The development and design of a highly discerning platform for data capture in a neonatal encephalopathy study

By

Thumbiko Clover Kalua

This dissertation is presented in a publication format

and is submitted in fulfilment of the requirements for:

Master of Science

in

Medical Immunology

Department of Immunology

Faculty of Health Sciences

University of Pretoria

August 2024

Declaration

I declare that the work contained in this dissertation is my own original work and has not been submitted previously, in whole or in part, in respect of any academic award to any other institution. All references to other works have been duly acknowledged and cited, and this work complies with the ethical standards of research. This work is submitted in fulfilment of the requirements for a Master of Science degree at the University of Pretoria.

Signed:  Date: August 2024

Acknowledgements

Firstly, I thank my heavenly Father for providing me with the strength, tenacity, and encouragement to keep going, especially on my darkest days.

I would like to extend my heartfelt thanks to my supervisors Professor Michael Pepper, Professor Fourie Joubert and Miss Jeanne van Rensburg, for their support, guidance, and encouragement. Your insightful feedback, patience, and expertise have been vital in shaping this work, and I am truly grateful for your mentorship.

I would also like to acknowledge the NESHIE study for providing the essential data and framework that formed the foundation of this research. My sincere appreciation goes to the Bill & Melinda Gates Foundation (BMGF) and the South African Medical Research Council (SAMRC) for their generous funding and support.

I am also deeply grateful to the students at the Institute for Cellular and Molecular Medicine (ICMM) who provided support, friendship, and invaluable advice throughout this journey. The sense of community at the ICMM has been a source of constant motivation and inspiration.

Finally, I would like to express my deepest gratitude to my family for their unwavering love, patience, and encouragement. To my parents, Rumbani Kalua and the late Professor Fetson Anderson Kalua, who fostered a love for knowledge and learning from a young age, your belief in my potential has been my driving force. Your enduring support and inspiration have been pivotal in my academic journey. To my sisters, Tai Kalua and Temmi Kalua, and my cousin, Towera Gondwe, your constant support, understanding, and encouragement have been invaluable. Your companionship and unwavering faith in me have provided the strength I needed to persevere.

To my friends, thank you for being my pillars of support. Your encouragement and understanding have been crucial in helping me navigate the challenges of this journey.

Thank you all for being a part of this journey.

Summary

NESHIE is a clinical condition defined by a restricted supply of oxygen and blood flow around the time of labour and delivery. Some neonates diagnosed with moderate and severe cases die in the infantile stage while surviving neonates may develop neurodevelopmental disorders, with cerebral palsy (CP) being one of the most adverse outcomes. For unknown reasons, South Africa experiences a 2-3-fold higher incidence than the global average, which prompted the launch of the NESHIE study. The NESHIE study aims to ascertain if there is a genetic predisposition to death and disability following moderate and severe NESHIE in newborns in several public institutions in South Africa. To achieve these aims, biomarker discovery, clinical data and samples are collected from enrolled patients across the various hospitals involved.

This research project aims to design and develop a national-level database capable of capturing and storing multiple data points from several sites. In addition, it will evaluate the data capture process to refine data quality assessments and enhance overall data accuracy, ensuring the database meets the needs of diverse users across all study sites.

The management of data in the NESHIE study has used an electronic data capture (EDC) software known as REDCap. Various features in REDCap were leveraged to design improved versions of the NESHIE REDCap database, resulting in a more user-friendly interface as well as a reduction in errors. Statistical analyses provided support and pinpointed sections requiring adjustments of the database and its supporting data. Furthermore, quantitative analyses focused on sub-sections of data requiring quality control. Trends in the data were explored to characterise observed errors, compare error rates between database iterations, and identify outliers.

The database design process carried out in REDCap demonstrated the ability to implement diverse design features for clinical research. These trends in the data revealed error occurrences and their sources. A chi-square analysis confirmed a reduction in errors in the refined A7 database. Lastly, the exploration of outliers provided valuable insights into error incidents by highlighting variables containing extreme observations.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A3	Amendment version 3
A4	Amendment version 4
A5	Amendment version 5
A7	Amendment version 7
aEEG	Amplitude integrated electroencephalogram
APGAR	Appearance, Pulse, Grimace, Activity, and Respiration.
ASSAf	Academy of Sciences of South Africa
ATP	Adenosine triphosphate
BD	Base Deficit
BMI	Body Mass Index
BMV	Bag and Mask Ventilation
BP	Blood Pressure
CFM	Cerebral Function Monitor
COH	Circumference of Head
COVID-19	Coronavirus Disease 2019
CP	Cerebral palsy
CPAP	Continuous Positive Airway Pressure
CRFs	Case Report Forms
CRIMS	Chinese National HIV/AIDS Data Repository
CSV	Comma-Separated Values
DAGs	Data access groups
eCRFs	electronic case report forms
EDC	Electronic Data Capture

EEG	Electroencephalogram
EHR	Electronic health record
EMA	European Medical Agency
EUS	Endoscopic Ultrasound
EVD	Ebola Virus Disease
FAQ	Frequently Asked Questions
FDA	Food and Drug Administration
FISMA	Federal Information Security Management Act
GA	Gestational Age
GCP	Good Clinical Practice
GUI	Graphical user interface
HELIX Trial	Hypothermia for Encephalopathy in Low- and Middle-Income Countries Trial
HIC	High-income countries
HIE	Hypoxic ischaemic encephalopathy
HIPAA	Health Insurance Probability and Accountability
HITECH Act	Health Information Technology for Economic and Clinical Health Act
HIV/AIDS	Human Immunodeficiency Virus/ Acquired Immunodeficiency Syndrome
HREC	Human Research Ethics Committee
HTML	Hypertext Markup Language
ICH-GCP	International Conference on Harmonisation Good Clinical Practice
ID	Identification
IPPV	Intermittent Positive Pressure Ventilation

IQR	Interquartile Range
IT	Information Technology
LMICs	Low- and Middle-Income Countries
NA	Not Applicable
NE	Neonatal encephalopathy
NESHIE	Neonatal encephalopathy with suspected hypoxic ischaemic encephalopathy
NICU	Neonatal Intensive Care Unit
PDF	Portable Document Format
pH	Potential of Hydrogen
PHI	Protected Health Information
PI	Principal Investigator
POC	Point of Care
POPI Act	Protection of Personal Information Act
PPHN	Persistent Pulmonary Hypertension of the Newborn
QC	Quality control
REDCap	Research Electronic Data Capture
ROS	Reactive Oxygen Species
SA	South Africa
SAS	Statistical Analysis System
SDV	Source document verification
SOP	Standard Operating Procedures
SPSS	Statistical Package for the Social Sciences
SSL	Secure Sockets Layer

SU	Stellenbosch University
TH	Therapeutic hypothermia
TN	Tennessee
UCT	University of Cape Town
UKZN	University of KwaZulu-Natal
UP	University of Pretoria
WHO	World Health Organisation

Chapter 1: Introduction

1.1 Background

Neonatal encephalopathy (NE) is a clinical condition characterised by disturbed neurologic function in infants born at or after 35 weeks of gestation [1]. It manifests as a subnormal level of consciousness and/or seizures often accompanied by reduced muscle tone and reflexes, and difficulties in initiating and maintaining respiration [2]. The American College of Obstetrics and Gynecology (ACOG) defines various terms related to NE such as asphyxia and hypoxic ischaemic encephalopathy (HIE). HIE is a subtype of NE that occurs as a consequence of a limited supply of oxygen and blood supply occurring in close temporal proximity to labour and delivery [1,2].

While HIE is estimated to affect 1-6 infants per 1 000 live births in high-income countries (HIC), in low-or middle-income countries(LMIC), this incidence may increase to approximately 20 infants per 1 000 live births and accounts for an estimated one million infant deaths annually [3]. Close to 40% of newborns with HIE do not survive the neonatal period (birth to one month), and around 30% develop neurological disorders [4]. In countries with limited resources, the incidence of HIE has been documented to rise to 26 cases per 1 000 term newborns, with even fewer survivors. HIE is directly responsible for 23-25 % of neonatal deaths globally [4] and is a global health concern, imposing a substantial burden on society at large. However, a lack of a universally accepted definition of HIE has contributed to degrees of variability in the reported incidence rates within local hospitals, and on national, regional and international levels [4]. This lack of consensus prompted Kurinczuk et al. (2010) to emphasise the cautious use of the term 'HIE', as it implies a clear understanding of the root cause of neurologic dysfunction, accompanied by substantial evidence of intrapartum hypoxia-ischaemia [1,2].

The NESHIE study is novel as it seeks to investigate the factors behind these statistics. This multi-institutional study involves seven hospitals and a team of various experts, with extensive data, sample collection and analysis. The study is dedicated to investigating potential genetic predispositions for mortality and disability in newborns who experience Sarnat grade 2 and 3 NESHIE in South Africa's public institutions. To effectively manage the extensive clinical data generated by the study, REDCap (Research Electronic Data Capture)

was employed as the data collection system. REDCap is a data management platform that offers online support and data-capturing capacity for research studies[6].

1.2 Purpose of the project and Research Problem

The understanding of NESHIE pathogenesis, both in South Africa and globally remains incomplete. Consequently, the NESHIE study has the potential to increase the knowledge base in this field, locally and on a global scale.

1.3 Research Problem

While this project forms part of the broader NESHIE study, its primary focus is the design and development of a database capable of accommodating the data management needs of the study. This database will be able to capture multiple data points across several national sites and enhance data accuracy by refining the data quality evaluation. The meticulous data capturing method, combined with robust monitoring, will ensure thorough evaluation and identification of data errors.

The objectives of this project will be to establish well-defined parameters for study collaborators, streamline the data capture process, and optimise the user-friendly interface for database users. This, in turn, will facilitate data entry, reduce errors, and promote data quality checks by the investigators. Additionally, advanced statistical analyses will be applied to further refine and support the data collection system of the database.

1.4 Hypothesis

While this project may not fit within the traditional framework of a testable(falsifiable) hypothesis, it nonetheless involves testing specific assumptions related to database performance. Thus, we propose that these assumptions are tested from the following research objectives:

- First, we wish to demonstrate whether a responsive and appropriately modified database design will yield a user-friendly interface, resulting in a measurable error reduction in user-generated errors.
- Second, we wish to determine if mean error rates will be significantly lower in the modified database compared to the initial database.

This dissertation includes a literature review on the pathophysiology of neonatal encephalopathy, data collection and management systems, database design principles, and data quality. The database setup chapter covers requirements gathering, schema development, and implementing REDCap features to meet project needs. The error analysis chapter (Analysis I) categorises compliance issues, invalid data, and deviations from protocol within all three iterations of the NESHIE project database. The outlier detection section (Analysis II) defines outliers, examines techniques for their identification, and explores their impact on data quality and management strategies. Finally, the discussion summarises the findings, highlights the implications of the work, and offers recommendations for future research while addressing any limitations encountered during the study.

Chapter 2: Literature Review

2.1 Pathophysiology of Neonatal Encephalopathy with Suspected Hypoxic-Ischaemic Encephalopathy

Sub-Saharan Africa has the highest mortality rate for children under 5 years of age when compared to other regions that have implemented Sustainable Development Goals. According to the World Health Organisation (WHO), 27 deaths per 1,000 live births were recorded in 2019 in sub-Saharan Africa followed by central and southern Asia with 24 deaths per 1,000 live births [1]; a child born in sub-Saharan Africa or southern Asia is ten times more likely to die in the first month of life than a child born in a high-income nation. Approximately 75% of all neonatal deaths ensue within the first week of life with a million newborn deaths occurring within the first 24 hours [1]. Most neonatal deaths in 2017 were attributed to factors that include preterm birth, intrauterine exposure to infections, birth defects and intrapartum-related complications such as perinatal asphyxia.

