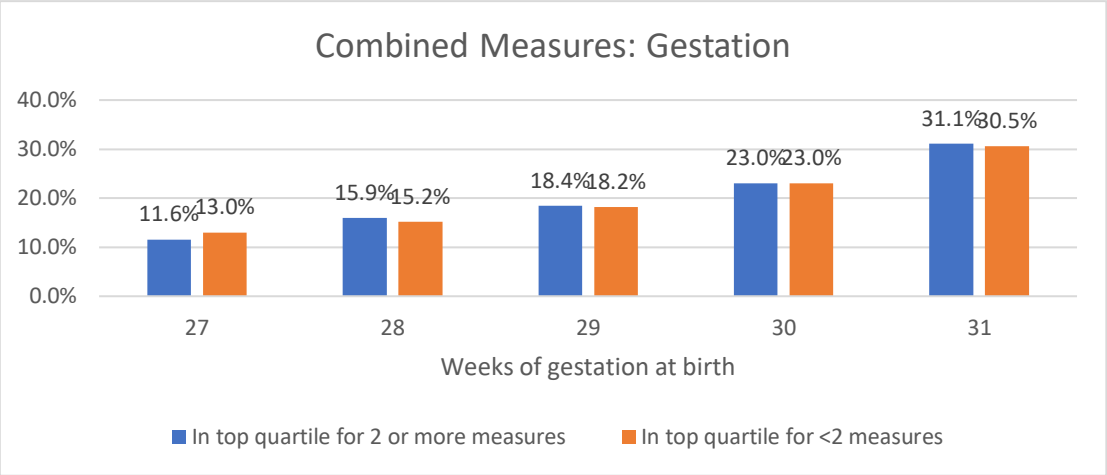


Figure 98 Proportion of babies born at each gestational week when categorising units according to the combination of my non-NNAP MQC Measures

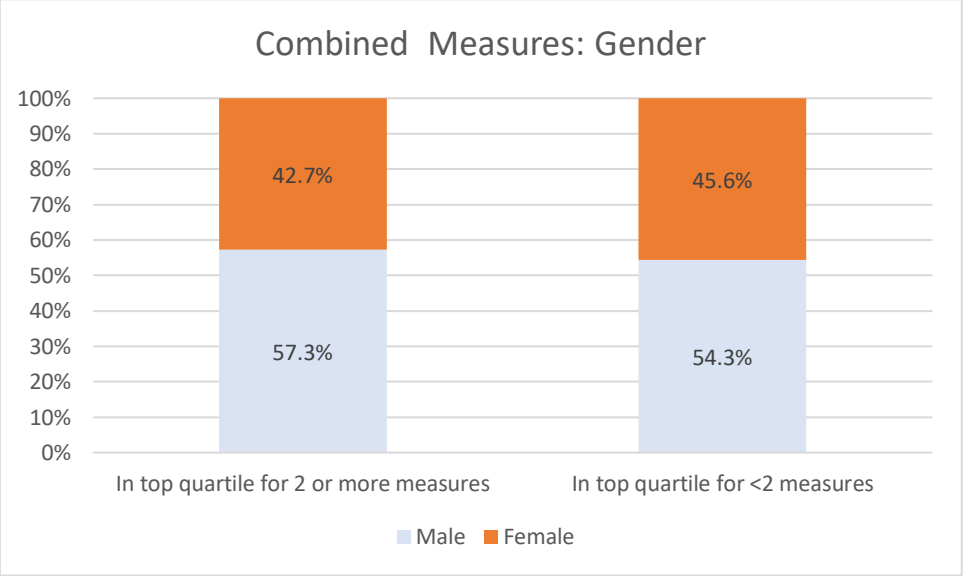


Figure 99 Proportion of male versus female babies when categorising units according to the combination of my non-NNAP MQC Measures

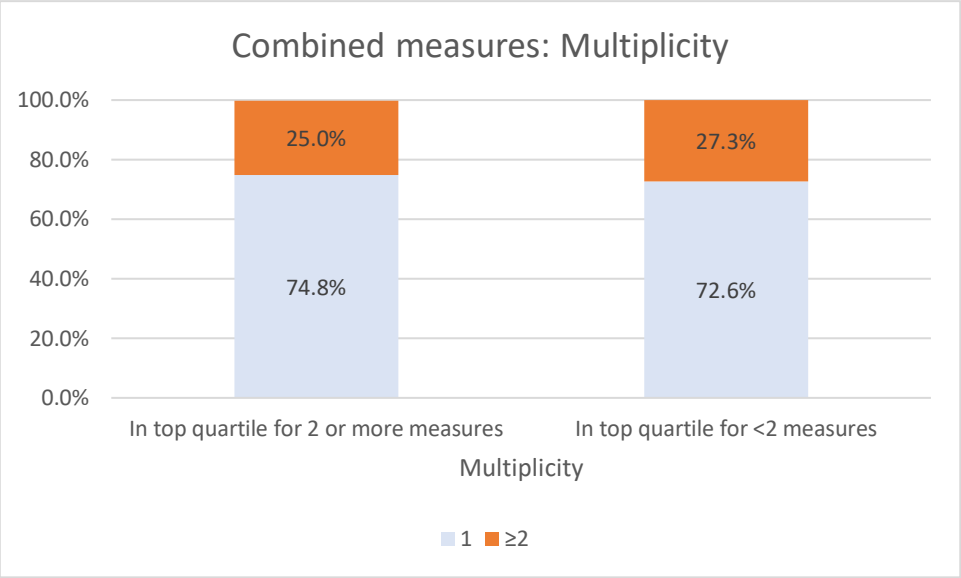


Figure 100 Proportion of singletons versus pregnancies with multiple foetuses when categorising units according to the combination of my non-NNAP MQC Measures

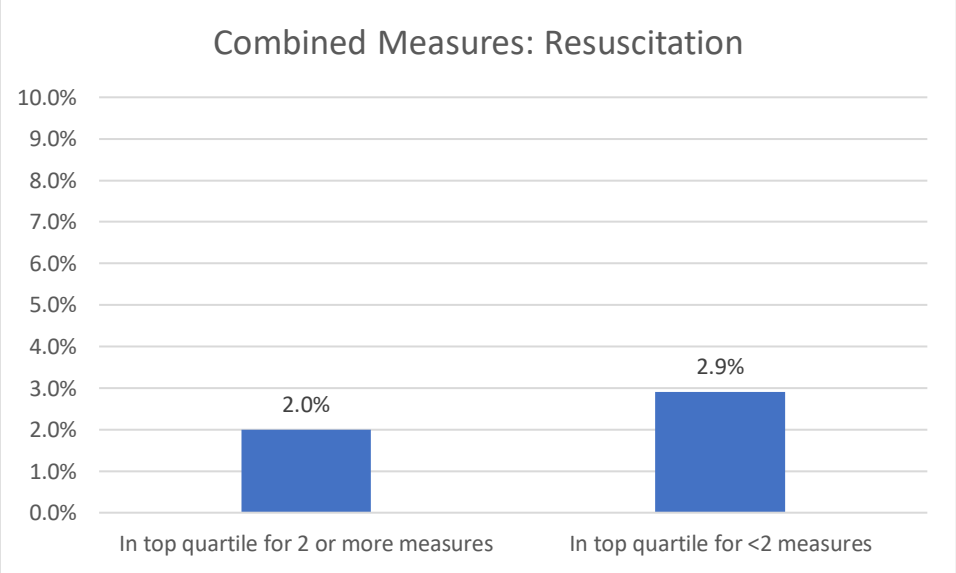


Figure 101 Proportion of babies requiring significant resuscitation when categorising units according to the combination of my non-NNAP MQC Measures

16 Appendix VI: assumptions for multivariate analyses

16.1 Collinearity

16.1.1 Outcome variable: mortality

16.1.1.1 Pearson correlations for non-NNAP MQC logistic regression

Correlations

	GWk	BBW	Gender	FetNum	IMDQ	MetOfRes	PoB	group
GWk	1	.629**	-.004	.075**	.019	-.048**	-.085**	-.011
BBW	.629**	1	-.109**	-.008	.033*	-.007	-.098**	-.029
Gender	-.004	-.109**	1	.010	-.030*	-.018	.010	.028
FetNum	.075**	-.008	.010	1	.063**	-.042**	.013	.019
IMDQ	.019	.033*	-.030*	.063**	1	-.019	-.063**	-.123**
MetOfRes	-.048**	-.007	-.018	-.042**	-.019	1	.006	.024
PoB	-.085**	-.098**	.010	.013	-.063**	.006	1	.265**
group	-.011	-.029	.028	.019	-.123**	.024	.265**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 51 Pearson correlations for non-NNAP MQC logistic regression

16.1.1.2 Pearson correlations for NNAP audit measures logistic regression

Correlations

	GWk	BBW	Gender	FetNum	IMDQ	MetOfRes	PoB	group
GWk	1	.627**	-.005	.075**	.018	-.051**	-.070**	-.047**
BBW	.627**	1	-.112**	-.005	.029	-.011	-.080**	-.014
Gender	-.005	-.112**	1	.005	-.034*	-.017	.019	-.018
FetNum	.075**	-.005	.005	1	.061**	-.046**	.016	.009
IMDQ	.018	.029	-.034*	.061**	1	-.018	-.065**	-.103**
MetOfRes	-.051**	-.011	-.017	-.046**	-.018	1	.010	.024
PoB	-.070**	-.080**	.019	.016	-.065**	.010	1	.236**
group	-.047**	-.014	-.018	.009	-.103**	.024	.236**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 52 Pearson correlations for NNAP audit measures logistic regression

16.1.2 Outcome variable: LOS

16.1.2.1 Pearson correlations and variance inflation factor (VIF) for non-NNAP MQC linear regression

Correlations

	GWk	BBW	Gender	FetNum	IMDQ	MetOfRes	PoB	Group
GWk	1	.625**	-.002	.075**	.018	-.033*	-.079**	-.013
BBW	.625**	1	-.108**	-.009	.034*	-.008	-.093**	-.027
Gender	-.002	-.108**	1	.017	-.034*	-.032*	.008	.026
FetNum	.075**	-.009	.017	1	.063**	-.040**	.014	.017
IMDQ	.018	.034*	-.034*	.063**	1	-.020	-.063**	-.125**
MetOfRes	-.033*	-.008	-.032*	-.040**	-.020	1	.000	.025
PoB	-.079**	-.093**	.008	.014	-.063**	.000	1	.267**
Group	-.013	-.027	.026	.017	-.125**	.025	.267**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 53 Pearson correlations for non-NNAP MQC linear regression

Coefficients[a]