Perinatal asphyxia, also known as birth asphyxia or neonatal asphyxia, is defined as a lack of blood flow and/or an impairment of gaseous exchange that has a negative impact on the foetus or neonate. Often resulting in hypoxaemia and hypercapnia complemented by metabolic acidosis and subsequent multi-organ failure, the clinical definition is presented as an inability of a newborn to initiate or maintain regular breathing at birth [2, 3].

Perinatal asphyxia leads to about 4 million deaths annually around the globe, as reported by the WHO in 2014 [4]; a more recent statistic does not appear to be available. In most high-

income countries, the frequency of perinatal asphyxia constitutes less than 0.1% of newborn deaths. However, it can range from 4.6 per 1 000 to 26 per 1 000 live births in low-middle-income countries [2, 3]. More than 25% of the world's newborn deaths have occurred in Africa, where perinatal asphyxia accounts for 24% of these deaths. Seventy five percent of the twenty countries in the world with the highest risk of neonatal deaths are found in Africa[1]. Furthermore, perinatal asphyxia, infections, and complications of preterm birth collectively account for 88% of newborn deaths in Africa[1]. In sub-Saharan Africa, 280 000 newborn deaths in the first day of life were attributed to perinatal asphyxia indicating an occurrence of asphyxia in east, central, and southern Africa of 22% [5]. Perinatal asphyxia has been noted in 80 of 1 000 live births worldwide; however, it is estimated that only 10 of 1 000 babies will go on to develop encephalopathy [2] .

The term neonatal encephalopathy (NE) is a clinical description of disrupted neurological function in newborns, which is not limited to a specific aetiology or mechanism of pathogenesis. It is used to describe the clinical condition of infant brain function for which there exists multiple causes, risk factors and treatment [6].

NE has been estimated to affect 1.15 million neonates around the world each year, of which 478 000 are found in sub-Saharan Africa. It is the third most prevalent cause of death in children under the age of five years and a weighty contributor to long-term neurological morbidities seen worldwide [7]. NE is characterised by decreased levels of consciousness, seizures, respiratory insufficiency and depression of tone and reflexes [6].

Determining poor outcomes in neonates with neonatal encephalopathy with suspected hypoxic ischaemic encephalopathy (NESHIE) around the time of birth is difficult; however, it is crucial to have a way of identifying babies at high risk [2]. Stage 1 encephalopathy is the mildest degree of disease in infants accompanied by hypertonia and poor feeding while more severely affected neonates with Stage 2 encephalopathy have reduced tone and seizures associated with poor outcome in 25% of cases. Stage 3 encephalopathy is observed in the most severely affected infants, which is characterised by profound stupor or coma as well as a severely suppressed amplitude integrated electroencephalography (aEEG). Death and/or a severely abnormal outcome is observed in more than 75% of stage 3 encephalopathy neonates [8].

However, even in moderate neonatal encephalopathy where there is a lower risk of cerebral palsy (CP), cognitive defects that manifest only when the child is of school age have been reported[9]. These cognitive defects comprise lags in spelling, reading and arithmetic, difficulties with language and sensorimotor functions, sentence and repetition narrative memory, impairment of episodic memory, recall deficits and verbal learning, and difficulties with visual recall [9].

There are three notable subtypes of NE, namely, 1) hypoxia-ischaemia, 2) asphyxia, and 3) hypoxic-ischaemic encephalopathy. While hypoxia-ischaemia is characterised as an inadequate volume of blood delivered to tissues that can lead to brain injury if glucose and oxygen levels fall below critical levels [10], asphyxia is defined as a process, rather than an end point, of alternating severity and duration of progressive hypoxemia, hypercapnia and suggestive metabolic acidosis [10]. Hypoxic-ischaemic encephalopathy (HIE) is considered to be caused by a limitation of oxygen and blood flow observable around the time of birth [10].

NE aetiologies and related risk factors can include antepartum events such as maternal hypotension, intrapartum events such as abruption of the placenta, and postnatal events such as severe respiratory distress or sepsis. Regardless of risk factor or cause, the Sarnat grading scale is a grading scale for scoring the severity of neurological encephalopathy based on clinical criteria [11]. The Modified Sarnat grade scoring scale is calculated by adding the scores assigned for six categories. These categories include: the level of consciousness, spontaneous activity, posture, tone, primitive reflexes, and autonomic nervous system at birth. Scores are assigned as being normal (0), mild (1), moderate (2), or severe (3) [12-13]. The approximate breakdown tends to be 39% mild, 39% moderate, and 22% severe, respectively[14].

2.1.1 The pathogenesis of Hypoxic-ischaemic encephalopathy

The pathogenesis of HIE is the consequence of impairment in cerebral blood flow which results in interrupted oxygen delivery to the brain. This prompts a cascade of events at both the cellular and systemic level [15-16]. Prenatal or perinatal factors, including uterine rupture, cord prolapse, placenta previa, placental abruption, breech presentation, shoulder dystocia or maternal hypotension, may act as a hypoxic-ischaemic event or insult [15-16].

A decrease in the concentration of inhaled oxygen, accompanied by hypotension or the occlusion of the umbilical or carotid arteries, results in a myriad of neurotoxic events involving energy failure and the formation of reactive oxygen species (ROS). A build-up of ROS is dangerous and leads to cellular and tissue damage, particularly in the immature developing brain that has a limited ability to protect itself against ROS [16]. As shown in Figure 1, such a hypoxic-ischaemic event is comprised of three phases of brain damage, namely, acute (primary), secondary and tertiary [15].

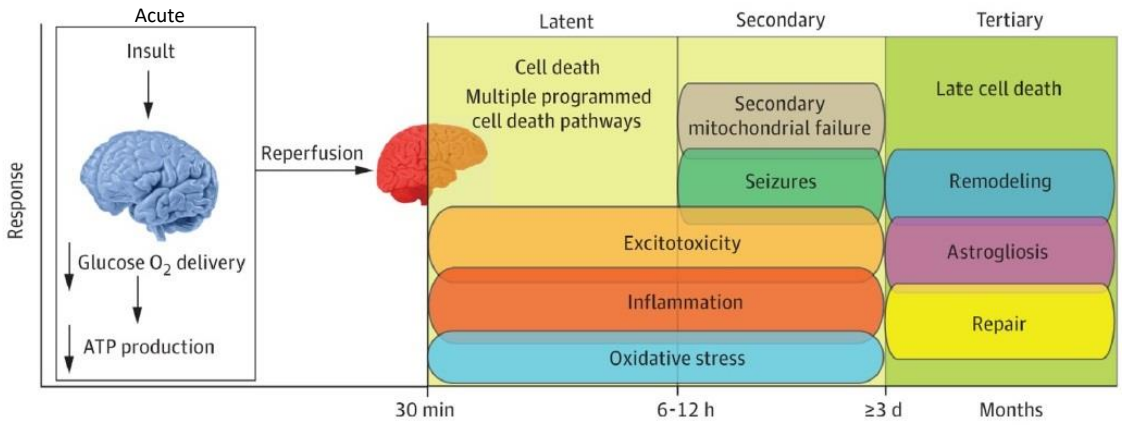

Figure 1 The three phases of brain damage following a hypoxic-ischaemic event as represented by Wachtel et al. (2019) [10]. The three phases, latent, secondary, and tertiary, all exhibit the different responses of the brain from 30 minutes after birth to weeks of life.

In the acute phase, an insult or injury prompts reduced cerebral blood flow, diminished oxygenation and decreased glucose leading to the exhaustion of adenosine triphosphate (ATP) and a decrease in oxidative metabolism [15]. This leads to immense neuronal depolarization with cerebral oedema contributing to hypoxic depolarization of neuronal cells and hence cytotoxic oedema, micro-vascular damage and necrosis/apoptosis [15]. The secondary phase occurs 6-48 hours following the initial injury and is categorised by injury to neuronal tissue. This injury is characterised by an increase in free radicals, excitotoxicity and ensuing cytotoxic oedema. Consequently, cytokine-induced inflammation occurs with accompanying cell death. This coordinates the disruptive effects of HIE [15]. The tertiary phase of brain damage consists of persistent activation of pro-inflammatory receptors. These processes may result in poor outcomes by predisposing a patient to additional injury and hindering repair or regeneration subsequent to the initial insult to the brain [15].

HIE is suggestive in 50-75% of term neonates with neonatal encephalopathy [16]. Nevertheless, most term newborns with NE ascribed to HIE don't have a recorded history of ischaemia or hypoxia during the antenatal and/or labour period. Moreover, there are no identifiable risk factors in more than half of term newborns with NE [16]. Hypoxemia, together with ischaemia (low cerebral blood flow) and hypoxia (low oxygen carriage in blood) are the fundamental physiologic processes that lead to HIE [17].

Several cognitive defects, such as problems in language, ability to learn, intellectual limitations, executive function and social skills development have been extensively noted to develop later in life following HIE injury. However, CP has been described to be the most adverse complication accounting for approximately 5-10% of noted HIE cases in high income countries[10]. In South Africa, this instance has been reported to be as high as 45% [18]. The number of survivors of extreme prematurity exceeds the number of survivors from HIE, however, the proportion of neonates with long-term sequelae has been found to be higher in the latter[19].

There are several noteworthy risk factors for HIE which include:

1) Antepartum events (such as maternal hypotension and infertility treatments). These have been reported in 25% of all cases of NE;
2) Intrapartum events (including abruption of placenta and forceps delivery). These are associated with fewer than 10% of cases; and
3) Postnatal events (such as severe respiratory distress and sepsis), which are found in less than 10% of cases.

Regardless of these risk factors, it has been reported that HIE is the cause of neonatal encephalopathy in 50-75% of term newborns [16].

2.1.2 Postnatal Therapeutic Care for Hypoxic-Ischaemic Insult: The Efficacy of Therapeutic Hypothermia and Diagnosis Criteria

As shown in Figure 1, postnatal therapeutic care becomes crucial in preventing secondary injury following a hypoxic-ischaemic insult. Therapeutic hypothermia (TH) is a widely used treatment resource [19]. With the exception of the recently published HELIX trial, early induction of hypothermia is the only therapy proven through randomised clinical trials to be

beneficial as it decreases brain post-asphyxial reperfusion injuries and it may reduce the incidence of death and the occurrence of future disabilities, particularly CP [20]. In most studies, when moderate hypothermia is initiated within 6 hours following the hypoxic-ischaemic insult, neonatal survival is improved by 40% (without CP or disability) while death or neurological disability is reduced by 30% [21]. In contrast to this, a team of researchers concluded a five-year clinical study (the HELIX trial) into the efficacy of TH in reducing death and disability in south Asia in 2020. The results of this study potentially have bearing on low-and- middle-income countries (LMICs), as most available data regarding NE and HIE is mined from high-income countries and high-resource settings. The HELIX trial is the largest clinical trial conducted in LMICs and in many ways, is setup similarly to the NESHIE study [22]. Several trials and systematic reviews have reported TH as beneficial, safe and effective; however, the HELIX trial has insisted that this is not true for LMICs. The HELIX trial researchers noted that the use of TH did not significantly reduce the outcome of death or disability in moderate and severe HIE neonates at 18-22 months [22]. Their findings have been queried extensively by peers. In a very recent publication by Nakwa et al (2023), the proportion of deaths was found to be higher in neonates who did not receive TH as opposed to neonates that did receive TH (53.4% vs 17.0%, $p< 0.05$) [23].

Typically, a diagnosis of HIE is based on a thorough and multidimensional assessment of maternal medical history, intrapartum factors (including foetal heart monitoring), placental histology, examination of umbilical cord gas directly after birth and physical examination of the neonate [12]. The following neonatal signs are important in the assessment: an APGAR score <5 at 5 and 10 minutes, the need for prolonged resuscitation up to and exceeding 10 minutes, as well as a neonatal blood gas pH < 7.0 and/or base deficit of 15mmol/L [12]. Additional features such as a Thompson score ≥ 7, and the presence of signs of encephalopathy and abnormal phenotype at birth, including abnormal reflexes, are usually noted. Collectively, since HIE can only definitively be diagnosed once the above criteria have been met, these presenting features justify the use of the term neonatal encephalopathy with suspected hypoxic ischaemic encephalopathy (NESHIE) until a definitive diagnosis has been confirmed. In 2016, Stats SA reported a 17.3% incidence associated with birth asphyxia in the category of Respiratory and Cardiovascular disorders specific to the perinatal period.

It is unknown why South Africa experiences such a high incidence of NESHIE; however, it is possible that genetic factors may a contribute [24].

2.1.3 Neonatal Encephalopathy in Sub-Saharan Africa: High Mortality and Burden of Disease

Children in sub-Saharan Africa with NE are at high risk of death and subsequent neurodevelopmental sequelae with a significant impact on affected children and families. Tann and her colleagues noted a high mortality in Ugandan children with NE up to the age of 27 months, with most deaths occurring in the neonatal period [7]. A third of these survivors were found to develop neurodevelopmental impairments such as CP and severe global developmental disability without motor impairment. Mortality at two years of age was highest in children diagnosed with CP [7].

Most statistical data quoted in literature is reflective of the HIE landscape in high income countries, but the greatest burden of disease remains in low-income countries. Lack of modern obstetric care, insufficient neonatal resuscitation and lack of TH have been cited as factors responsible for the current and progressing high incidence of HIE in these countries [15]. However, South African hospitals (as followed in this study as well as those seen in other South African studies) employ similar, if not the same standards of care compared to high income countries, alongside well-trained specialists in the field. In a 2015 study conducted in South Africa, Bruckmann and Velaphi studied the incidence of asphyxia and HIE at Chris Hani Baragwanath Hospital in Johannesburg. Approximately 30 000 births occur in total per year from both the hospital itself and referrals from its clinics [25]. They noted an incidence of HIE ranging from 8.5 to 13.3 per 1 000 births of which 60% were moderate to severe. The overall mortality rate was noted at 7.8%. This high incidence of HIE is one of a few documented studies of HIE in the South African setting and possibly reflects the incidence in similar hospitals in the public healthcare system [25]. This has prompted the exploration of other causative aspects of HIE. Much of this exploration hinges on the availability of data capturing systems.

2.2 Data collection and management systems

2.2.1 The importance of data collection and management systems in clinical research

Anderson et al. (2007) found that researchers ‘‘do not want to spend much time finding data management solutions or solving technology support problems—they would rather spend their time doing research’’ [26]. This statement has formed the backbone on which up-and-coming IT (information technology) systems depend upon to provide fitting solutions, notably in the development of electronic data capture systems (EDCs). Such systems allow for the effective management of information so that it benefits its users and is suitable for the tasks that need to be performed [27]. This is because data collection in clinical research or trials is a manually laborious process that is monotonous and error prone. However, constraints such as time and money, often cause researchers to use general-purpose office applications such as spreadsheets that are poorly suited for data management on a large scale [28].

Traditionally, investigators located at study sites typically transcribe the data from hospital records, which serve as the source documents, onto case report forms (CRFs) [29]. This allows monitors to verify and establish whether the CRFs match source document data and any additional attached documents. When discrepancies are found and queried, it is up to the delegated parties in charge of data management and monitoring to resolve the queries in a predetermined period [29]. This constitutes a form of quality control and is a multi-tiered process essential in the data collection system. It ensures that data entered on the CRFs are accurate through data review/cleaning, and real-time alert for subject safety [29].

Nevertheless, in an extensive qualitative study of over 100 researchers, Franklin and his team found that there was a strong need for data management expertise as well as infrastructure that includes database design and management, backups, security and storage [26]. As such, IT has come a long way in developing systems that fit a wide range of needs, including those in the clinical research landscape. Particularly, IT standards have been married to system methodologies to create and build clinical databases. For example, Britain saw an increased diversion of funds from 2% in the 1990s, to 5-7% in recent years towards the introduction of electronic patient records. This has notably benefited the health care and clinical research fields in the realm of data collection, even though the full impact on research data quality is insufficiently understood [30]. This growing investment into the use

of electronic case report forms (eCRFs) has accompanied the ever expanding incorporation of IT systems into clinical research [28].

While the types and uses of various IT systems may differ, the commonality of data storage, safe operational environments, and confidentiality of patient data in a controlled format, remains a shared principle. Electronic systems developed specifically for clinical research have the capability of providing support for the security of and access to research data. They also potentially increase the quality of care delivered to research-enrolled individuals [31]. In the development of these systems, care must be given to the design of the eCRF as it must be of superior data quality and aligned to the guidelines established by good clinical practice (GCP) [32]. The design, development and support of a clinical data management system is crucial in maintaining compliance regarding the quality of data to be entered into the database.

The objectives of clinical research are to discover new treatments, develop preventative measures against a range of diseases, and generate screening and diagnostic tools necessary for reducing the risk of disease in humans [33]. The results from clinical trials are frequently used to formulate the foundational structure of clinical guidelines. For this reason, it is imperative that clinical research generate high quality data through the reduction of errors associated with inaccurate and missing data. Strategies in use for data quality refinement place emphasis on the design of the trial, governance and conduct, and data collection procedures restricted only to essential items, thus limiting the potential for erroneous and missing data [30].

2.2.2 The use and disclosure of Protected Health Information

Clinical research often involves the collection of personal information from research subjects. For this reason, researchers must use measured care in protecting the confidentiality and privacy of research subjects. Explicit guidelines regarding the use of and disclosure of protected health information (PHI) originate from United State of America laws, specifically the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act [34]. In South Africa, the National Health Act (Act No. 61 of 2003) and the Protection of Personal Information Act (POPI Act of 2009, No. 4 of 2013 and its 2020 regulations) protects the PHI

of patients by prohibiting the disclosure of such data, unless clear consent is provided by the patient or health care providers can justify the disclosure thereof [34].

Although PHI laws, particularly HIPAA, set a high standard for the management and dissemination of personal health information, they are often difficult to enforce. However, most management systems reviewed in this dissertation were developed in Europe and North America and therefore adhere to these acts. The POPI Act safeguards the personal information of individuals processed by public and private health institutions. Some stipulations of this law ensure that sensitive patient information can only be accessed by healthcare providers for the express use of proper treatment, provided that patient consent is given. In the case of clinical research studies, and for teaching purposes, de-identified health information may be accessed and utilised by healthcare providers. This PHI may only be disclosed as required by law [35]. The Academy of Sciences of South Africa (ASSAf) code of conduct for research guidelines defines PHI in clinical research studies as "personal information that includes an individual's medical history, healthcare services received, genetic information, and other related information". This also encompasses sensitive personal information such as race, sex life, health, biometrics, or religious or philosophical beliefs [36-37].

2.2.3 Electronic Data Capture Systems

2.2.3.1 Overview of Electronic Data Capture systems in clinical research

As previously described, easy and low-cost data capture methods such as paper-based case report forms (CRFs), data entry into spreadsheets and offline databases have long been used. However, in the realm of large clinical trials and studies, these methods are not customisable, secure, or available to users across wide geographical distributions [38]. The development of electronic data capture systems (EDCs) has subsequently provided faster and more reliable means of data collection [26]. They have also reduced workloads and improved the quality of collected data.

EDCs have been found to be more useful than paper-based methods in clinical trials, specifically when considering the number of queries that are made, their associated responses, and the time taken to lock databases (a method of restricting access to data fields or variables in order to secure the data for analysis). While these benefits encourage

the use of EDCs in clinical trials and studies, Welker et al. (2007) have nevertheless described the conversion to EDCs as a "slow progression" chiefly because, at the time, a mere one third of clinical trials utilised them. Two of the key factors identified in the delays to adoption were directly associated with the failure to implement the system(s) [39]. In an international survey conducted to ascertain reasons for adoption delay, 46% of respondents identified inertia or concern with changing processes, while 40% identified resistance to system implementation from investigative sites. Further details associated with adoption delays include the initial expense and the accompanying ignorance regarding the eventual reduction in costs as well as the overall inability to efficiently introduce the system. These, among other reasons, contribute to the slow dissemination of new technologies into reputable environments [39]. In South Africa (and elsewhere), strict enforcement of good clinical practice guidelines makes implementation of EDCs complex and rarely achievable[40] .

2.2.3.1 Considerations for choosing an Electronic Data Capture system for clinical research

The data collection process for clinical trials or research is vital, so careful consideration must be taken when choosing a system or platform in order to best serve the needs of the study. In this way, error-prone processes that may end up as barriers to the initiation of a study can be avoided [26]. EDCs can prove to be an invaluable resource in either foregoing the manual completion of CRFs or storing their information in an electronic format [26]. For an EDC system to be reliable it must however boast a myriad of features, including real-time data availability and flexible and auditable system features[32]. Features such as minimising data entry errors and an audit trail are those most looked for in a good EDC system. User-friendly graphical user interfaces, reporting tools to analyse collected data, and data validation utilities are further desirable features [32]. These built-in features afford the user time to review captured data; this is perhaps the main reason Lu (2010) found that an error is more likely to be captured on a source document than on an electronic system.

Since EDCs employ different development and distribution systems, they are often subdivided into open source and commercial categories. Each boasts different features that have their own advantages and disadvantages [38]. Since the internet is regarded as reliable and more accessible, more web-based solutions have quickly become the preferred method for utilising an EDC system [39]. Nevertheless, software installation requires time consuming

server maintenance, customisation and error correction in real-time. Moreover, individual sites may need to install software components to their computers, if necessary [39].

In contrast to this, open-source systems use has made internet-based applications available at affordable rates to both small and large organisations. It also allows multiple users simultaneous access and use of the same EDC system, and affords provision of study documentation, site training information and trial management procedures [39]. Although several top line licence-based EDC systems offer the very same benefits, open-source software increases data intensity and provides all users the greatest productivity as an output. Secondly, permitting the software company to act as the application service provider with the crucial responsibility of maintaining the EDC system on their own servers, removes unscheduled interruptions in server access and software installation for researchers [39]. For this reason, numerous commercially available clinical trial management systems utilise EDC platforms. Open-source solutions typically do not offer formal guarantees or liability for software defects, and users cannot expect the same level of support as provided by commercial software vendors. Nonetheless, major open-source projects often benefit from robust communities of developers who actively contribute to the project. These communities frequently offer substantial support and are responsive in addressing and rectifying bugs.

The ever-growing demand for “bigger and better”, “sleeker and faster” EDC technology has contributed to a drive towards continuous improvement, smart features, and the quick ability to configure these systems, with the added benefit of improving clinical proficiencies while decreasing the time spent conducting the study [41]. It has been noted repeatedly in the literature that users prefer to utilise a system that is user friendly, therefore, a considerable amount of computing time and design must be rightly allocated to this attribute. To achieve this, processing power must be considered as it allows for the manoeuvring of data. This ensures that the database can perform a wide range of functions, not limited to match, sort, link, aggregate, calculate, skip fields and arrange [27]. Because of the ever-increasing options available through EDCs, standardisation and its associated implementation is a challenge, as the principal focus centres on the proficiency and versatility of the data entry process [31]. This is vital for research quality, reproducibility, and consistency. Nonetheless, deciding on and selecting common standards for clinical research is a difficult task because of diverse

research needs. These needs include research interests, technical problems, the intended use of the data and terminology criteria [31].