Model		Collinearity Statistics: Tolerance	VIF
1	Group	.916	1.092
	GWk	.598	1.674
	BBW	.593	1.687
	Gender	.979	1.021
	FetNum	.983	1.017
	IMDQ	.977	1.023
	MetOfRes	.995	1.005
	PoB	.920	1.088

a. Dependent Variable: LOS

Table 54 Variance inflation factor (VIF) for non-NNAP MQC linear regression

16.1.2.2 Pearson correlations and variance inflation factor (VIF) for NNAP audit measures linear regression

Correlations

	GWk	BBW	Gender	FetNum	IMDQ	MetOfRes	PoB	group
GWk	1	.623**	-.003	.075**	.019	-.035*	-.065**	-.045**
BBW	.623**	1	-.111**	-.007	.031	-.012	-.074**	-.017
Gender	-.003	-.111**	1	.011	-.037*	-.032*	.017	-.017
FetNum	.075**	-.007	.011	1	.060**	-.047**	.017	.006
IMDQ	.019	.031	-.037*	.060**	1	-.016	-.066**	-.105**
MetOfRes	-.035*	-.012	-.032*	-.047**	-.016	1	.003	.010
PoB	-.065**	-.074**	.017	.017	-.066**	.003	1	.238**
group	-.045**	-.017	-.017	.006	-.105**	.010	.238**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 55 Pearson correlations for NNAP audit measures linear regression

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	group	.933	1.071
	GWk	.600	1.668
	BBW	.596	1.678
	Gender	.978	1.023
	FetNum	.984	1.017
	IMDQ	.981	1.019
	MetOfRes	.996	1.004
	PoB	.936	1.068

a. Dependent Variable: LOS

Table 56 Variance inflation factor (VIF) for NNAP audit measures linear regression

16.2 Linearity to the logit for continuous confounding variables (for logistic regression)

16.2.1 Box-Tidwell test for non-NNAP MQC

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	GWk	11.215	15.671	.512	1	.474	74207.651
	BBW	-.046	.010	19.119	1	<.001	.955
	Gender(1)	-.025	.193	.017	1	.897	.975
	FetNum(1)	-.100	.237	.178	1	.673	.905
	IMDQ			.382	4	.984	
	IMDQ(1)	.018	.256	.005	1	.945	1.018
	IMDQ(2)	.029	.284	.010	1	.919	1.029
	IMDQ(3)	-.162	.316	.263	1	.608	.851
	IMDQ(4)	-.045	.320	.020	1	.888	.956
	MetOfRes(1)	2.123	.276	59.079	1	<.001	8.355
	PoB			7.683	2	.021	
	PoB(1)	.339	.249	1.852	1	.174	1.403
	PoB(2)	.689	.249	7.634	1	.006	1.992
	group(1)	-.208	.234	.786	1	.375	.812
	ln_GWk	-2.681	3.597	.556	1	.456	.068
	ln_BBW	.006	.001	18.491	1	<.001	1.006
	Constant	-60.032	102.859	.341	1	.559	.000

a. Variable(s) entered on step 1: GWk, BBW, Gender, FetNum, IMDQ, MetOfRes, PoB, group, ln_GWk, ln_BBW.

Table 57 Box-Tidwell test for non-NNAP MQC

16.2.2 Box-Tidwell test for NNAP audit measures

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	GWk	5.433	16.247	.112	1	.738	228.939
	BBW	-.046	.011	18.226	1	<.001	.955
	Gender(1)	-.018	.201	.008	1	.929	.982
	FetNum(1)	-.177	.251	.499	1	.480	.838
	IMDQ			.553	4	.968	
	IMDQ(1)	.104	.268	.150	1	.698	1.109
	IMDQ(2)	.128	.294	.189	1	.664	1.136
	IMDQ(3)	-.102	.329	.097	1	.756	.903
	IMDQ(4)	.018	.330	.003	1	.957	1.018
	MetOfRes(1)	2.024	.288	49.235	1	<.001	7.567
	PoB			5.951	2	.051	
	PoB(1)	.375	.264	2.021	1	.155	1.454
	PoB(2)	.640	.262	5.945	1	.015	1.896
	group(1)	-.330	.298	1.222	1	.269	.719
	ln_GWk	-1.360	3.729	.133	1	.715	.257
	ln_BBW	.006	.001	17.718	1	<.001	1.006
	Constant	-21.746	106.631	.042	1	.838	.000

a. Variable(s) entered on step 1: GWk, BBW, Gender, FetNum, IMDQ, MetOfRes, PoB, group, ln_GWk, ln_BBW.

Table 58 Box-Tidwell test for NNAP audit measures

16.3 Normality and homoscedasticity (for linear regression)

16.3.1 Non-NNAP MQC

16.3.1.1 Scatterplot matrix for non-NNAP MQC pre-log transformation

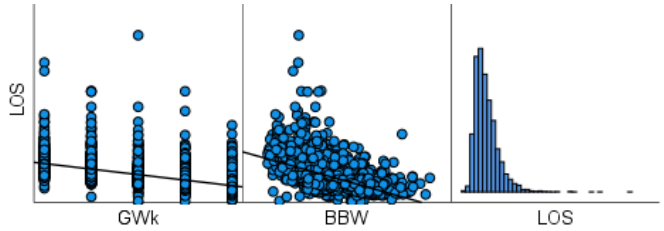


Figure 102 Scatterplot matrix for non-NNAP MQC pre-log transformation

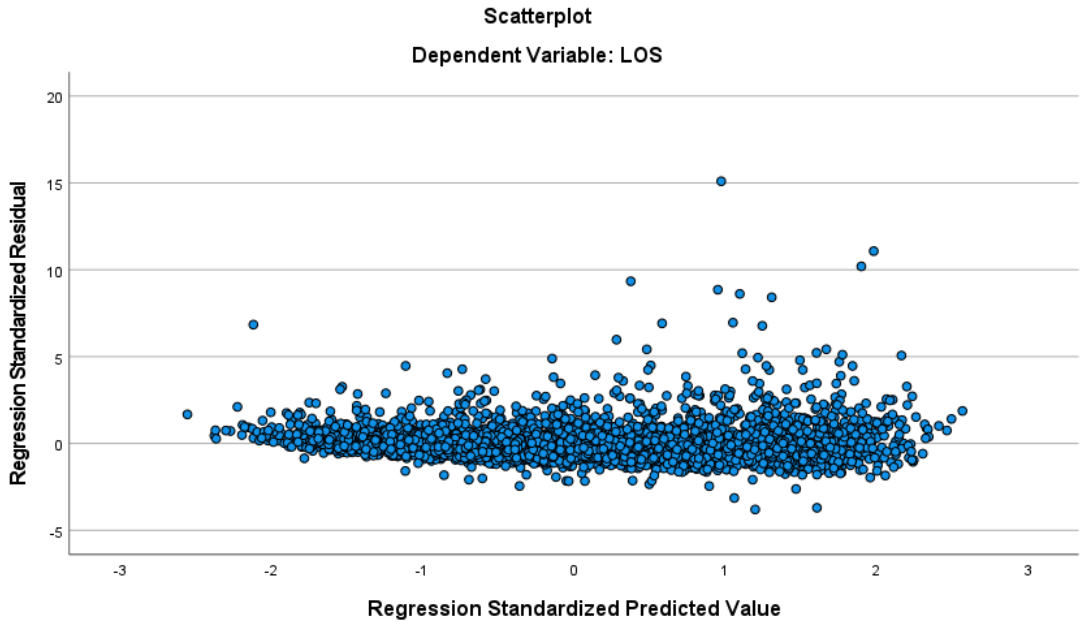


Figure 103 Scatterplot of Regression Standardized Residual versus Regression Standardized Predicted Value for non-NNAP MQC pre-log transformation

16.3.1.2 Scatterplot matrix for non-NNAP MQC post-log transformation

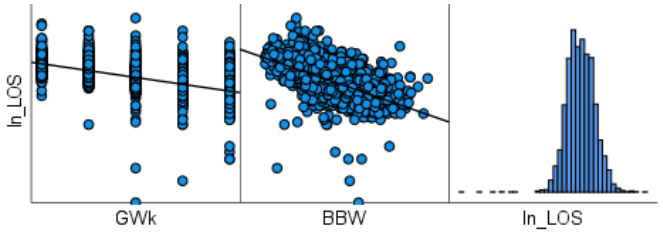


Figure 104 Scatterplot matrix for non-NNAP MQC post-log transformation

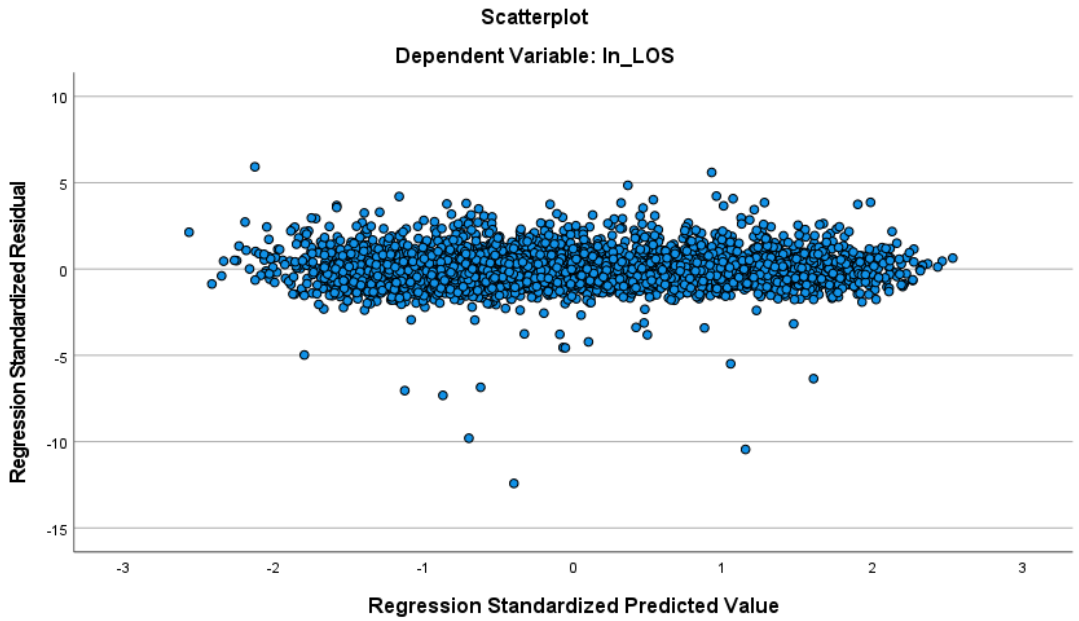


Figure 105 Scatterplot of Regression Standardized Residual versus Regression Standardized Predicted Value for non-NNAP MQC post-log transformation