While there are various database models, the most commonly used include object-oriented, relational and network database models. The relational database model is the most widely used model as it is the preferred choice in most information systems [42]. This database model contains a set of tables that describes the relationship(s) among and across data fields known as relations [42].

2.2.3.2 Comparison of Electronic Data Capture systems for clinical research

Over eighty competitors exist in the field of clinical research data capture with several standout platforms such as IBM Clinical Developments, REDCap and Videoc. These platforms offer features like remote-access data entry and data capture design capabilities, specifically for use in electronic data capture in clinical trials [39]. REDCap, which shares common features with survey creation platforms such as SurveyMonkey and Qualtrics, was developed by clinical researchers for clinical research, making it useful primarily as a database platform for clinical research with survey tools also available. A lead REDCap administrator at the University of Washington's Institute of Translational Health Sciences emphasises this distinction [39]. The University of Pretoria currently holds licences for both REDCap and Qualtrics to meet the clinical research needs of multiple fields. While Qualtrics is a survey tool, it has limited capabilities for clinical trials or research.

In a study conducted by Franklin et al. (2011), the Catalyst and OpenClinica systems were compared to REDCap to determine the optimal electronic data capture platform for clinical research. While Catalyst was found to be a simple, open-source system with extensive online tutorials and in-person training courses, it lacked appropriate support for de-identifying sensitive patient data, a crucial aspect of clinical research [26]. OpenClinica, on the other hand, offered strong site management, de-identification of PHI and an in-depth multifunctional CRF design suitable for large-scale studies [28]. Nonetheless, the team concluded that the steep learning curve and significant time investment placed OpenClinica at a disadvantage. The study highlights the importance of choosing electronic data capture systems appropriate for supporting the needs of the scientific research being conducted [28].

In addition to well-known EDC platforms, there are several lesser-known systems that can aid in clinical data management. The Clinical Data Reporting Tool is a custom Windows application that creates several complex data management reports for active data-cleaning. These reports are organised and accessible by assigned personnel. ePharma is a training tool that offers investigators protocol and safety training, along with document trafficking [32]. Clinitrial and Oracle Clinical are roles-based clinical database management systems that can unify electronic or paper-based clinical data management systems into one platform [32].

To setup a secure database or project, REDCap, unlike OpenClinica, does not require any prior programming knowledge or skills. Unlike OpenClinica, REDCap and Catalyst can be accessed from any device with an internet connection and web browser, as long as remote access privileges are enabled [6]. Lastly, while Catalyst and OpenClinica are both open-source systems, OpenClinica has a specific fee structure for support services that Catalyst does not have. In comparison, REDCap is not open source, but academic institutions can partner with Vanderbilt University to obtain a REDCap licence at no charge.

Because REDCap facilitates remote access and collaboration, provides a secure file repository, and offers extensive training materials including online tutorials and pre-recorded webinars [5], it was selected as the preferred EDC system for this study. The REDCap website contains a counter which shows that it is employed by 5 223 institutions across 141 countries. There are over 1 230 000 million projects currently using REDCap with 1 786 000 million users, and REDCap has been cited in 14 677 thousand articles as of 15 July 2021 [43]. In the case of this study, REDCap incorporates a relational database system and therefore, the database design will conform to this model.

REDCap was developed to support health-related research in the biological sciences and particularly regarding clinical studies. It contains a secure login page, an excellent graphical user interface (GUI) and a variety of coding choices on the back end [44]. In addition to the creation of databases, REDCap can be used to perform surveys as well as blinded expert tools and auditing trails, among many other uses. It is designed to address the needs of both investigators and participants; however, in the NESHIE study it will strictly be used by authenticated investigators involved in the study [45]. REDCap can be used to collect any type of data compliant with the Federal Information Security Management Act (FISMA) and Health Insurance Probability and Accountability (HIPAA) privacy and security guidelines [46].

While these acts were founded in the United States of America, their principles may be applied in South Africa. The HIPAA Act of 1996 operates with the objective of safeguarding patient PHI through five key principles [47]. These include: 1) disclosure of a research or corporate entity's privacy and data protection practices to a patient or study participant, 2) research or corporate entities are obliged to disclose what PHI that will be collected as part of the study aims and research objectives, as well as explain how this will be used [47], 3) the participant's written consent for the collection and disclosure of research-specific, aggregated PHI, 4) patient are entitled to and may request a full account of PHI that is disclosed by the research or corporate entity, and 5) in certain conditions, patients have the right to access the research or corporate entity's designated record set [47].

REDCap is a PHP-based system, storing data in a MariaDB database (previously MySQL). It uses SSL certificates for secure connections. Some vital data protection tools offered on REDCap include the auto-logout function which is triggered when no activity is detected on the site, and strict user right designation that ensures that users can access data assigned only to their user profile and account [46]. One of the notable advantages of REDCap is that it allows researchers to swiftly define and design project-specific data capture forms and quickly initiate the data collection process [48]. Additionally, it also offers basic support through training videos and a comprehensive FAQ section, daily backups, and user-friendly data export mechanisms that grant users with significant control over their data [39].

REDCap's versatility and utility have been demonstrated in various healthcare and research contexts. During the Ebola Virus Disease (EVD) outbreak that occurred between March 2014 and June 2016 in the Democratic Republic of the Congo, Stefanie Hossmann and a team from CTU Bern in Switzerland conducted two vaccine trials. REDCap was the data capture system of choice [49]. To track the success or failure of the trials, a data management strategy was employed in which 575 variables and 648 additional data points were collected [49]. A team from the Carolinas Medical Centre recognised the need to develop a predictive model that allowed health care workers to perform real-time risk assessment during hospitalisation with the goal of reducing readmission rates as well as identifying high-risk patients in the postoperative setting. Patient data, which included demographics, family history and physical health status, were collected from the medical records of 400 patients, and then transferred to the REDCap database. This collation of information allowed the

researchers to identify 21 input variables as a basis for a predictive model that was able to compute a numeric risk probability score [34].

Despite its robustness, REDCap is not without its challenges. At the University of Washington Health Translational Library, librarians using the system reported a steep learning curve when using the system [39]. Amongst a number of errors, they reported that they accidentally deleted data from records and events as well as accidentally renamed fields [39].

In addition to these systems, software clouds provide a framework for supporting and integrating different applications and offering partially integrated web-based services. There are more than a thousand such platforms including Salesforce.com and Google Apps which use their large infrastructures to offer cloud computing services and encourage businesses to adopt them [50]. LogicalDOC is a cloud-based document management platform which offers an easy-to-use interface in the move to store data. Due to the advanced technology employed in its development, LogicalDOC can easily integrate with REDCap to provide data storage and management solutions [51] making it an attractive option in complementing its implementation for this particular study.

In more detail, LogicalDOC is a Java-based web application running in TomCat and employing SSL certificates for secure communication. This entirely web-based application is compatible with most web browsers such as Chrome and Microsoft Edge and is capable of indexing the most common file types such as Microsoft Office documents and PDF files. LogicalDOC provides a web interface through which storage of document-based data on a dedicated server is possible. Servers can be either private (preferred for the NESHIE study) or commercial (hosted on third-party servers). The LogicalDOC interface does not store documents itself, but stores links to the documents stored on the file system. Provided that access permissions, particularly those for remote access, have been enabled, the NESHIE team is able to share and access documents in a secure and safe way across the country or globe. While an audit trail tracks the document repository and adjustments, backup and firewall protocols ensure that LogicalDOC is a safe environment that allows only authorised users access to its content. It is able to handle large volumes of data and is an invaluable resource to the NESHIE study [51].

2.3 Database design

The fundamental goal of database design is to produce normalised, comprehensive, and fully integrated conceptual, logical, and physical database models. This means that the database should be able to handle data efficiently, with minimal redundancy and data inconsistencies. Database design is a complex process that involves several stages, starting with the definition of all the data elements that need to be measured, and ending with the interpretation of results. Three core elements dictate database design. These include: 1) requirement analysis and conceptual design, 2) logical design and finally, 3) physical design[41, 52]. Several principles, which include accurate data collection and analysis support, govern this process. In order to facilitate accurate data collection, consideration must be given to the structure utilised for the storage and management of data, rather than the actual database software in use [53]. Adherence to these principles depends on the database's ability to be flexible, while also addressing any design limitations [53]. Nevertheless, after the design stage, implementation is a key part of the database development process. This process is best summarised in Figure 2.

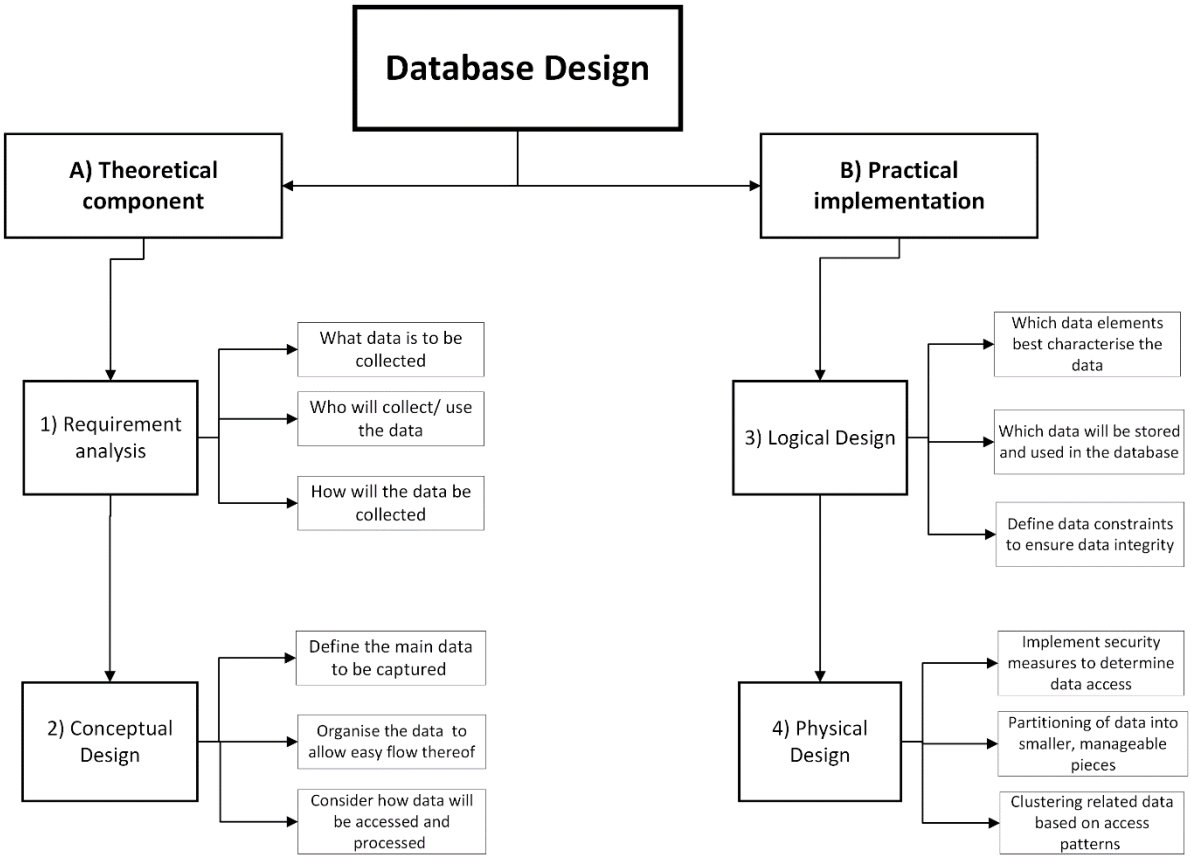

Figure 2: A breakdown of the process of database design (created by T Kalua). Firstly, database design branches into two components: theoretical and practical implementation. the theoretical component consists of the requirement analysis and conceptual design, and the practical implementation consists of logical and physical design elements.