16.3.2 NNAP audit measures

16.3.2.1 Scatterplot matrix for NNAP audit measures pre-log transformation

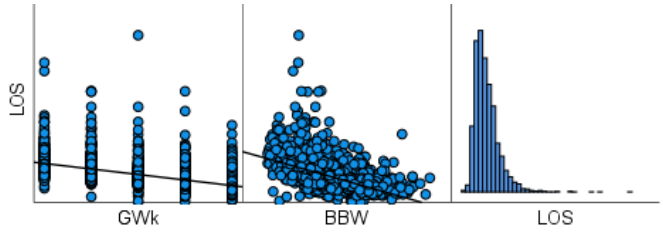


Figure 106 Scatterplot matrix for NNAP audit measures pre-log transformation

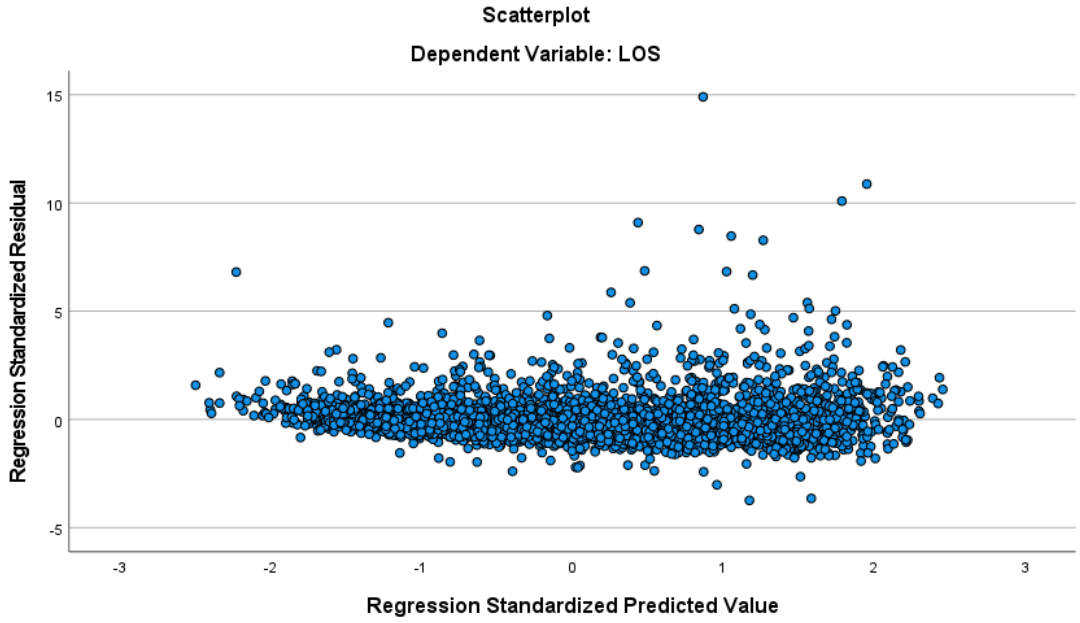


Figure 107 Scatterplot of Regression Standardized Residual versus Regression Standardized Predicted Value for NNAP audit measures pre-log transformation

16.3.2.2 Scatterplot matrix for NNAP audit measures post-log transformation

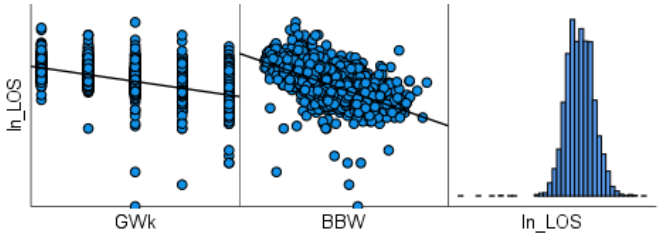


Figure 108 Scatterplot matrix for NNAP audit measures post-log transformation

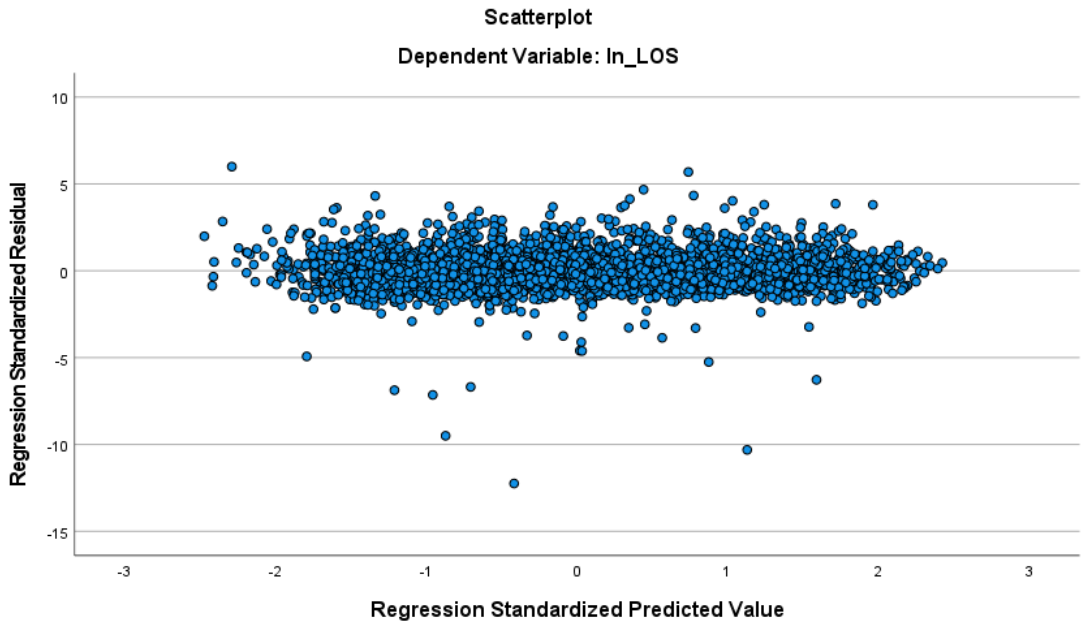


Figure 109 Scatterplot of Regression Standardized Residual versus Regression Standardized Predicted Value for NNAP audit measures post-log transformation

17 References

1. Ismail AQT, Boyle EM, Pillay T, OptiPrem Study G. The impact of level of neonatal care provision on outcomes for preterm babies born between 27 and 31 weeks of gestation, or with a birth weight between 1000 and 1500 g: a review of the literature. BMJ Paediatr Open. 2020;4(1):e000583.
2. UK Clinical Trials Gateway. OPTI-Prem: Optimising Neonatal Service Provision for Preterm Babies Born Between 27 and 31 Weeks of Gestation in England 2017 [Available from: https://ukctg.nihr.ac.uk/trials/trial-details/trial-details?trialNumber=NCT02994849.
3. Neonatal Data Analysis Unit. NDAU 2014 Report. 2014.
4. Operational Delivery Networks. [Available from: https://www.england.nhs.uk/ourwork/part-rel/odn/.
5. Pillay T, Modi N, Rivero-Arias O, Manktelow B, Seaton SE, Armstrong N, et al. Optimising neonatal service provision for preterm babies born between 27 and 31 weeks gestation in England (OPTI-PREM), using national data, qualitative research and economic analysis: a study protocol. BMJ Open. 2019;9(8):e029421.
6. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med. 1979;301(25):1364-9.
7. Luft HS, Garnick DW, Mark DH, McPhee SJ. Hospital Volume, Physician Volume, and Patient Outcomes: Assessing the Evidence: Health Administration Press; 1990.
8. Pearson G, Shann F, Barry P, Vyas J, Thomas D, Powell C, et al. Should paediatric intensive care be centralised? Trent versus Victoria. Lancet. 1997;349(9060):1213-7.
9. Tilford JM, Simpson PM, Green JW, Lensing S, Fiser DH. Volume-outcome relationships in pediatric intensive care units. Pediatrics. 2000;106(2 Pt 1):289-94.
10. Jiang S, Warre R, Qiu X, O'Brien K, Lee SK. Parents as practitioners in preterm care. Early Hum Dev. 2014;90(11):781-5.
11. Bliss. Financial costs facing families [Available from: https://www.bliss.org.uk/research-campaigns/campaigns/families-kept-apart-take-action-now/financial-costs-facing-families.
12. POPPY Steering Group. Family-centred care in neonatal units. A summary of research results and recommendations from the POPPY project. 2009.
13. Marlow N, Bryan Gill A. Establishing neonatal networks: the reality. Arch Dis Child Fetal Neonatal Ed. 2007;92(2):F137-42.
14. Perkins BB. Rethinking perinatal policy: history and evaluation of minimum volume and level-of-care standards. J Public Health Policy. 1993;14(3):299-319.
15. McCormick MC, Richardson DK. Access to neonatal intensive care. Future Child. 1995;5(1):162-75.
16. Ryan GM, Jr. Toward improving the outcome of pregnancy Recommendations for the regional development of perinatal health services. Obstet Gynecol. 1975;46(4):375-84.
17. Yeast JD, Poskin M, Stockbauer JW, Shaffer S. Changing patterns in regionalization of perinatal care and the impact on neonatal mortality. Am J Obstet Gynecol. 1998;178(1 Pt 1):131-5.
18. American Academy of Pediatrics Committee on Fetus and Newborn. Levels of neonatal care. Pediatrics. 2012;130(3):587-97.

19. Peddle LJ, Brown H, Buckley J, Dixon W, Kaye J, Muise M, et al. Voluntary regionalization and associated trends in perinatal care: The Nova Scotia Reproductive Care Program. American Journal of Obstetrics and Gynecology. 1983;145(2):170-6.
20. Committee on Reproductive Care. Reproductive Care: Towards the 1990s: Second Report of the Advisory Committee on Reproductive Care. Toronto, Ontario; 1988.
21. Perinatal Services BC. Perinatal Tiers of Service Module. 2016.
22. Chien LY, Whyte R, Aziz K, Thiessen P, Matthew D, Lee SK, et al. Improved outcome of preterm infants when delivered in tertiary care centers. Obstet Gynecol. 2001;98(2):247-52.
23. Papiernik E, Keith LG. The regionalization of perinatal care in France--description of a missing policy. Eur J Obstet Gynecol Reprod Biol. 1995;61(2):99-103.
24. Viisainen K, Gissler M, Hemminki E. Birth outcomes by level of obstetric care in Finland: a catchment area based analysis. J Epidemiol Community Health. 1994;48(4):400-5.
25. Blondel B, Papiernik E, Delmas D, Kunzel W, Weber T, Maier RF, et al. Organisation of obstetric services for very preterm births in Europe: results from the MOSAIC project. BJOG. 2009;116(10):1364-72.
26. National Health and Medical Research Council. Statements adopted at 85th session, Adelaide, June 1978. Med J Aust. 1978;2 Suppl 3:1-4.
27. Lumley J. Better perinatal health. Australia. Lancet. 1980;1(8159):79-81.
28. Kitchen W, Ford G, Orgill A, Rickards A, Astbury J, Lissenden J, et al. Outcome in infants with birth weight 500 to 999 gm: a regional study of 1979 and 1980 births. J Pediatr. 1984;104(6):921-7.
29. Changing outcome for infants of birth-weight 500-999 g born outside level 3 centres in Victoria. The Victorian Infant Collaborative Study Group. Aust N Z J Obstet Gynaecol. 1997;37(3):253-7.
30. Rosenblatt RA, Reinken J, Shoemack P. Is obstetrics safe in small hospitals? Evidence from New Zealand's regionalised perinatal system. Lancet. 1985;2(8452):429-32.
31. Ministry of Health New Zealand. Report on Maternity 1999 Wellington; 2001.
32. Shaffer ER. State Policies and Regional Neonatal Care: Progress and Challenges 25 Years After TIOP. White Plains, NY: March of Dimes; 2001.
33. State of Victoria. Defining levels of care for Victorian newborn services. In: Department of Health and Human Services, editor. 2015.
34. OFFICE OF KIDS AND FAMILIES. NSW Maternity and Neonatal Service Capability Framework. In: NSW MINISTRY OF HEALTH, editor. 2016.
35. Department of Health Western Australia. Framework for the care of neonates in Western Australia. 2009.
36. South Australia. Dept. for Health and Aging. Standards for Maternal and Neonatal Services in South Australia. 2015.
37. Ministry of Health New Zealand. A Review of Neonatal Intensive Care Provision in New Zealand. 2004.
38. Finnish Medical Society Duodecim and the Finnish Gynecologist Association. Premature birth. 2018.
39. Swedish Neonatal Quality Register. Årsrapport SNQ 2016. 2016.
40. Parant O, Maillard F, Tsatsaris V, Delattre M, Subtil D, Goffinet F, et al. Management of threatened preterm delivery in France: a national practice survey (the EVAPRIMA study). BJOG. 2008;115(12):1538-46.