2.3.1 Requirement analysis and conceptual design

To design a usable database, it is essential to begin with an abstract overview of the general data environment and gradually add specifications as the project progresses [54]. The first step in this process is the requirement analysis, which involves interviewing producers and end-users of the database to determine the database's needs. End-users are the people who utilise the application programs to work with the data and produce information [54]. Requirement analysis is a crucial step that precedes the database design build. It ensures the efficacy of subsequent steps, such as physical design, to guarantee the functionality of the final product. [27].

Requirement analysis involves collecting, analysing, and reporting data in the form of statements or specifications, and should be conveyed independently of how the requirements may be achieved [27]. The first part of the requirement analysis is a thorough

plan for data collection, analysis, and reporting, which can be informed by documents like the study protocol, statistical analysis plan and the data management plan [52]. If these documents do not exist, the relevant experts are tasked to create examples of the type of information to be documented [55]. Simple diagrams or illustrations are often compiled to delineate and prioritise the way the information system's key operations will interact with the study environment. It is also essential to document estimates of the expected data volumes, including the frequency of data queries and their respective response times. This information can determine the necessary hardware and software resources required to support the information system effectively [55].

The output of the requirements analysis is an essential document that outlines the non-functional and functional requirements of the information system. The document also includes maintenance information that details how the system will be maintained over time. [55]. To ensure the usefulness of the output document, it must be "unambiguous, complete, verifiable, consistent, modifiable and traceable" as stated by Halpin (2008) [55]. Creating a well-documented requirements analysis is crucial to the success of any information system project. It provides a clear roadmap for the development team and ensures that all stakeholders are aligned regarding the system's requirements and functionality.

The completion of the requirements analysis affords the opportunity to commence with the conceptual design; it serves as a communication tool between the database designer and end-user and bridges the requirements analysis and logical design phases associated with database design [41]. The requirements specification, which focuses on data relationships, is used to produce the conceptual data model [27,41]. The requirements analysis and conceptual data modelling phases are closely related and frequently overlap. Experienced database designers or modellers can leverage their expertise to identify similar features between previous and current models or applications. This allows them to reuse strategies and design elements from prior models, saving time and effort in database development [27]. To facilitate the reuse of design elements and strategies, online forums and web pages are available that offer best practices, design patterns, and templates. By adopting these resources, designers and modellers can ensure that their designs are efficient, effective, and in line with industry standards [55].

2.3.2 Logical design

After the conceptual database design is produced, the next step in the database development process is the logical design. The goal of this phase is to produce a company-wide database that is based on data models developed during the conceptual design phase. The focus of the logical design is on the structure of the database and is separate from physical-level considerations [56]. The logical design is a crucial step in the database design process as it produces a database schema, which is the blueprint for the actual database. The schema specifies the data elements, their relationships, and the constraints that govern them. With the schema in place, the first stage in executing the database has been completed [27].

The logical design phase involves making choices about which data elements would be most appropriate for characterising the data in the database. These choices are based on design principles such as limiting duplication of data fields and ensuring flexibility to incorporate changes as needed and according to the needs of the study [27]. During this phase, the relational tables selected during the conceptual design phase are defined in the logical schema, which determine what data is stored and how it can be employed in the database [27]. The logical schema establishes relationships between the data elements and how it can be used in the database. The goal of the logical design phase is to produce a comprehensive and efficient database schema that can be used as the basis for the physical implementation of the database.

As an example of a logical design element, coding of data is crucial in the design process. Coding involves assigning meaningful categories to the collected data. To ensure that coding is consistent and accurate, a coding schema should be determined at the start of the study and incorporated into the CRFs and database as much as possible [52]. It is imperative that coding schemas be carefully documented and precisely tracked. While there are many other logical design considerations, their importance relative to REDCap will be discussed elsewhere.

2.3.3 Physical design

The final phase of the database design involves selecting access methods, partitioning and clustering of data to build the physical database, with the goal of optimizing performance before launching it for widespread use [52]. While the logical design focuses on "what", the physical design focuses on "how" the database is structured and implemented [57]. The objective of this step is to optimise performance to the fullest before the database is launched for widespread use. Physical modelling uses the input of the database schema from the previous step and creates storage structures to introduce the schema in computer systems [56].

By focusing on one aspect at a time, complexity can be managed, and each level produces an output that flows into the next tier, creating an easy-to-follow structure (abstraction) [56]. This phase is the final tier in the abstraction process and is carefully related to the performance of the database by describing the way data are saved on storage media to maximise space utilisation, response time and transaction throughput [57]. Moreover, the database must be safeguarded against data loss, and recovery plans tested [54]. Security measures should also be introduced to ensure that all users of the database are authenticated and access to the data is restricted as necessary [54]. Although user management such as defining user and security groups and roles is more of a function of database administration than database design itself, it is an important consideration.

Database design can be difficult because it may include problems that cannot be foreseen at the onset of the design process. In other words, the design process must adapt to ongoing, changing requirements presented by database queries [56]. These needs or problems may only be realised through use of the database in a "test" environment. Testing of the databases objectively assesses its performance and precedes its final implementation. More specifically, testing of the database determines if it is fit for its intended purpose thereby fulfilling the criteria established during the requirements analysis phase [54]. Any challenges encountered during the integration and deployment plans are addressed, followed by user training and system documentation. After the physical design has been deemed satisfactory, the database enters the production phase in which the database design is executed. At this point, end-users may actively engage with the tested system. Once the system has been thoroughly tested and receives final approval, it should become a useful resource for the

organisation. However, there are several factors that can affect the database, such as power outages and unintended data deletion. Therefore, data backup and recovery procedures are essential to ensure consistent data availability [54].

2.3.4 REDCap: Practical applications in logical and physical design

The general principles of database design provide a foundation for building efficient and purpose-driven systems. REDCap, as a versatile platform, boasts a number of features that align with the logical and physical database design phases of development. A summary of some of the core elements to consider in database design when using REDCap are summarised in Table 1. These elements create the framework that is used throughout the database to meet the criteria stipulated in the requirements analysis phase of the design process.

Table 1: A summary of the logical and physical database design elements in REDCap

Logical Design Elements				
Database design elements	Variable name	Field type	Field label	Field choices
	Field notes	Field validation	Data validation	Field upload
Advanced design elements	HTMLs			
Physical Design Elements				
Data grouping and input control elements	Matrix of fields	Branching logic	Piping	Custom alignment
Data partitioning elements	Data instruments	Data events	Repeatable instruments	Data arms
Data quality & security elements	Data resolution workflow	Double-data entry	Data quality module	User rights & permissions
Optional 'other' elements	Auto-numbering of records	Randomisation module	Scheduling module	
Logical and Physical Design Elements				
Shared design elements	Identifying fields	Field annotation	Data access groups	Codebook & Data Dictionary

Regarding the logical design elements in REDCap, the variable name, field type, field label and field choices options allow for the stipulation of the type of data to be captured in the envisaged database. The use of field notes, field validation and data validation elaborate and provide guidance on the type of data to be captured by the end-user. As a separate consideration, Hypertext Markup Language (HTMLs), describes a string of code that assigns

colour, typographical emphasis, style, font size and several other features to a body of text. This feature can be integrated into a REDCap database in the following ways:

1) Distinct colour-coded HTML prompts could be displayed as headers to provide information to users regarding data input per instrument. Similarly, if access has been granted or denied, section headers can be set up using HTMLs to provide users with pertinent information to either proceed or return to previous instruments to correct any mistakes. Following the correct and full completion of an instrument, a distinct colour-coded HTML prompt could be used to inform users that they may proceed to the following section, as well as to remind users to upload corresponding physical source documentation.
2) Similarly, bold, distinct colour-coded HTML prompts could act as partitions improving the readability and organisation of data. These partitions separate the data into clear sections.

Physical design elements in REDCap are useful to group or control the flow of the data fields incorporated in the database through the logical design process. Those physical design elements that are incorporated during the logical design phase include the matrix of fields, branching logic, piping and custom alignment options. When a section contains a group of data points with the same answer choices, the “Matrix of fields” feature can group the fields into an ordered flow of data. The answer choices can be specified as checkboxes or radio buttons, and a group name can be assigned to the matrix. Branching logic is a specific syntactical language that creates custom data input pathways based on user responses. Simply put, it governs the flow or order of data entry in the envisaged database. Lastly, while piping allows user-captured data from a data field to be channelled into a field label [58], custom alignment affords the designer to format the alignment of data fields on the final database as “left”, “centre”, or “right”. Once these elements have been incorporated into the design of the database, decisions regarding the further partitioning of data are required.

Data partitioning is achieved through the data instrument, repeatable instrument, data event and data arm functions within REDCap. Data instruments partition data fields into distinct groupings of related information, while data events represent a collection of instruments. Instruments may be repeated when multiple instances of the same data need to be collected. An example of this is seen when collecting data on administered

medications in patients during clinical trials. Data arms facilitate the collection of data relative to participant designation within the study (e.g. control participant vs. patient). It also affords specification regarding relative timelines for data collection of each group of study participants.

Physical design functions that facilitate data quality include the data resolution workflow, double-data entry and data quality module elements. Data resolution workflows is a module that can be used to raise data queries or comments. Data queries may be addressed to a specific end-user, while comments serve the function of qualifying a particular data field. The data query function allows the end-user to specify a reason for the observed query and automatically captures the time taken for this to occur. While double-data entry is an option that facilitates the capturing of the same data by two independent end-users as a form of quality control and validation, the data quality module affords a means whereby specific restrictions on data may be implemented. This is useful in identifying real-time discrepancies during the data capture process. Regarding security functions, REDCap incorporates user rights and privileges as a crucial feature to grant data access to authenticated users only. The project creator automatically receives full rights and can grant other users access to different modules and functionalities while limiting their database design and data access privileges as needed [58]. For example, this controls the user's ability to enter and export data, alter or add database fields, create or run reports, adjust user privileges and view logging records [58].

Other physical design elements in REDCap that can be considered includes the auto-numbering of records, randomisation module and scheduling module functions. The "auto-numbering of records" function automatically generates sequential numeric record identifiers for each participant's data captured to the database. Useful in clinical trials, the randomisation module will randomly assign participants to a treatment or control group by way of in-built randomisation models. These include stratified, group/site, or data field-related models. The scheduling module function is useful in longitudinal studies that require follow-up of participants and is an optional feature in relevant database designs.

Lastly, shared logical and physical design elements in REDCap include identifying fields, field annotation, data access groups, and the codebook and data dictionary. Identifying fields are a means through which PHI can be protected during the data export process – those data

fields marked as a field containing identifiable information will not be shared when de-identified data are exported during the course of the study. Field annotation on the other hand provides a further means by which the flow of data can be controlled. In contrast to field notes, field annotations are not necessarily visible to the end-user. As an added measure of data security, data access groups (DAGs) can be utilised to compartmentalise users' access to dedicated group data. Lastly, in REDCap, the codebook feature provides a full description of every code built into the database during the logical design phase, which can be downloaded in PDF format during the physical design phase. The governance of the codebook is determined by the data dictionary. The data dictionary is a tool that stores definitions and metadata of the data elements and their relationships within the database and can be used to search for the necessary data component and associated data framework. Any changes or adjustments made to or on the database are immediately reflected in the data dictionary. This frees the data designer from manually modifying the instruments or programs that access the changed form or structure. In this way, REDCap provides for data abstraction, removing structural and data dependencies from the database system [54]. Proper documentation and tracking of the coding schema and data dictionary is essential for maintaining data quality and facilitating analysis.

While not strictly falling into the physical design phase, the ability of the EDC system to export data is an important consideration. Therefore, database designs need to be compatible with other database platforms and data analysis software or packages. REDCap offers this benefit. The data generated in REDCap can be easily exported and analysed in statistical packages such as Strata, SAS, SPSS, and R.

2.4 Data Quality

2.4.1 The importance of data quality

Irrespective of the purpose of data collection, high quality of collected data is essential. The early definition of data quality was restricted simply to measuring accuracy of data capturing. Currently, definitions of data quality remain inconsistent; however, Harrison, Rahimi and Danovaro-Holliday suggest that it should be explained as "the degree to which it [the data] is fit for its intended purpose" [59-60].

Wang and Strong analysed data quality by using the viewpoint of data users and compiled more than 100 traits of data quality into four specific categories that include relevance, accuracy, representation and accessibility [60]. Therefore, it can be said that quality data is data that is relevant, accurate, precise, complete and timely enough to fulfil its envisioned objective.

Limited studies recognize or offer evidence on when, where and how data quality challenges arise, or the causative factors of poor data quality or strategies that can be applied to improve data quality. Davenport and Harris observed that even organisations that consider data to be a strategic asset, "struggle with data quality" [61]. They further noted that some business people are taking up to 30 hours per week to deal with data quality issues, in some cases leading to hundreds of millions of dollars of previously unmeasured 'commercial damage' [60]. There is a pressing need to recognise components affecting and influencing the data process particularly because data quality problems most often emerge in the data collection process. This offers pertinent implications and insight for data quality management [62].

Poor data quality is a key problem encountered in clinical research, as it affects data integrity and the reliability of research findings. The most prominent issue with data quality is that it presents a substantial risk that can result in operational inefficiencies and monetary loss for organisations that rely on data for decision-making [60]. In the context of the medical sciences, this has the potential to impact on patient care and outcomes. Therefore, poor data quality must be reduced to generate accurate and reliable data. Ideally, this would be achieved through active and progressive data monitoring, rather than relying on retrospective searches and fixes [33].

As shown in Figure 3, there are several types of data errors and reasons for poor data quality. While these include factors such as data accuracy and completeness, behavioural and technical competencies also have a role to play.

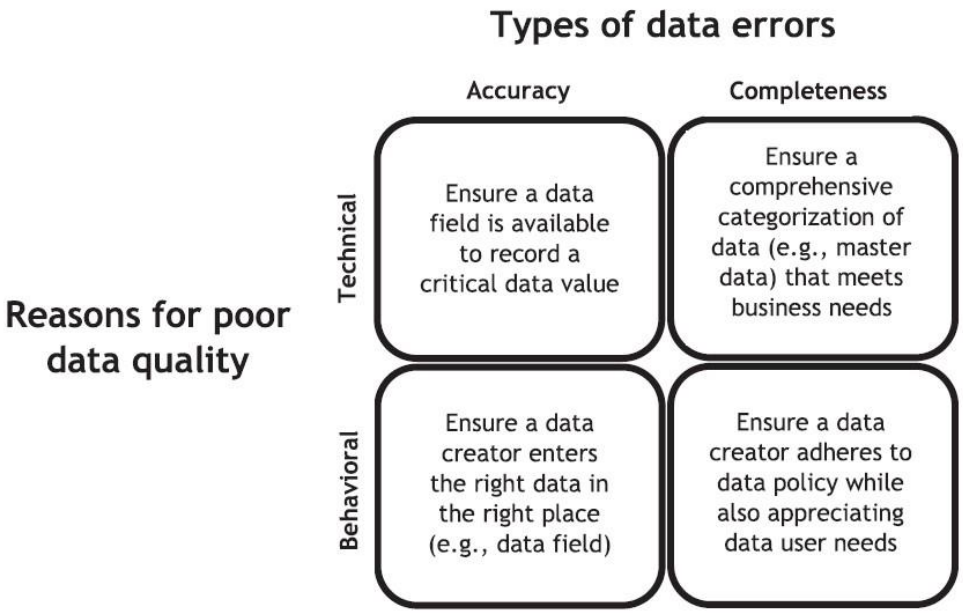

Figure 3: The types of data errors as published by Nagle, Redman and Sammon (2020) [63]. The two types of data errors, accuracy and completeness, intersect with technical and behavioural reasons for poor data quality.

2.4.2 Data quality in clinical research studies

In a clinical study targeted at ascertaining an accurate picture of immunisation coverage, Harrison and colleagues met the unfortunate obstacle encountered in most clinical trials – poor data quality. Chief among their observations were missing values and inaccuracies in entered data [62]. It was expected that poor quality immunisation data would be observed in mostly disadvantaged and low-income countries; however, the health infrastructure systems of several high-income countries were implicated, proving that the problem of good data quality is a challenge faced in all spheres [63].

Researchers looking to qualitatively evaluate the quality of the data collection process used by the Chinese national HIV/ AIDS data repository (CRIMS) recognised that data is vital, particularly for decision making and interventions associated with public health programs [61]. Prior to this study, Chen *et al.* (2021) identified that numerous data quality assessment methods in national public health information systems were directed towards the data collection procedure while little attention was focussed on how external contextual factors (such as personnel and the environment) influenced and affected the actual quality of the data collection process [62].

One of the ways hospital-based trials can regulate research data collection is to leverage existing data within the electronic health record (EHR), for example, patient files. In complex, large-scale community-based trials, data quality depends heavily upon the accuracy at which data collectors conduct their administerial duties [63]. This includes the completion of questionnaires and assessments with participants, extracting data from clinical records (when relevant and where available), and the processes involved in the follow up and resolution of queries when odd or unusual data is encountered. This combined responsibility is a challenge that holds risks for data quality [63].

Poor data quality occurs for a variety of reasons, such as inadequate operational support for data collection, dense data abstraction procedures requiring data collectors to interpret and adhere to definitions, and manual calculation of values. The first two points are significant contributors to poor quality data [33]. Poor usability was noted to be associated with frustration, inefficiency, confusion and stress for study staff. All of these factors may negatively impact overall data quality. Contrarily, enhancements in usability have been established to improve data quality and minimise errors seen in a hospital setting [62].

Modern-day data collection systems have been shown to improve data quality through enhanced data validation. However, it should be noted that they are only effectual to the level that they are intuitive and the ease with which assigned end users can work with the system [63]. The degree to which technology can be effectively and successfully implemented by users is an important limiting factor in the update of EHR systems [63].

2.4.3 Monitoring techniques in clinical research

As mentioned, data monitoring is one of the means through which data quality principles can be enforced. The International Council for Harmonisation (ICH) defines monitoring in clinical trials as supervising the progress of a clinical trial to ensure that it complies with the protocol, Standard Operating Procedures (SOPs), Good Clinical Practice (GCP), and applicable regulatory requirements [64]. However, this definition is broad and does not specify the frequency or extent of monitoring required to ensure data quality. The nature and scope of monitoring should depend on factors such as the objective, purpose and size of the study [64].

There are several ways to conduct data monitoring in clinical trials or research studies. One of these is source document verification (SDV), a traditional method that involves comparing original data such as source documents and medical records with the study's CRF. The review by Nahm et al. (2008) reported an error rate of 976 errors per 10 000 variables in a source-database comparison while a 14 per 10 000 variables was noted in a CRF-to-database comparison [65]. Although SDV can effectively identify errors, it is a rigorous and time-consuming process that focuses on detecting errors rather than preventing them. Despite its limitations, researchers consistently practice SDV to ensure better quality data in the long run. Typically, data in a research trial encompasses non-critical and critical data points, including inclusion-exclusion criteria, primary and secondary endpoints, and informed consent [64] However, regulations do not provide clear guidance on the frequency and extent of monitoring activities. Therefore, researchers rely on their discretion and the study's needs to determine the appropriate monitoring approach.

Central monitoring is another technique that provides an overview of all the trial sites to ensure trial parameters comply with the protocol, SOPs, GCP and applicable regulatory requirements [64]. This technique focuses on critical aspects, such as adherence to enrolment criteria, safety trends, protocol deviations, and primary efficacy outcomes. While these features can also be monitored on-site, central monitoring offers the advantage of viewing all the sites simultaneously. The FDA recommends that a "representative" amount of study records should be reviewed instead of all the records [64]. While predefined SDV criteria guide the choice of the representative sample of records, a flexible approach can also be taken by the researchers.

EDC software is widely used in clinical trials to manage study data. However, in some studies, such as the NESHIE study, paper CRFs are still used and are an essential component of the study. Nahm et al. (2008) reviewed the errors that occurred during medical record abstraction or transcription, and their findings revealed that most errors occurred during these processes for both source-to-database and CRF-to-database comparisons. Despite the wide use of EDC software, its adverse effects have yet to be fully investigated [65]. Although the NESHIE study is an observational clinical research investigation and not a clinical trial, monitoring and data validation practices similar to those used in clinical trials have been applied in this study. Therefore, the use of paper CRFs and monitoring practices in the

NESHIE study can provide valuable insights into necessary data management practices required throughout the study period.

Although there is no evidence to support the notion that more data monitoring leads to better data quality, SDV is still a viable option [66]. The European Medical Agency (EMA) recommends that sponsors shift towards a systematic and risk-based approach to quality management. This approach involves recognising the continual risks associated with various activities throughout the clinical trial process, from design to reporting [67]. These include the design, conduct, execution, evaluation, and reporting of clinical trials or studies. To incorporate data monitoring into the trial protocol, the process should begin when the protocol is designed, and other relevant documents are prepared [67]. In 2016, the GCP E6(R2) guideline suggested that on-site monitoring be reduced and replaced with risk-based approaches or a combination of on-site and centralised monitoring [66]. These approached offer effective and flexible monitoring procedures. However, limited guidance is provided on how to conduct these procedures [66]. Nevertheless, research indicates that errors in primary or critical data points in clinical research have a greater impact than errors in secondary or non-critical data [66].

2.5 Conclusion

The aim of this Master's project was to develop and implement a REDCap database for the NESHIE study while considering database maintenance, data quality and error reporting. The project presents a methodology for developing and implementing a REDCap database for large-scale research projects in neonatal and maternal health with a specific focus on NESHIE. It also presents means through which data quality and identification of errors can be ensured. The methodology could serve as a guideline for researchers and practitioners who plan to use REDCap for their projects within the scope in which this database has been developed. Moreover, the error reporting technique used in this study can be used to identify genuine errors that evade the data verification process, which can improve the accuracy of research findings.

2.6 References

1. World Health Organisation. Newborns: Improving survival and well-being [Internet]. World Health Organisation; 2020 [cited 2021 Jul 9]. Available from: https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality.