41. British Association of Perinatal Medicine (BAPM). Standards for Hospitals Providing Neonatal Intensive and High Dependency Care (Second Edition). London; 2001.
42. Department of Health (DH). Report of the Neonatal Intensive Care Services Review Group. UK; 2003.
43. Black DP, Fyfe IM. The safety of obstetric services in small communities in northern Ontario. Can Med Assoc J. 1984;130(5):571-6.
44. Hein HA, Burmeister LF. The effect of ten years of regionalized perinatal health care in Iowa, U.S.A. Eur J Obstet Gynecol Reprod Biol. 1986;21(1):33-48.
45. NHS and Department of Health (DH). Toolkit for High-Quality Neonatal Services. London; 2009.
46. British Association of Perinatal Medicine (BAPM). Service Standards For Hospitals Providing Neonatal Care (3rd edition) 2010.
47. Marlow N, Bennett C, Draper ES, Hennessy EM, Morgan AS, Costeloe KL. Perinatal outcomes for extremely preterm babies in relation to place of birth in England: the EPICure 2 study. Arch Dis Child Fetal Neonatal Ed. 2014;99(3):F181-8.
48. Group E, Fellman V, Hellstrom-Westas L, Norman M, Westgren M, Kallen K, et al. One-year survival of extremely preterm infants after active perinatal care in Sweden. JAMA. 2009;301(21):2225-33.
49. Anderson JG, Baer RJ, Partridge JC, Kuppermann M, Franck LS, Rand L, et al. Survival and Major Morbidity of Extremely Preterm Infants: A Population-Based Study. Pediatrics. 2016;138(1).
50. Boland RA, Dawson JA, Davis PG, Doyle LW. Why birthplace still matters for infants born before 32 weeks: Infant mortality associated with birth at 22-31 weeks' gestation in non-tertiary hospitals in Victoria over two decades. Aust N Z J Obstet Gynaecol. 2015;55(2):163-9.
51. Report of working group of the British Association of Perinatal Medicine and Neonatal Nurses Association on categories of babies requiring neonatal care. Arch Dis Child. 1992;67(7 Spec No):868-9.
52. Field D, Draper ES. Survival and place of delivery following preterm birth: 1994-96. Arch Dis Child Fetal Neonatal Ed. 1999;80(2):F111-4.
53. Field D, Hodges S, Mason E, Burton P. Survival and place of treatment after premature delivery. Arch Dis Child. 1991;66(4 Spec No):408-10; discussion 10-1.
54. NHS England. Neonatal Critical Care Transformation Review. 2015.
55. Lasswell SM, Barfield WD, Rochat RW, Blackmon L. Perinatal regionalization for very low-birth-weight and very preterm infants: a meta-analysis. JAMA. 2010;304(9):992-1000.
56. Rautava L, Lehtonen L, Peltola M, Korvenranta E, Korvenranta H, Linna M, et al. The effect of birth in secondary- or tertiary-level hospitals in Finland on mortality in very preterm infants: a birth-register study. Pediatrics. 2007;119(1):e257-63.
57. Bolisetty S, Legge N, Bajuk B, Lui K, New South W, the Australian Capital Territory Neonatal Intensive Care Units' Data C. Preterm infant outcomes in New South Wales and the Australian Capital Territory. J Paediatr Child Health. 2015;51(7):713-21.
58. Empana JP, Subtil D, Truffert P. In-hospital mortality of newborn infants born before 33 weeks of gestation depends on the initial level of neonatal care: the EPIPAGE study. Acta Paediatr. 2003;92(3):346-51.
59. Phibbs CS, Baker LC, Caughey AB, Danielsen B, Schmitt SK, Phibbs RH. Level and volume of neonatal intensive care and mortality in very-low-birth-weight infants. N Engl J Med. 2007;356(21):2165-75.

60. Warner B, Musial MJ, Chenier T, Donovan E. The effect of birth hospital type on the outcome of very low birth weight infants. Pediatrics. 2004;113(1 Pt 1):35-41.
61. Cochrane Library (neonatal). Search filter for neonatal population [Available from: http://neonatal.cochrane.org/resources-review-authors.
62. Committee on Perinatal Health. Toward Improving the Outcome of Pregnancy: Recommendations for the Regional Development of Maternal and Perinatal Health Services. White Plains, NY: March of Dimes National Foundation; 1976.
63. Lamont RF, Dunlop PD, Crowley P, Levene MI, Elder MG. Comparative mortality and morbidity of infants transferred in utero or postnatally. J Perinat Med. 1983;11(4):200-3.
64. Johansson S, Montgomery SM, Ekbom A, Olausson PO, Granath F, Norman M, et al. Preterm Delivery, Level of Care, and Infant Death in Sweden: A Population-Based Study. Pediatrics. 2004;113(5):1230-5.
65. Truffert P, Goujard J, Dehan M, Vodovar M, Breart G. Outborn status with a medical neonatal transport service and survival without disability at two years. A population-based cohort survey of newborns of less than 33 weeks of gestation. Eur J Obstet Gynecol Reprod Biol. 1998;79(1):13-8.
66. Hauspy J, Jacquemyn Y, Van Reempts P, Buytaert P, Van Vliet J. Intrauterine versus postnatal transport of the preterm infant: a short-distance experience. Early Hum Dev. 2001;63(1):1-7.
67. Lee SK, McMillan DD, Ohlsson A, Boulton J, Lee DS, Ting S, et al. The benefit of preterm birth at tertiary care centers is related to gestational age. Am J Obstet Gynecol. 2003;188(3):617-22.
68. Holmgren PA, Hogberg U. The very preterm infant - a population-based study. Acta Obstet Gynecol Scand. 2001;80(6):525-31.
69. Jonas HA, Lumley J. Trends in stillbirths and neonatal deaths for very pre-term infants (< 32 weeks' gestation) born in Victoria, 1986-1993. Aust N Z J Obstet Gynaecol. 1997;37(1):59-66.
70. Ozminkowski RJ, Wortman PM, Roloff DW. Inborn/outborn status and neonatal survival: a meta-analysis of non-randomized studies. Stat Med. 1988;7(12):1207-21.
71. Improvement of outcome for infants of birth weight under 1000 g. The Victorian Infant Collaborative Study Group. Arch Dis Child. 1991;66(7 Spec No):765-9.
72. Lumley J. The safety of small maternity hospitals in Victoria 1982-84. Community Health Stud. 1988;12(4):386-93.
73. Chung JH, Phibbs CS, Boscardin WJ, Kominski GF, Ortega AN, Gregory KD, et al. Examining the effect of hospital-level factors on mortality of very low birth weight infants using multilevel modeling. J Perinatol. 2011;31(12):770-5.
74. Goodman DC, Fisher ES, Little GA, Stukel TA, Chang CH, Schoendorf KS. The relation between the availability of neonatal intensive care and neonatal mortality. N Engl J Med. 2002;346(20):1538-44.
75. Matthews TJ, MacDorman MF, Thoma ME. Infant Mortality Statistics From the 2013 Period Linked Birth/Infant Death Data Set. Natl Vital Stat Rep. 2015;64(9):1-30.
76. Lubchenco LO, Searls DT, Brazie JV. Neonatal mortality rate: relationship to birth weight and gestational age. J Pediatr. 1972;81(4):814-22.
77. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6.
78. Lee SK, McMillan DD, Ohlsson A, Boulton J, Lee DSC, Ting S, et al. The benefit of preterm birth at tertiary care centers is related to gestational age. American Journal of Obstetrics and Gynecology. 2003;188(3):617-22.

79. Gortmaker S, Sobol A, Clark C, Walker DK, Geronimus A. The survival of very low-birth weight infants by level of hospital of birth: a population study of perinatal systems in four states. Am J Obstet Gynecol. 1985;152(5):517-24.
80. Gould JB, Marks AR, Chavez G. Expansion of community-based perinatal care in California. J Perinatol. 2002;22(8):630-40.
81. Miller TC, Densberger M, Krogman J. Maternal transport and the perinatal denominator. Am J Obstet Gynecol. 1983;147(1):19-24.
82. Mohamed MA, Aly H. Transport of premature infants is associated with increased risk for intraventricular haemorrhage. Arch Dis Child Fetal Neonatal Ed. 2010;95(6):F403-7.
83. Obladen M, Luttkus A, Rey M, Metze B, Hopfenmuller W, Dudenhausen JW. Differences in morbidity and mortality according to type of referral of very low birthweight infants. J Perinat Med. 1994;22(1):53-64.
84. Powell SL, Holt VL, Hickok DE, Easterling T, Connell FA. Recent changes in delivery site of low-birth-weight infants in Washington: impact on birth weight-specific mortality. Am J Obstet Gynecol. 1995;173(5):1585-92.
85. Powell TG, Pharoah PO. Regional neonatal intensive care: bias and benefit. Br Med J (Clin Res Ed). 1987;295(6600):690-2.
86. Sanderson M, Sappenfield WM, Jespersen KM, Liu Q, Baker SL. Association between level of delivery hospital and neonatal outcomes among South Carolina Medicaid recipients. Am J Obstet Gynecol. 2000;183(6):1504-11.
87. Watkinson M, McIntosh N. Outcome of neonatal intensive care: obstetric implications for a regional service. Br J Obstet Gynaecol. 1986;93(7):711-6.
88. Szklo M. Population-based cohort studies. Epidemiol Rev. 1998;20(1):81-90.
89. Manktelow BN. MBRRACE-UK perinatal mortality surveillance report: UK perinatal deaths for births from January to December 2014: Department of Health Sciences, University of Leicester; 2016.
90. Berger TM, Steurer MA, Woerner A, Meyer-Schiffer P, Adams M, Swiss Neonatal N. Trends and centre-to-centre variability in survival rates of very preterm infants (<32 weeks) over a 10-year-period in Switzerland. Arch Dis Child Fetal Neonatal Ed. 2012;97(5):F323-8.
91. Abdel-Latif ME, Nowak G, Bajuk B, Glass K, Harley D. Variation in hospital mortality in an Australian neonatal intensive care unit network. Arch Dis Child Fetal Neonatal Ed. 2017.
92. Horbar JD, Badger GJ, Lewit EM, Rogowski J, Shiono PH. Hospital and patient characteristics associated with variation in 28-day mortality rates for very low birth weight infants. Vermont Oxford Network. Pediatrics. 1997;99(2):149-56.
93. Vohr BR, Wright LL, Dusick AM, Perritt R, Poole WK, Tyson JE, et al. Center differences and outcomes of extremely low birth weight infants. Pediatrics. 2004;113(4):781-9.
94. Verloove-Vanhorick SP, Verwey RA, Ebeling MC, Brand R, Ruys JH. Mortality in very preterm and very low birth weight infants according to place of birth and level of care: results of a national collaborative survey of preterm and very low birth weight infants in The Netherlands. Pediatrics. 1988;81(3):404-11.
95. Kirby RS. Perinatal mortality: the role of hospital of birth. J Perinatol. 1996;16(1):43-9.
96. Merlo J, Gerdtham UG, Eckerlund I, Hakansson S, Otterblad-Olausson P, Pakkanen M, et al. Hospital level of care and neonatal mortality in low- and high-risk deliveries: reassessing the question in Sweden by multilevel analysis. Med Care. 2005;43(11):1092-100.