2. Alake O, Hardman S, Chakkarapani E. Managing hypoxic ischaemic encephalopathy in term newborn infant. Paediatrics and Child Health. 2018; 28(9):399-404. doi: 10.1016/j.paed.2018.07.002

3. Workineh Y, Semachew A, Ayalew E, Animaw W, Tirfie M, Birhanu M. Prevalence of perinatal asphyxia in east and central Africa: systematic review and meta-analysis. Heliyon. 2020; 6(4):e03793. doi:10.1016/j.heliyon.2020.e03793

4. Aslam HM, Saleem S, Afzal R, Iqbal U, Saleem SM, Shaikh MWA, et al. Risk factors of birth asphyxia. Italian Journal of Pediatrics. 2014; 40(1):94. doi:10.1186/s13052-014-0094-2

5. Gosar D, Tretnjak V, Bregant T, Neubauer D, Derganc M. Reduced white-matter integrity and lower speed of information processing in adolescents with mild and moderate neonatal hypoxic-ischaemic encephalopathy. Eur J Paediatr Neurol. 2020; 28:205-13. doi:https://doi.org/10.1016/j.ejpn.2020.06.003

6. Kurinczuk JJ, White-Koning M, Badawi N. Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy. Early Hum Dev. 2010; 86(6):329-38. doi:10.1016/j.earlhumdev.2010.05.010

7. Tann CJ, Webb EL, Lassman R, Ssekyewa J, Sewegaba M, Musoke M, et al. Early childhood outcomes after neonatal encephalopathy in Uganda: A cohort study. EClinicalMedicine. 2018; 6:26-35. doi:10.1016/j.eclinm.2018.12.001

8. Azzopardi D. Clinical management of the baby with hypoxic ischaemic encephalopathy. Early Hum Dev. 2010; 86(6):345-50. doi:10.1016/j.earlhumdev.2010.05.008

9. Eunson P. The long-term health, social, and financial burden of hypoxic-ischaemic encephalopathy. Dev Med Child Neurol. 2015; 57 Suppl 3:48-50. doi:10.1111/dmcn.12727

10. Wachtel EV, Verma S, Mally PV. Update on the current management of newborns with neonatal encephalopathy. Curr Probl Pediatr Adolesc Health Care. 2019; 49(7):100636. doi:10.1016/j.cppeds.2019.07.001

11. Sarnat HB, Flores-Sarnat L, Fajardo C, Leijser LM, Wusthoff C, Mohammad K. Sarnat grading scale for neonatal encephalopathy after 45 years: An update proposal. Pediatr Neurol. 2020; 113:75-9. doi:10.1016/j.pediatrneurol.2020.08.014

12. Wilson LA, Fell DB, Hawken S, Wong CA, Murphy MSQ, Little J, et al. Association between newborn screening analytes and hypoxic ischemic encephalopathy. Sci Rep. 2019; 9(1):15704. doi:10.1038/s41598-019-51919-x

13. Chalak LF, Adams-Huet B, Sant'Anna G. A total Sarnat score in mild hypoxic-ischemic encephalopathy can detect infants at higher risk of disability. J Pediatr. 2019; 214:217-21.e1. doi:10.1016/j.jpeds.2019.06.026

14. Ahearne CE. Short and long term prognosis in perinatal asphyxia: An update. World J Clin Pediatr. 2016; 5(1):67. doi:10.5409/wjcp.v5.i1.67

15. Cánovas-Ahedo M, Alonso-Alconada D. Combined therapy in neonatal hypoxic-ischaemic encephalopathy. An Pediatr (Engl Ed). 2019; 91(1):59-.e6. doi:10.1016/j.anpede.2019.04.007

16. Bruun TUJ, Desroches C-L, Wilson D, Chau V, Nakagawa T, Yamasaki M, et al. Prospective cohort study for identification of underlying genetic causes in neonatal encephalopathy using whole-exome sequencing. Genetics in Medicine. 2018; 20(5):486-94. doi:10.1038/gim.2017.129

17. Garrido MMB, Jordi Gascón. Hypoxic encephalopathy, miscellanea on encephalopathies - a second look. In: Tanasescu R, editor. Hypoxic Ischemic Encephalopathy - From Bench to Bedside. IntechOpen; 2012.

18. van Toorn R, Laughton B, van Zyl N, Doets L, Elsinger F. Aetiology of cerebral palsy in children presenting at Tygerberg hospital. South African Journal of Child Health. 2007; 1(2): 38-42.

19. Martinello K, Hart AR, Yap S, Mitra S, Robertson NJ. Management and investigation of neonatal encephalopathy: 2017 update. Arch Dis Child Fetal Neonatal Ed. 2017; 102(4):F346-F58. doi:10.1136/archdischild-2015-309639

20. Pereira CO, Dias A, Nunes Vicente I, Pinto JT, Marques C, Dinis A, et al. Prognostic value of near-infrared spectroscopy in hypoxic-ischaemic encephalopathy. An Pediatr (Barc). 2021; 94(3):136-43. doi:10.1016/j.anpede.2020.07.001

21. James A, Patel V. Hypoxic ischaemic encephalopathy. Paediatr Child Health. 2014; 24(9):385-9. doi:10.1016/j.paed.2014.02.003

22. Thayyil S, Pant S, Montaldo P, Shukla D, Oliveira V, Ivain P, et al. Hypothermia for moderate or severe neonatal encephalopathy in low-income and middle-income countries (HELIX): a randomised controlled trial in India, Sri Lanka, and Bangladesh. Lancet Glob Health. 2021; 9(9):e1273-e85. doi:10.1016/S2214-109X(21)00264-3

23. Nakwa FL, Sepeng L, van Kwawegen A, Thomas R, Seake K, Mogajane T, et al. Characteristics and outcomes of neonates with intrapartum asphyxia managed with therapeutic hypothermia in a public tertiary hospital in South Africa. BMC Pediatr. 2023; 23(1):51. doi:10.1186/s12887-023-03852-2

24. Conway JM, Walsh BH, Boylan GB, Murray DM. Mild hypoxic ischaemic encephalopathy and long term neurodevelopmental outcome - a systematic review. Early Hum Dev. 2018; 120:80-7. doi:10.1016/j.earlhumdev.2018.02.007

25. Bruckmann EK, Velaphi S. Intrapartum asphyxia and hypoxic ischaemic encephalopathy in a public hospital: Incidence and predictors of poor outcome. South African Medical Journal. 2015; 105(4):298. doi:10.7196/samj.9140

26. Franklin JD, Guidry A, Brinkley JF. A partnership approach for electronic data capture in small-scale clinical trials. Journal of Biomedical Informatics. 2011; 44:S103-S8. doi:10.1016/j.jbi.2011.05.008

27. Watt A, Eng N, editors. Database design [Internet]. 2014 [cited 2021 Jan 28]. Available from: https://opentextbook.site/informationsystems2019/chapter/chapter-8-database-design/

28. Rorie DA, Flynn RWV, Grieve K, Doney A, Mackenzie I, MacDonald TM, et al. Electronic case report forms and electronic data capture within clinical trials and pharmacoepidemiology. Br J Clin Pharmacol. 2017; 83(9):1880-95. doi:10.1111/bcp.13285

29. Babre D. Electronic data capture - narrowing the gap between clinical and data management. Perspect Clin Res. 2011; 2(1):1-3. doi:10.4103/2229-3485.76282

30. Lockery JE, Collyer TA, Reid CM, Ernst ME, Gilbertson D, Hay N, et al. Overcoming challenges to data quality in the ASPREE clinical trial. Trials. 2019; 20(1) doi:10.1186/s13063-019-3789-2

31. Barlow C. Oncology research: Clinical trial management systems, electronic medical record, and artificial intelligence. Semin Oncol Nurs. 2020; 36(2):151005. doi:https://doi.org/10.1016/j.soncn.2020.151005

32. Lu Z. Electronic data-capturing technology for clinical trials: experience with a global postmarketing study. IEEE Eng Med Biol Mag. 2010; 29(2):95-102. doi:10.1109/MEMB.2009.935726

33. Houston L, Yu P, Martin A, Probst Y. Heterogeneity in clinical research data quality monitoring: A national survey. J Biomed Inform. 2020; 108:103491. doi:10.1016/j.jbi.2020.103491

34. Cochran AR, Raub KM, Murphy KJ, Iannitti DA, Vrochides D. Novel use of REDCap to develop an advanced platform to display predictive analytics and track compliance with enhanced recovery after surgery for pancreaticoduodenectomy. Int J Med Inform. 2018; 119:54-60. doi:10.1016/j.ijmedinf.2018.09.001

35. Government of South Africa. Protection of personal information Act 4. Johannesburg: Government Gazette; 2013. p. 1-148.

36. Academy of Science of South Africa (ASSAf). Update on the code of conduct for research to stakeholders. 2021.

37. Adams R, Adeleke F, Anderson D, Bawa A, Branson N, Christoffels A, et al. POPIA Code of Conduct for Research (with Corrigendum). S Afr J Sci. 2021; 117(5/6). doi:10.17159/sajs.2021/10933

38. Shah J, Rajgor D, Pradhan S, McCready M, Zaveri A, Pietrobon R. Electronic data capture for registries and clinical trials in orthopaedic surgery: Open source versus commercial systems. Clin Orthop Relat Res. 2010; 468(10):2664-71. doi:10.1007/s11999-010-1469-3

39. Patridge EF, Bardyn TP. Research electronic data capture (REDCap). J Med Libr Assoc. 2018; 106(1):3. doi:10.5195/jmla.2018.319

40. Department of Health, South Africa. South African good clinical practice : Clinical trial Guidelines. Pretoria; 2020.

41. Teorey TJ, Lightstone SS, Nadeau T, Jagadish HV. Database modelling and design : Logical design. San Francisco: Elsevier; 2011.

42. Uduwela W, Wijayarathna G. An approach to automate the relational database design process. Int J Database Manage Syst. 2015; 7(6):49-56. doi:10.5121/ijdms.2015.7604

43. Research Electronic Data Capture (REDCap) [Internet]. Vanderbilt University; [cited 2021]. Available from: https://www.project-redcap.org/.

44. Martin-Willett R, McCormick Z, Newman W, Larsen LD, Ortiz Torres MA, Bidwell LC. The transformation of a gold standard in-person substance use assessment to a web-based, REDCap integrated data capture tool. J Biomed Inform. 2019; 94:103186. doi:10.1016/j.jbi.2019.103186

45. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap Consortium: Building an international community of software platform partners. J Biomed Inform. 2019; 95:103208. doi:https://doi.org/10.1016/j.jbi.2019.103208

46. Bowman MA, Maxwell RA. A beginner's guide to avoiding protected health information (PHI) issues in clinical research - with how-to's in REDCap data management software. J Biomed Inform. 2018; 85:49-55. doi:10.1016/j.jbi.2018.07.008

47. Grandison T, Bhatti R. HIPAA compliance and patient privacy protection. Stud Health Technol Inform. 2010; 160(Pt 2):884-8.

48. Obeid JS, McGraw CA, Minor BL, Conde JG, Pawluk R, Lin M, et al. Procurement of shared data instruments for Research Electronic Data Capture (REDCap). J Biomed Inform. 2013; 46(2):259-65. doi:10.1016/j.jbi.2012.10.006

49. Hossmann S, Haynes AG, Spoerri A, Diatta ID, Aboubacar B, Egger M, et al. Data management of clinical trials during an outbreak of Ebola virus disease. Vaccine. 2019; 37(48):7183-9. doi:10.1016/j.vaccine.2017.09.094

50. Fu J, Wang C, Yu Z, Wang J, Sun J, editors. A watermark-aware trusted running environment for software clouds. In: Proceedings of the 2010 Fifth Annual ChinaGrid Conference; 2010 Jul 16-18; Beijing, China.

51. LogicalDOC [Internet]. LogicalDOC document management. c2021. [cited 2021 Mar]. Available from: https://www.logicaldoc.com/

52. Pomerantseva V. Coding & designing a clinical database. Appl Clin Trials. 2012; 21(3):44-50,2-3.

53. Gunjal B. Database management: Concepts and design. New Delhi: Firewall Media; 2003.

54. Coronel C, Morris S, Rob P. Database systems: Design, implementation, and management: 10th ed. Boston: Course Technology Press; 2009.

55. Halpin T, Morgan T. Information levels and frameworks. In: Halpin T, Morgan T, editors. Information modelling and relational databases. 2nd ed. San Francisco: Morgan Kaufmann; 2008. p. 27-57.