97. Brunet O, Lézine I. Le développement psychologique de la première enfance. 1951.
98. Watson SI, Arulampalam W, Petrou S, Marlow N, Morgan AS, Draper ES, et al. The effects of designation and volume of neonatal care on mortality and morbidity outcomes of very preterm infants in England: retrospective population-based cohort study. BMJ Open. 2014;4(7):e004856.
99. Office For National Statistics. Pregnancy and ethnic factors influencing births and infant mortality: 2013 2013 [Death rates of pre-term, full-term and post-term babies and various factors that may influence their survival]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/pregnancyandethnicfactorsinfluencingbirthsandinfantmortality/2015-10-14#gestational-age.
100. Centers for Disease Control and Prevention. Linked Birth / Infant Death Records 2015 [Available from: https://wonder.cdc.gov/lbd.html.
101. Ge WJ, Mirea L, Yang J, Bassil KL, Lee SK, Shah PS, et al. Prediction of neonatal outcomes in extremely preterm neonates. Pediatrics. 2013;132(4):e876-85.
102. Brouwer A, Groenendaal F, van Haastert IL, Rademaker K, Hanlo P, de Vries L. Neurodevelopmental outcome of preterm infants with severe intraventricular hemorrhage and therapy for post-hemorrhagic ventricular dilatation. J Pediatr. 2008;152(5):648-54.
103. Synnes AR, Chien LY, Peliowski A, Baboolal R, Lee SK, Canadian NN. Variations in intraventricular hemorrhage incidence rates among Canadian neonatal intensive care units. J Pediatr. 2001;138(4):525-31.
104. Baud O, Foix-L'Helias L, Kaminski M, Audibert F, Jarreau PH, Papiernik E, et al. Antenatal glucocorticoid treatment and cystic periventricular leukomalacia in very premature infants. N Engl J Med. 1999;341(16):1190-6.
105. Ancel PY, Livinec F, Larroque B, Marret S, Arnaud C, Pierrat V, et al. Cerebral palsy among very preterm children in relation to gestational age and neonatal ultrasound abnormalities: the EPIPAGE cohort study. Pediatrics. 2006;117(3):828-35.
106. Battersby C, Santhalingam T, Costeloe K, Modi N. Incidence of neonatal necrotising enterocolitis in high-income countries: a systematic review. Arch Dis Child Fetal Neonatal Ed. 2018;103(2):F182-F9.
107. Walker MW, Clark RH, Spitzer AR. Elevation in plasma creatinine and renal failure in premature neonates without major anomalies: terminology, occurrence and factors associated with increased risk. J Perinatol. 2011;31(3):199-205.
108. Larsson E, Carle-Petrelius B, Cernerud G, Ots L, Wallin A, Holmstrom G. Incidence of ROP in two consecutive Swedish population based studies. Br J Ophthalmol. 2002;86(10):1122-6.
109. Clyman RI. Patent Ductus Arteriosus in the Preterm Infant2012. 751-61 p.
110. Stoll BJ, Hansen N, Fanaroff AA, Wright LL, Carlo WA, Ehrenkranz RA, et al. Changes in pathogens causing early-onset sepsis in very-low-birth-weight infants. N Engl J Med. 2002;347(4):240-7.
111. Group SI. Early enteral feeding strategies for very preterm infants: current evidence from Cochrane reviews. Arch Dis Child Fetal Neonatal Ed. 2013;98(6):F470-2.
112. Royal College of Paediatrics and Child Health (RCPCH). Early years - UK-WHO growth charts and resources 2016 [Available from: https://www.rcpch.ac.uk/child-health/research-projects/uk-who-growth-charts/uk-who-growth-chart-resources-0-4-years/uk-who-0.
113. Vohr B. Long-term outcomes of moderately preterm, late preterm, and early term infants. Clin Perinatol. 2013;40(4):739-51.

114. Lebenthal E, Lee PC, Heitlinger LA. Impact of development of the gastrointestinal tract on infant feeding. J Pediatr. 1983;102(1):1-9.
115. Neu J. Gastrointestinal development and meeting the nutritional needs of premature infants. Am J Clin Nutr. 2007;85(2):629S-34S.
116. Bisset WM, Watt JB, Rivers RP, Milla PJ. Ontogeny of fasting small intestinal motor activity in the human infant. Gut. 1988;29(4):483-8.
117. Riezzo G, Indrio F, Montagna O, Tripaldi C, Laforgia N, Chiloiro M, et al. Gastric electrical activity and gastric emptying in term and preterm newborns. Neurogastroenterol Motil. 2000;12(3):223-9.
118. Lebenthal A, Lebenthal E. The ontogeny of the small intestinal epithelium. JPEN J Parenter Enteral Nutr. 1999;23(5 Suppl):S3-6.
119. Manson WG, Weaver LT. Fat digestion in the neonate. Arch Dis Child Fetal Neonatal Ed. 1997;76(3):F206-11.
120. Rouwet EV, Heineman E, Buurman WA, ter Riet G, Ramsay G, Blanco CE. Intestinal permeability and carrier-mediated monosaccharide absorption in preterm neonates during the early postnatal period. Pediatr Res. 2002;51(1):64-70.
121. Kempley S, Gupta N, Linsell L, Dorling J, McCormick K, Mannix P, et al. Feeding infants below 29 weeks' gestation with abnormal antenatal Doppler: analysis from a randomised trial. Arch Dis Child Fetal Neonatal Ed. 2014;99(1):F6-F11.
122. Choi YY. Necrotizing enterocolitis in newborns: update in pathophysiology and newly emerging therapeutic strategies. Korean J Pediatr. 2014;57(12):505-13.
123. Battersby C, Longford N, Mandalia S, Costeloe K, Modi N, group UKNCNEs. Incidence and enteral feed antecedents of severe neonatal necrotising enterocolitis across neonatal networks in England, 2012-13: a whole-population surveillance study. Lancet Gastroenterol Hepatol. 2017;2(1):43-51.
124. Harding R, Hooper SB. Regulation of lung expansion and lung growth before birth. J Appl Physiol (1985). 1996;81(1):209-24.
125. Fraga MV, Guttentag S. Lung development: embryology, growth, maturation, and developmental biology. Avery's Diseases of the Newborn: Elsevier; 2012. p. 571-83.
126. Henderson-Smart DJ. The effect of gestational age on the incidence and duration of recurrent apnoea in newborn babies. Aust Paediatr J. 1981;17(4):273-6.
127. Eichenwald EC, Aina A, Stark AR. Apnea frequently persists beyond term gestation in infants delivered at 24 to 28 weeks. Pediatrics. 1997;100(3 Pt 1):354-9.
128. Chung HR. Adrenal and thyroid function in the fetus and preterm infant. Korean J Pediatr. 2014;57(10):425-33.
129. Pickerd N, Kotecha S. Pathophysiology of respiratory distress syndrome. Paediatrics and Child Health. 2009;19(4):153-7.
130. Lista G, Castoldi F, Fontana P, Reali R, Reggiani A, Bianchi S, et al. Lung inflammation in preterm infants with respiratory distress syndrome: effects of ventilation with different tidal volumes. Pediatr Pulmonol. 2006;41(4):357-63.
131. Carvalho CG, Silveira RC, Procianoy RS. Ventilator-induced lung injury in preterm infants. Rev Bras Ter Intensiva. 2013;25(4):319-26.
132. Chakraborty M, McGreal EP, Kotecha S. Acute lung injury in preterm newborn infants: mechanisms and management. Paediatr Respir Rev. 2010;11(3):162-70; quiz 70.
133. Egreteau L, Pauchard JY, Semama DS, Matis J, Liska A, Romeo B, et al. Chronic oxygen dependency in infants born at less than 32 weeks' gestation: incidence and risk factors. Pediatrics. 2001;108(2):E26.
134. Jetton JG, Boohaker LJ, Sethi SK, Wazir S, Rohatgi S, Soranno DE, et al. Incidence and outcomes of neonatal acute kidney injury (AWAKEN): a multicentre,

multinational, observational cohort study. Lancet Child Adolesc Health. 2017;1(3):184-94.
135. Stritzke A, Thomas S, Amin H, Fusch C, Lodha A. Renal consequences of preterm birth. Mol Cell Pediatr. 2017;4(1):2.
136. Hinchliffe SA, Sargent PH, Howard CV, Chan YF, van Velzen D. Human intrauterine renal growth expressed in absolute number of glomeruli assessed by the disector method and Cavalieri principle. Lab Invest. 1991;64(6):777-84.
137. Rodriguez MM, Gomez AH, Abitbol CL, Chandar JJ, Duara S, Zilleruelo GE. Histomorphometric analysis of postnatal glomerulogenesis in extremely preterm infants. Pediatr Dev Pathol. 2004;7(1):17-25.
138. Black MJ, Sutherland MR, Gubhaju L, Kent AL, Dahlstrom JE, Moore L. When birth comes early: effects on nephrogenesis. Nephrology (Carlton). 2013;18(3):180-2.
139. Sutherland MR, Gubhaju L, Moore L, Kent AL, Dahlstrom JE, Horne RS, et al. Accelerated maturation and abnormal morphology in the preterm neonatal kidney. J Am Soc Nephrol. 2011;22(7):1365-74.
140. Taeusch HW, Ballard RA, Avery ME, Gleason CA. Avery's Diseases of the Newborn: Saunders Book Company; 2005.
141. Hellstrom A, Smith LE, Dammann O. Retinopathy of prematurity. Lancet. 2013;382(9902):1445-57.
142. Nussbaum C, Gloning A, Pruenster M, Frommhold D, Bierschenk S, Genzel-Boroviczeny O, et al. Neutrophil and endothelial adhesive function during human fetal ontogeny. J Leukoc Biol. 2013;93(2):175-84.
143. Carr R, Huizinga TW. Low soluble FcRIII receptor demonstrates reduced neutrophil reserves in preterm neonates. Arch Dis Child Fetal Neonatal Ed. 2000;83(2):F160.
144. Carr R, Huizinga TW, Kleijer M, Davies JM. Changes in plasma FcRIII demonstrate increasing receptor production during late pregnancy and after preterm birth. Pediatr Res. 1992;32(5):505-8.
145. Kan B, Razzaghian HR, Lavoie PM. An Immunological Perspective on Neonatal Sepsis. Trends Mol Med. 2016;22(4):290-302.
146. Marchant EA, Kan B, Sharma AA, van Zanten A, Kollmann TR, Brant R, et al. Attenuated innate immune defenses in very premature neonates during the neonatal period. Pediatr Res. 2015;78(5):492-7.
147. McGreal EP, Hearne K, Spiller OB. Off to a slow start: under-development of the complement system in term newborns is more substantial following premature birth. Immunobiology. 2012;217(2):176-86.
148. van den Berg JP, Westerbeek EA, van der Klis FR, Berbers GA, van Elburg RM. Transplacental transport of IgG antibodies to preterm infants: a review of the literature. Early Hum Dev. 2011;87(2):67-72.
149. Dineen R, Jaspan T. The neonatal brain. 22011. p. 1253-93.
150. Soul JS, Hammer PE, Tsuji M, Saul JP, Bassan H, Limperopoulos C, et al. Fluctuating pressure-passivity is common in the cerebral circulation of sick premature infants. Pediatr Res. 2007;61(4):467-73.
151. Tsuji M, Saul JP, du Plessis A, Eichenwald E, Sobh J, Crocker R, et al. Cerebral intravascular oxygenation correlates with mean arterial pressure in critically ill premature infants. Pediatrics. 2000;106(4):625-32.
152. Luan-ying T. Incidence and Risk Factors of Brain Injury in Preterm Infants [J]. Journal of Applied Clinical Pediatrics. 2011;2.
153. Volpe JJ. The encephalopathy of prematurity--brain injury and impaired brain development inextricably intertwined. Semin Pediatr Neurol. 2009;16(4):167-78.

154. Jakovcevski I, Filipovic R, Mo Z, Rakic S, Zecevic N. Oligodendrocyte development and the onset of myelination in the human fetal brain. Front Neuroanat. 2009;3:5.
155. Tau GZ, Peterson BS. Normal development of brain circuits. Neuropsychopharmacology. 2010;35(1):147-68.
156. Menassa DA, Gomez-Nicola D. Microglial Dynamics During Human Brain Development. Front Immunol. 2018;9:1014.
157. Gould SJ, Howard S. An immunohistological study of macrophages in the human fetal brain. Neuropathol Appl Neurobiol. 1991;17(5):383-90.
158. Billiards SS, Haynes RL, Folkerth RD, Trachtenberg FL, Liu LG, Volpe JJ, et al. Development of microglia in the cerebral white matter of the human fetus and infant. J Comp Neurol. 2006;497(2):199-208.
159. Merrill JE, Ignarro LJ, Sherman MP, Melinek J, Lane TE. Microglial cell cytotoxicity of oligodendrocytes is mediated through nitric oxide. J Immunol. 1993;151(4):2132-41.
160. Haynes RL, Folkerth RD, Keefe RJ, Sung I, Swzeda LI, Rosenberg PA, et al. Nitrosative and oxidative injury to premyelinating oligodendrocytes in periventricular leukomalacia. J Neuropathol Exp Neurol. 2003;62(5):441-50.
161. Himmelmann K, Uvebrant P. The panorama of cerebral palsy in Sweden. XI. Changing patterns in the birth-year period 2003-2006. Acta Paediatr. 2014;103(6):618-24.
162. Bouyssi-Kobar M, du Plessis AJ, McCarter R, Brossard-Racine M, Murnick J, Tinkleman L, et al. Third Trimester Brain Growth in Preterm Infants Compared With In Utero Healthy Fetuses. Pediatrics. 2016;138(5).
163. Peterson BS, Vohr B, Staib LH, Cannistraci CJ, Dolberg A, Schneider KC, et al. Regional brain volume abnormalities and long-term cognitive outcome in preterm infants. JAMA. 2000;284(15):1939-47.
164. Johnson S. Cognitive and behavioural outcomes following very preterm birth. Semin Fetal Neonatal Med. 2007;12(5):363-73.
165. Snowden JM, Cheng YW, Kontgis CP, Caughey AB. The association between hospital obstetric volume and perinatal outcomes in California. Am J Obstet Gynecol. 2012;207(6):478 e1-7.
166. Pyykonen A, Gissler M, Jakobsson M, Petaja J, Tapper AM. Determining obstetric patient safety indicators: the differences in neonatal outcome measures between different-sized delivery units. BJOG. 2014;121(4):430-7.
167. Joyce R, Webb R, Peacock JL. Associations between perinatal interventions and hospital stillbirth rates and neonatal mortality. Arch Dis Child Fetal Neonatal Ed. 2004;89(1):F51-6.
168. Hamilton KE, Redshaw ME, Tarnow-Mordi W. Nurse staffing in relation to risk-adjusted mortality in neonatal care. Arch Dis Child Fetal Neonatal Ed. 2007;92(2):F99-F103.
169. Rogowski JA, Staiger D, Patrick T, Horbar J, Kenny M, Lake ET. Nurse staffing and NICU infection rates. JAMA Pediatr. 2013;167(5):444-50.
170. Callaghan LA, Cartwright DW, O'Rourke P, Davies MW. Infant to staff ratios and risk of mortality in very low birthweight infants. Arch Dis Child Fetal Neonatal Ed. 2003;88(2):F94-7.
171. Jaques E. The changing culture of a factory. New York: Garland; 1987. xvii, 341 p. p.

172. Mahl S, Lee SK, Baker GR, Cronin CM, Stevens B, Ye XY, et al. The Association of Organizational Culture and Quality Improvement Implementation With Neonatal Outcomes in the NICU. J Pediatr Health Care. 2015;29(5):435-41.
173. Rogowski JA, Horbar JD, Staiger DO, Kenny M, Carpenter J, Geppert J. Indirect vs direct hospital quality indicators for very low-birth-weight infants. JAMA. 2004;291(2):202-9.
174. NICE. Preterm Labour and Birth. National Institute for Health and Care Excellence: Clinical Guidelines. London2015.
175. National Neonatal Audit Programme (NNAP). Annual report. RCPCH; 2018.
176. Ojha S, Gridley E, Dorling J. Use of heated humidified high-flow nasal cannula oxygen in neonates: a UK wide survey. Acta Paediatr. 2013;102(3):249-53.
177. Hagadorn JI, Brownell EA, Trzaski JM, Johnson KR, Lainwala S, Campbell BT, et al. Trends and variation in management and outcomes of very low-birth-weight infants with patent ductus arteriosus. Pediatr Res. 2016;80(6):785-92.
178. Carbajal R, Eriksson M, Courtois E, Boyle E, Avila-Alvarez A, Andersen RD, et al. Sedation and analgesia practices in neonatal intensive care units (EUROPAIN): results from a prospective cohort study. Lancet Respir Med. 2015;3(10):796-812.
179. Dempsey EM, Barrington KJ. Treating hypotension in the preterm infant: when and with what: a critical and systematic review. J Perinatol. 2007;27(8):469-78.
180. Wong J, Shah PS, Yoon EW, Yee W, Lee S, Dow K. Inotrope use among extremely preterm infants in Canadian neonatal intensive care units: variation and outcomes. Am J Perinatol. 2015;32(1):9-14.
181. Liem TB, Krediet TG, Fleer A, Egberts TC, Rademaker CM. Variation in antibiotic use in neonatal intensive care units in the Netherlands. J Antimicrob Chemother. 2010;65(6):1270-5.
182. Kusuda S, Fujimura M, Sakuma I, Aotani H, Kabe K, Itani Y, et al. Morbidity and mortality of infants with very low birth weight in Japan: center variation. Pediatrics. 2006;118(4):e1130-8.
183. Lee SK, McMillan DD, Ohlsson A, Pendray M, Synnes A, Whyte R, et al. Variations in practice and outcomes in the Canadian NICU network: 1996-1997. Pediatrics. 2000;106(5):1070-9.
184. Seppanen AV, Draper ES, Petrou S, Barros H, Aubert AM, Andronis L, et al. High Healthcare Use at Age 5 Years in a European Cohort of Children Born Very Preterm. J Pediatr. 2022;243:69-77 e9.
185. Achana F, Johnson S, Ni Y, Marlow N, Wolke D, Khan K, et al. Economic costs and health utility values associated with extremely preterm birth: Evidence from the EPICure2 cohort study. Paediatr Perinat Epidemiol. 2022.
186. Newnham JP, Schilling C, Petrou S, Morris JM, Wallace EM, Brown K, et al. The health and educational costs of preterm birth to 18 years of age in Australia. Aust N Z J Obstet Gynaecol. 2022;62(1):55-61.
187. Donabedian A. The seven pillars of quality. Arch Pathol Lab Med. 1990;114(11):1115-8.
188. Lohr KN, Schroeder SA. A strategy for quality assurance in Medicare. N Engl J Med. 1990;322(10):707-12.
189. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med. 2000;51(11):1611-25.
190. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. Int J Qual Health Care. 2000;12(4):281-95.
191. Department of Health & Social Care. The NHS Constitution for England. 2015.

192. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q. 1966;44(3):Suppl:166-206.
193. Tawfik DS, Sexton JB, Adair KC, Kaplan HC, Profit J. Context in Quality of Care: Improving Teamwork and Resilience. Clin Perinatol. 2017;44(3):541-52.
194. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. 2013.
195. Kirkup B. The Report of the Morecambe Bay Investigation. 2015.
196. Independent Maternity Review. Ockenden Report: Emerging Findings and Recommendations from the Independent Review of Maternity Services at the Shrewsbury and Telford Hospital NHS Trust. 2020.
197. Chassin MR. Quality of care. Time to act. JAMA. 1991;266(24):3472-3.
198. Liu VX, Fielding-Singh V, Greene JD, Baker JM, Iwashyna TJ, Bhattacharya J, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. Am J Respir Crit Care Med. 2017;196(7):856-63.
199. Fonarow GC, Zhao X, Smith EE, Saver JL, Reeves MJ, Bhatt DL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. JAMA. 2014;311(16):1632-40.
200. Foo CY, Bonsu KO, Nallamothu BK, Reid CM, Dhippayom T, Reidpath DD, et al. Coronary intervention door-to-balloon time and outcomes in ST-elevation myocardial infarction: a meta-analysis. Heart. 2018.
201. Royal College of Paediatrics and Child Health (RCPCH). NNAP Online [Available from: http://www.nnap.rcpch.ac.uk/.
202. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. Int J Qual Health Care. 2003;15 Suppl 1:i5-11.
203. Kotter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators--a systematic review. Implement Sci. 2012;7:21.
204. Rubin HR, Pronovost P, Diette GB. From a process of care to a measure: the development and testing of a quality indicator. Int J Qual Health Care. 2001;13(6):489-96.
205. NHS Improvement. Never Events list 2018. 2018.
206. Profit J, Gould JB, Zupancic JA, Stark AR, Wall KM, Kowalkowski MA, et al. Formal selection of measures for a composite index of NICU quality of care: Baby-MONITOR. J Perinatol. 2011;31(11):702-10.
207. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care. 1986;2(1):53-63.
208. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. N Engl J Med. 1998;338(26):1888-95.
209. Kowalkowski M, Gould JB, Bose C, Petersen LA, Profit J. Do practicing clinicians agree with expert ratings of neonatal intensive care unit quality measures? J Perinatol. 2012;32(4):247-52.
210. Reiter A, Fischer B, Kotting J, Geraedts M, Jackel W, Barlag H, et al. QUALIFY: Instrument for the Assessment of Quality Indicators. Dusseldorf: Bundes Geschafts Stelle Qualitats Sicherung. 2007.
211. National Quality Forum (NQF). Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement 2018 [Available from: https://www.qualityforum.org/measuring_performance/submitting_standards/measure_evaluation_criteria.aspx.

212. Altilio T, Otis-Green S, Dahlin CM. Applying the National Quality Forum Preferred Practices for Palliative and Hospice Care: a social work perspective. J Soc Work End Life Palliat Care. 2008;4(1):3-16.
213. Wilke LG, Ballman KV, McCall LM, Giuliano AE, Whitworth PW, Blumencranz PW, et al. Adherence to the National Quality Forum (NQF) breast cancer measures within cancer clinical trials: a review from ACOSOG Z0010. Ann Surg Oncol. 2010;17(8):1989-94.
214. Backus LI, Boothroyd DB, Phillips BR, Belperio PS, Halloran JP, Valdiserri RO, et al. National quality forum performance measures for HIV/AIDS care: the Department of Veterans Affairs' experience. Arch Intern Med. 2010;170(14):1239-46.
215. Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM. Expert consensus on the desirable characteristics of review criteria for improvement of health care quality. Qual Health Care. 2001;10(3):173-8.
216. Kelley E, Hurst J. Health care quality indicators project conceptual framework paper2006.
217. Little TD, Jorgensen TD, Lang KM, Moore EW. On the joys of missing data. J Pediatr Psychol. 2014;39(2):151-62.
218. Rosenthal GE. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. Am J Public Health. 1997;87(3):429-33.
219. Profit J, Zupancic JA, Gould JB, Pietz K, Kowalkowski MA, Draper D, et al. Correlation of neonatal intensive care unit performance across multiple measures of quality of care. JAMA Pediatr. 2013;167(1):47-54.
220. Profit J, Typpo KV, Hysong SJ, Woodard LD, Kallen MA, Petersen LA. Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. Implement Sci. 2010;5:13.
221. Barclay M, Dixon-Woods M, Lyratzopoulos G. The problem with composite indicators. BMJ Qual Saf. 2019;28(4):338-44.
222. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. JAMA. 2006;295(10):1168-70.
223. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med Care. 2007;45(6):489-96.
224. Staffordshire Shropshire and Black Country Neonatal Operational Delivery Network. Neonatal Guidelines 2017-2019 [Available from: https://www.networks.nhs.uk/nhs-networks/staffordshire-shropshire-and-black-country-newborn/neonatal-guidelines.
225. Elliott R, Timulak L. Descriptive and interpretive approaches to qualitative research. A handbook of research methods for clinical and health psychology. 2005;1(7):147-59.
226. Hill CE, Thompson BJ, Williams EN. A guide to conducting consensual qualitative research. The counseling psychologist. 1997;25(4):517-72.
227. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. BMJ. 2004;328(7451):1312-5.
228. Heberlein TA, Baumgartner R. Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. American sociological review. 1978:447-62.
229. Goyder JC. Further evidence on factors affecting response rates to mailed questionnaires. American Sociological Review. 1982:550-3.

230. Hartge P. Raising response rates: getting to yes. Epidemiology. 1999;10(2):105-7.
231. Battersby C, Statnikov Y, Santhakumaran S, Gray D, Modi N, Costeloe K, et al. The United Kingdom National Neonatal Research Database: A validation study. PLoS One. 2018;13(8):e0201815.
232. Costeloe K, Bowler U, Brocklehurst P, Hardy P, Heal P, Juszczak E, et al. A randomised controlled trial of the probiotic Bifidobacterium breve BBG-001 in preterm babies to prevent sepsis, necrotising enterocolitis and death: the Probiotics in Preterm infantS (PiPS) trial. Health Technol Assess. 2016;20(66):1-194.
233. RCPCH. National Neonatal Audit Programme [Available from: https://www.rcpch.ac.uk/work-we-do/quality-improvement-patient-safety/national-neonatal-audit-programme.
234. NDAU. Neonatal-dataset-ISB1595-release-1-version-22 [Available from: https://www.imperial.ac.uk/media/imperial-college/medicine/dept-medicine/infectious-diseases/neonatology/Neonatal-dataset-ISB1595-release-1-version-22.pdf.
235. British Association of Perinatal Medicine (BAPM). Antenatal Optimisation for Preterm Infants less than 34 weeks: A Quality Improvement Toolkit. 2020.
236. Vyas J, Kotecha S. Effects of antenatal and postnatal corticosteroids on the preterm lung. Arch Dis Child Fetal Neonatal Ed. 1997;77(2):F147-50.
237. Schwab M, Roedel M, Anwar MA, Muller T, Schubert H, Buchwalder LF, et al. Effects of betamethasone administration to the fetal sheep in late gestation on fetal cerebral blood flow. J Physiol. 2000;528(Pt 3):619-32.
238. Roberts D, Brown J, Medley N, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database Syst Rev. 2017;3:CD004454.
239. Katheria AC, Lakshminrusimha S, Rabe H, McAdams R, Mercer JS. Placental transfusion: a review. J Perinatol. 2017;37(2):105-11.
240. McDonald SJ, Middleton P, Dowswell T, Morris PS. Effect of timing of umbilical cord clamping of term infants on maternal and neonatal outcomes. Cochrane Database Syst Rev. 2013(7):CD004074.
241. Committee on Obstetric P. Committee Opinion No. 684: Delayed Umbilical Cord Clamping After Birth. Obstet Gynecol. 2017;129(1):e5-e10.
242. Sommers R, Stonestreet BS, Oh W, Laptook A, Yanowitz TD, Raker C, et al. Hemodynamic effects of delayed cord clamping in premature infants. Pediatrics. 2012;129(3):e667-72.
243. Mercer JS, Erickson-Owens DA, Deoni SCL, Dean DC, 3rd, Collins J, Parker AB, et al. Effects of Delayed Cord Clamping on 4-Month Ferritin Levels, Brain Myelin Content, and Neurodevelopment: A Randomized Controlled Trial. J Pediatr. 2018;203:266-72 e2.
244. Fogarty M, Osborn DA, Askie L, Seidler AL, Hunter K, Lui K, et al. Delayed vs early umbilical cord clamping for preterm infants: a systematic review and meta-analysis. Am J Obstet Gynecol. 2018;218(1):1-18.
245. Rabe H, Diaz-Rossello JL, Duley L, Dowswell T. Effect of timing of umbilical cord clamping and other strategies to influence placental transfusion at preterm birth on maternal and infant outcomes. Cochrane Database Syst Rev. 2012(8):CD003248.
246. Rabe H, Reynolds G, Diaz-Rossello J. A systematic review and meta-analysis of a brief delay in clamping the umbilical cord of preterm infants. Neonatology. 2008;93(2):138-44.

247. NICE. Intrapartum Care: Care of Healthy Women and Their Babies During Childbirth. National Institute for Health and Care Excellence: Clinical Guidelines. London2014.
248. Resuscitation Council (UK). Resuscitation and support of transition of babies at birth 2015 [Available from: https://www.resus.org.uk/resuscitation-guidelines/resuscitation-and-support-of-transition-of-babies-at-birth/.
249. Dahm LS, James LS. Newborn temperature and calculated heat loss in the delivery room. Pediatrics. 1972;49(4):504-13.
250. Malin SW, Baumgart S. Optimal thermal management for low birth weight infants nursed under high-powered radiant warmers. Pediatrics. 1987;79(1):47-54.
251. Mance MJ. Keeping infants warm: challenges of hypothermia. Adv Neonatal Care. 2008;8(1):6-12.
252. Gandy GM, Adamsons K, Jr., Cunningham N, Silverman WA, James LS. Thermal Environment and Acid-Base Homeostasis in Human Infants during the First Few Hours of Life. J Clin Invest. 1964;43:751-8.
253. Benumof JL, Wahrenbrock EA. Dependency of hypoxic pulmonary vasoconstriction on temperature. J Appl Physiol Respir Environ Exerc Physiol. 1977;42(1):56-8.
254. de Almeida MF, Guinsburg R, Sancho GA, Rosa IR, Lamy ZC, Martinez FE, et al. Hypothermia and early neonatal mortality in preterm infants. J Pediatr. 2014;164(2):271-5 e1.
255. Laptook AR, Bell EF, Shankaran S, Boghossian NS, Wyckoff MH, Kandefer S, et al. Admission Temperature and Associated Mortality and Morbidity among Moderately and Extremely Preterm Infants. J Pediatr. 2018;192:53-9 e2.
256. Lyu Y, Shah PS, Ye XY, Warre R, Piedboeuf B, Deshpandey A, et al. Association between admission temperature and mortality and major morbidity in preterm infants born at fewer than 33 weeks' gestation. JAMA Pediatr. 2015;169(4):e150277.
257. Berseth CL, Nordyke C. Enteral nutrients promote postnatal maturation of intestinal motor activity in preterm infants. Am J Physiol. 1993;264(6 Pt 1):G1046-51.
258. Meetze WH, Valentine C, McGuigan JE, Conlon M, Sacks N, Neu J. Gastrointestinal priming prior to full enteral nutrition in very low birth weight infants. J Pediatr Gastroenterol Nutr. 1992;15(2):163-70.
259. Fernandez L, Langa S, Martin V, Maldonado A, Jimenez E, Martin R, et al. The human milk microbiota: origin and potential roles in health and disease. Pharmacol Res. 2013;69(1):1-10.
260. Leaf A, Dorling J, Kempley S, McCormick K, Mannix P, Linsell L, et al. Early or delayed enteral feeding for preterm growth-restricted infants: a randomized trial. Pediatrics. 2012;129(5):e1260-8.
261. Abbott J, Berrington J, Bowler U, Boyle E, Dorling J, Embleton N, et al. The Speed of Increasing milk Feeds: a randomised controlled trial. BMC Pediatr. 2017;17(1):39.
262. Subramaniam P, Ho JJ, Davis PG. Prophylactic nasal continuous positive airway pressure for preventing morbidity and mortality in very preterm infants. Cochrane Database Syst Rev. 2016(6):CD001243.
263. National Neonatal Audit Programme (NNAP). Neonatal Unit and Neonatal Network Annual Summary Reports 2017 [Available from: https://nnap.rcpch.ac.uk/annual-reports.aspx.
264. Royal College of Paediatrics and Child Health (RCPCH). National Neonatal Audit Programme (NNAP): A guide to the 2018 audit measures. 2017.

265. Wheelan SA, Burchill CN, Tilin F. The link between teamwork and patients' outcomes in intensive care units. Am J Crit Care. 2003;12(6):527-34.
266. Huang DT, Clermont G, Kong L, Weissfeld LA, Sexton JB, Rowan KM, et al. Intensive care unit safety culture and outcomes: a US multicenter study. Int J Qual Health Care. 2010;22(3):151-61.
267. Roch G, Dubois CA, Clarke SP. Organizational climate and hospital nurses' caring practices: a mixed-methods study. Res Nurs Health. 2014;37(3):229-40.
268. Brazil K, Wakefield DB, Cloutier MM, Tennen H, Hall CB. Organizational culture predicts job satisfaction and perceived clinical effectiveness in pediatric primary care practices. Health Care Manage Rev. 2010;35(4):365-71.
269. Pollack MM, Koch MA, Network NI-DoCN. Association of outcomes with organizational characteristics of neonatal intensive care units. Crit Care Med. 2003;31(6):1620-9.
270. Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. Pediatrics. 1993;91(3):617-23.
271. Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. J Pediatr. 2001;138(1):92-100.
272. Yates M, Bechman K, Dennison EM, MacGregor AJ, Ledingham J, Norton S, et al. Data quality predicts care quality: findings from a national clinical audit. Arthritis Res Ther. 2020;22(1):87.
273. Mathioudakis A, Rousalova I, Gagnat AA, Saad N, Hardavella G. How to keep good clinical records. Breathe (Sheff). 2016;12(4):369-73.
274. Bland JM, Kerry SM. Statistics notes. Weighted comparison of means. BMJ. 1998;316(7125):129.
275. MBRRACE-UK. MBRRACE-UK Perinatal Confidential Enquiry: Stillbirths and Neonatal Deaths in Twin Pregnancies. 2021.
276. Smith LK, Draper ES, Manktelow BN, Dorling JS, Field DJ. Socioeconomic inequalities in very preterm birth rates. Arch Dis Child Fetal Neonatal Ed. 2007;92(1):F11-4.
277. Taylor-Robinson D, Agarwal U, Diggle PJ, Platt MJ, Yoxall B, Alfirevic Z. Quantifying the impact of deprivation on preterm births: a retrospective cohort study. PLoS One. 2011;6(8):e23163.
278. Bauer J, Groneberg DA, Maier W, Manek R, Louwen F, Bruggmann D. Accessibility of general and specialized obstetric care providers in Germany and England: an analysis of location and neonatal outcome. Int J Health Geogr. 2017;16(1):44.
279. Ministry of Housing CLG. The English Indices of Deprivation 2019 (IoD2019) 2019.
280. Williams OA. These Are The U.K. Towns And Cities Where The Wealthiest Live. Forbes. 2019.
281. Osborne H. Wealthy homebuyers prefer to live in cities. The Guardian. 2015.
282. Williams J. An investigation into the effect of a plastic wrap on the temperature regulation of the very low birth weight and premature infant during transfer to the neonatal intensive care unit: a systematic review. Health Care Reports. 2004;2(3):53-77.
283. Cole A. Neonatal care: time for action on the shortage of nurses. J Fam Health Care. 2009;19(2):42-4.
284. Pillay T, Nightingale P, Owen S, Kirby D, Spencer A. Neonatal nurse staffing and delivery of clinical care in the SSBC Newborn Network. Arch Dis Child Fetal Neonatal Ed. 2012;97(3):F174-8.

285. Royal College of Paediatrics and Child Health (RCPCH). State of Child Health (Short report series: The Paediatric Workforce)2017. Available from: https://www.rcpch.ac.uk/sites/default/files/2018-03/2015_rcpch_state_of_child_health_the_paediatric_workforce_v1.1.pdf.
286. Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. In: Behrman R, Butler, AS., editor. Preterm Birth - Causes, Consequences, and Prevention. Washington (DC): National Academies Press (US),; 2007.
287. Parish A, Bhatia J. Feeding strategies in the ELBW infant. J Perinatol. 2008;28 Suppl 1:S18-20.
288. Weightman AL, Morgan HE, Shepherd MA, Kitcher H, Roberts C, Dunstan FD. Social inequality and infant health in the UK: systematic review and meta-analyses. BMJ Open. 2012;2(3).
289. O'Dowd A. Poverty status is linked to worse quality of care. BMJ. 2020;368:m303.
290. Nuffield Trust and the Health Foundation. How have inequalities in the quality of care changed over the last 10 years? ; 2020.
291. GIRFT; RCPCH. A snapshot of neonatal services and workforce in the UK. 2020.
292. Seaton SE, Barker L, Draper ES, Abrams KR, Modi N, Manktelow BN, et al. Estimating neonatal length of stay for babies born very preterm. Arch Dis Child Fetal Neonatal Ed. 2019;104(2):F182-F6.
293. MBRRACE-UK. MBRRACE-UK Perinatal Mortality Surveillance Report for births in 2019. 2021.
294. Maier RF, Blondel B, Piedvache A, Misselwitz B, Petrou S, Van Reempts P, et al. Duration and Time Trends in Hospital Stay for Very Preterm Infants Differ Across European Regions. Pediatr Crit Care Med. 2018;19(12):1153-61.
295. NHS England. the national schedules of NHS costs. 2019/20.
296. Lee SK, Beltempo M, McMillan DD, Seshia M, Singhal N, Dow K, et al. Outcomes and care practices for preterm infants born at less than 33 weeks' gestation: a quality-improvement study. CMAJ. 2020;192(4):E81-E91.
297. Lui K, Lee SK, Kusuda S, Adams M, Vento M, Reichman B, et al. Trends in Outcomes for Neonates Born Very Preterm and Very Low Birth Weight in 11 High-Income Countries. J Pediatr. 2019;215:32-40 e14.
298. Kaempf JW, Wang L, Dunn M. Using a composite morbidity score and cultural survey to explore characteristics of high proficiency neonatal intensive care units. Arch Dis Child Fetal Neonatal Ed. 2019;104(1):F13-F7.
299. Zeitlin J, Manktelow BN, Piedvache A, Cuttini M, Boyle E, van Heijst A, et al. Use of evidence based practices to improve survival without severe morbidity for very preterm infants: results from the EPICE population based cohort. BMJ. 2016;354:i2976.
300. Ismail AQT, Boyle EM, Pillay T, Modi N, Rivero-Arias O, Manktelow B, et al. Clinical outcomes for babies born between 27–31 weeks of gestation: Should they be regarded as a single cohort? Journal of Neonatal Nursing. 2022.
301. Ismail AT, Boyle EM, Oddie S, Pillay T. Exploring variation in quality of care and clinical outcomes between neonatal units: a novel use for the UK National Neonatal Audit Programme (NNAP). BMJ Open Qual. 2022;11(4).
302. Shane AL, Stoll BJ. Neonatal sepsis: progress towards improved outcomes. J Infect. 2014;68 Suppl 1:S24-32.
303. Dong Y, Speer CP. Late-onset neonatal sepsis: recent developments. Arch Dis Child Fetal Neonatal Ed. 2015;100(3):F257-63.

304. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. BMC Med Res Methodol. 2012;12:21.
305. NHS. National Patient And Staff Survey [Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/patient-surveys/.
306. Staniszewska S, Brett J, Redshaw M, Hamilton K, Newburn M, Jones N, et al. The POPPY study: developing a model of family-centred care for neonatal units. Worldviews Evid Based Nurs. 2012;9(4):243-55.
307. Burger S, King, J.,. Parents' experiences of neonatal care in England. Patient Experience Journal. 2015;2:45-52.
308. Profit J, Sharek PJ, Cui X, Nisbet CC, Thomas EJ, Tawfik DS, et al. The Correlation Between Neonatal Intensive Care Unit Safety Culture and Quality of Care. J Patient Saf. 2020;16(4):e310-e6.
309. Lake ET, Smith JG, Staiger DO, Hatfield LA, Cramer E, Kalisch BJ, et al. Parent Satisfaction With Care and Treatment Relates to Missed Nursing Care in Neonatal Intensive Care Units. Front Pediatr. 2020;8:74.
310. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. Bull World Health Organ. 2017;95(5):368-74.
311. Tawfik DS, Scheid A, Profit J, Shanafelt T, Trockel M, Adair KC, et al. Evidence Relating Health Care Provider Burnout and Quality of Care: A Systematic Review and Meta-analysis. Ann Intern Med. 2019;171(8):555-67.
312. Humphries N, Morgan K, Conry MC, McGowan Y, Montgomery A, McGee H. Quality of care and health professional burnout: narrative literature review. Int J Health Care Qual Assur. 2014;27(4):293-307.
313. Koch P, Zilezinski M, Schulte K, Strametz R, Nienhaus A, Raspe M. How Perceived Quality of Care and Job Satisfaction Are Associated with Intention to Leave the Profession in Young Nurses and Physicians. Int J Environ Res Public Health. 2020;17(8).
314. British Association of Perinatal Medicine (BAPM). Antenatal Optimisation for Preterm Infants less than 34 weeks. 2020.
315. Ohlsson A SP. Determinants and Prevention of Low Birth Weight: A Synopsis of the Evidence2008.
316. Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic disparities in adverse birth outcomes: a systematic review. Am J Prev Med. 2010;39(3):263-72.
317. Garner S, Bhattacharyya G. Place, ethnicity and poverty in England'. Joseph Rowntree Foundation; 2011.
318. Statistics OfN. Pregnancy and ethnic factors influencing births and infant mortality: 2013. ONS Statistical Bulletin; 2015.
319. Sigurdson K, Mitchell B, Liu J, Morton C, Gould JB, Lee HC, et al. Racial/Ethnic Disparities in Neonatal Intensive Care: A Systematic Review. Pediatrics. 2019;144(2).
320. Bliss. Bliss Baby Report 2015: Hanging in the Balance. Bliss London; 2015.
321. Buchan J, Ball J, Shembavnekar N, Charlesworth A. Building the NHS nursing workforce in England. The Health Foundation. 2020.
322. Murray R. The King's Fund,. 2017. [cited July 2021]. Available from: https://www.kingsfund.org.uk/blog/2017/10/falling-number-nurses-nhs-paints-worrying-picture?gclid=CjwKCAjw3MSHBhB3EiwAxcaEuyB6qdBLro6jV-B9bfm3RHj01KHCY0HF8u0yes44nTiQSyS45SS7LhoCEz0QAvD_BwE.
323. Imison C. Nuffield Trust, editor2016. [cited July 2021]. Available from: https://www.nuffieldtrust.org.uk/news-item/can-we-solve-the-nursing-workforce-crisis.

324. Royal College of Paediatrics and Child Health (RCPCH). National Neonatal Audit Programme A guide to the 2020 audit measures. 2019.
325. Royal college of Paediatrics and Child Health (RCPCH). National Neonatal Audit Programme 2021 Measures Guide. 2020.
326. Tarnow-Mordi W, Parry G. The CRIB score. Lancet. 1993;342(8883):1365.
327. Boyle EM, Poulsen G, Field DJ, Kurinczuk JJ, Wolke D, Alfirevic Z, et al. Effects of gestational age at birth on health outcomes at 3 and 5 years of age: population based cohort study. BMJ. 2012;344:e896.
328. Johnston KM, Gooch K, Korol E, Vo P, Eyawo O, Bradt P, et al. The economic burden of prematurity in Canada. BMC Pediatr. 2014;14:93.
329. Mangham LJ, Petrou S, Doyle LW, Draper ES, Marlow N. The cost of preterm birth throughout childhood in England and Wales. Pediatrics. 2009;123(2):e312-27.
330. Petrou S. Economic consequences of preterm birth and low birthweight. BJOG. 2003;110 Suppl 20:17-23.