56. Badia A, Lemire D. A call to arms. ACM SIGMOD Record. 2011; 40(3):61-9. doi:10.1145/2070736.2070750

57. Al-Barak M, Bahsoon R, editors. Database design debts through examining schema evolution. IEEE; 2016:

58. Consortium R. Redcap user guide for Vanderbilt/Meharry end users. 2019.

59. Harrison K, Rahimi N, Carolina Danovaro-Holliday MC. Factors limiting data quality in the expanded programme on immunization in low and middle-income countries: a scoping review. Vaccine. 2020; 38(30):4652-63. doi:10.1016/j.vaccine.2020.02.091

60. Liu Q, Feng G, Zhao X, Wang W. Minimizing the data quality problem of information systems: a process-based method. Decis Support Syst. 2020; 137:113381. doi:https://doi.org/10.1016/j.dss.2020.113381

61. van Eijk AM, Sutton PL, Ramanathapuram L, Sullivan SA, Kanagaraj D, Priya GSL, et al. The burden of submicroscopic and asymptomatic malaria in India revealed from epidemiology studies at three varied transmission sites in India. Sci Rep. 2019; 9(1):17095. doi:10.1038/s41598-019-53386-w

62. Chen H, Yu P, Hailey D, Cui T. Application of a four-dimensional framework to evaluate the quality of the HIV/AIDS data collection process in China. Int J Med Inform. 2021; 145:104306. doi:10.1016/j.ijmedinf.2020.104306

63. Nagle T, Redman T, Sammon D. Assessing data quality: A managerial call to action. Bus Horiz. 2020; 63:325-37.

64. De S. Hybrid approaches to clinical trial monitoring: Practical alternatives to 100% source data verification. PPerspect Clin Res. 2011; 2(3):100-4. doi:10.4103/2229-3485.83226

65. Nahm ML, Pieper CF, Cunningham MM. Quantifying data quality for clinical trials using electronic data capture. PLoS ONE. 2008; 3(8):e3049. doi:10.1371/journal.pone.0003049

66. Houston L, Martin A, Yu P, Probst Y. Time-consuming and expensive data quality monitoring procedures persist in clinical trials: A national survey. Contemp Clin Trials. 2021; 103:106290. doi:10.1016/j.cct.2021.106290

67. European Medicines Agency (EMA). Reflection paper on risk based quality management in clinical trials. [Internet]. 2013. [cited 2022 May|]|. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en.pdf

Chapter 3: Database Setup

This chapter will address the study aim of the design and maintenance of a national-level database capable of capturing and storing multiple data points across several sites.

3.1 Introduction

This chapter aims to describe the construction of the NESHIE REDCap database from the induction of the NESHIE study in 2019 to the introduction of the A7 database in September 2020. The database's features have allowed the database to evolve in form and function thus making it a platform not only useful for data capture but also for identifying data errors. Aspects of this process performed by other members of the NESHIE team and that do not constitute original work for the purposes of this MSc degree will be described in the Materials and Methods section.

According to the International Conference on Harmonisation-Good Clinical Practice (ICH-GCP), data should be accurate, complete, legible and collected in a timely manner [1]. Databases form part of information systems which cater for data collection, storage, retrieval and administration. While an information system exists in four phases, namely requirements analysis, and conceptual, logical and physical design, an external factor interface design is also important to consider. At this level, end-users interact with the system by entering data values, performing changes to the data, or deleting the data in a screen version of a table or form. Functions at the external level are developed by the designer into operations in the system at the physical level [2].

A complete information system comprises people, software, hardware, the database(s), application programs and procedures. The people involved in an information system may range from the end-users (the users of the data) to the system engineers and database administrators, depending on the infrastructure in use [1]. Application programs are used to transform data into useful information while databases are the repositories in which data is stored.

An organised collection of related information but more importantly, the manner in which information is organised for efficient use, qualifies as a database. A database stores information with the main objective of making access to information easy, fast, flexible and inexpensive [3]. End-user data (raw facts) and metadata (data about data) are important

elements that appear in a database. While end-user data and metadata are consolidated and maintained in the shared, integrated computer framework that is the database[1], end-user data forms the focus of this dissertation.

Every database must contain basic properties to be considered useful. A database must:

- Be a collection of facts or data elements that represent real-world information;
- Be logical, comprehensive, and internally consistent;
- Be designed, constructed, and filled with data for a distinctive purpose;
- Have each data point stored in a field; and
- Have a combination of fields which constitute a table [1].

Even a good database management system will perform poorly if a database is not adequately designed. Proper and efficient database design requires that the designer or architect precisely identifies the database's expected and required use(s) [1]. The emphasis of a good design rests on exact and consistent data[1, 4]. The main goals of database design are:

- To meet specified information content requirements;
- To offer a way to structure information that is both simple and easy to understand; and;
- To sustain performance goals and processing requisites including storage space and processing time [3].

This type of database design is outcomes-centred and serves two primary purposes; firstly, to accurately collect all necessary information and secondly, to provide analysis support [5]. Since database design is an ongoing process and requires continuous refinement in order to achieve the desired outcome, the requirements analysis, conceptual design, logical design and physical design elements can be viewed as a cyclical process (Figure 4).

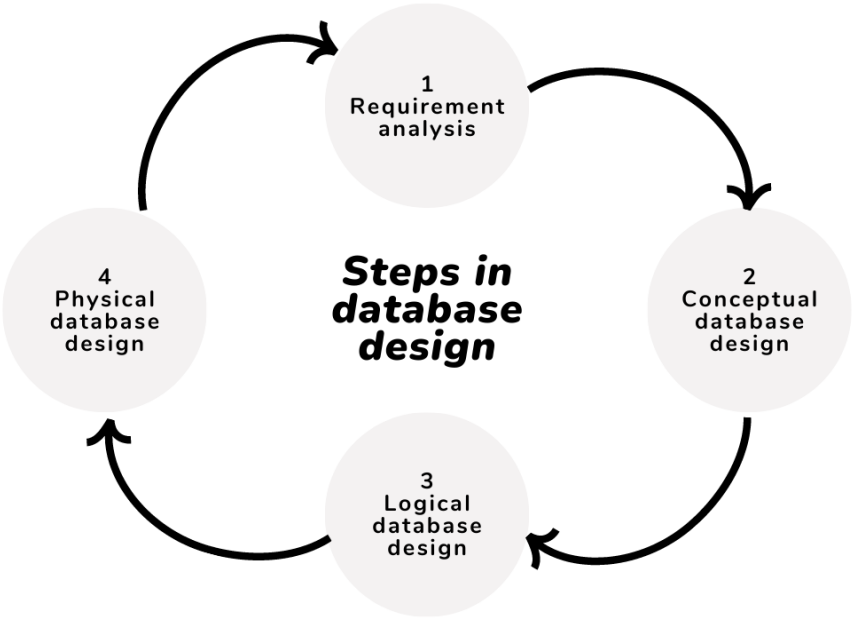

Figure 4: The cycle of database design phases. The database design starts with the requirement analysis followed by the conceptual, logical and physical design steps.

3.2 Materials & Methods

3.2.1 Study design

The NESHIE study is an observational study that is collecting clinical data and samples from neonates, as well as their parents, affected by moderate to severe NESHIE in South African hospitals. The study aims to better understand the clinical and molecular aspects of NESHIE. Through this, it is hoped that biomarkers for disease will be identified, thereby aiding in improved patient management and treatment practices. The database through which the data is collected focuses on medical history, demographics and other clinical data. It also allows for secure storage and analysis of the data.

3.2.1.1 NESHIE study Inclusion & Exclusion criteria

The NESHIE study was initiated by several neonatal and maternal health specialists with the following inclusion and exclusion criteria having been decided on:

1. Have a minimum weight of 1800g **AND** more than 36 weeks gestation;
2. Have a blood gas in the first hour greater than 16 with a pH 7 or less **OR** a blood gas in the first hour greater than 10 with a pH less than 7.15 and a peripartum/ sentinel

event **OR** a 5-minute Apgar score of less than 7 **OR** Required resus/ assisted ventilation at 10 minutes.

3. At least one sign of encephalopathy (lethargy, stupor, or coma) **AND** a Thompson score of at least 7 **OR** Seizures.

Presence of at least one abnormal sign (hypotonia, abnormal reflexes or abnormal suck) were required if seizures were **NOT** observed.

If all inclusion criteria were met, participants were screened for exclusion criteria, which included:

- The neonate not cooled within 6 hours of life or not cooled at all;
- The mother was under the age of 18;
- Consent not obtained because the parents refused, the neonate died before consent could be obtained or for other reasons;
- Encephalopathy primarily due to non-hypoxic cause;
- Hypotension or bleeding that is not responding to treatment;
- An aEEG was not performed;
- A congenital infection or abnormality;
- A surgical anomaly or suspected chromosomal abnormality; and
- Presence of asystole, severe PPHN or was moribund and unlikely to benefit from cooling.

3.2.1.2 Case Report Forms (CRFs) & their Evolution

CRFs were constructed as part of the NESHIE study by experts on neonatal and maternal health, including placental pathology, within South Africa. The neonatologists involved in creating the Neonatal CRFs included the following experts: Dr. Shakti Pillay (UCT), Prof. Alan Horn (UCT), Dr. Gugu Kali (SU), Dr. Khomotso Masemola (UP), Prof. Melantha Coetzee (UP), Prof. Daynia Ballot (Wits), Prof. Sithembiso Velaphi (Wits), Dr. Firdose Nakwa (Wits) and Dr. Rupesh Bhoola (UKZN). The Obstetrics and Gynaecology team responsible for creating the Maternal CRFs included Prof. Eckhart Buchman (Wits), Prof. Priya Soma-Pillay (UP), Prof. Tasneem Adams (Wits) and Dr. Valerie Vanneval (UP). The placental pathologists responsible for creating the CRFs for the placental pathology included Prof. Colleen Wright

(SU) and Prof. Pawel Schubert (SU). In each iteration of the CRFs and the associated database, these experts were involved.

3.2.1.3 Ethical Considerations and participant enrolment

Since the NESHIE study is a multi-site national study, several approvals were provided through the following institutions:

- University of Pretoria (HREC approval number 481/2017) associated with Kalafong and Steve Biko Academic Hospitals;
- Stellenbosch University (HREC approval number N1803041_RECIP_UP-481/2017) associated with Tygerberg Hospital;
- University of the Witwatersrand (HREC approval number M180541) associated with Charlotte Maxeke, Johannesburg Academic and Chris Hani Baragwanath Hospitals; and
- University of Cape Town (HREC approval number M180541622/2018) associated with Groote Schuur and Mowbray Maternity Hospitals.
- National Research Database approval was obtained for Gauteng (reference number GP_20180_022) and the Western Cape (reference number WC_201803_009).

For the purpose of this dissertation, approval was obtained from the Faculty of Health Science Research Ethics committee at the University of Pretoria under application number 739/2020 (Appendix A).

Informed consent was obtained from all the participants in the study. The study protocol adhered to the declaration of Helsinki and applied Good Clinical Practice (GCP) principals. Per Figure 5, participants were only approached for consent into the NESHIE study if all inclusion criteria were satisfied and no exclusion were present. Prior to obtaining consent for enrolment into the study, potential participants received counselling as part of standard of care. A minimum of 24 hours was given for this purpose. Following counselling, parent(s) were approached for consent. This consent covered both the parent(s) and the baby's inclusion into the study. Participant enrolment was considered complete upon written consent. If consent was not obtained, the patient was not recruited into the study, but the screening data was retained. If consent was provided, data and sample collection could be initiated.

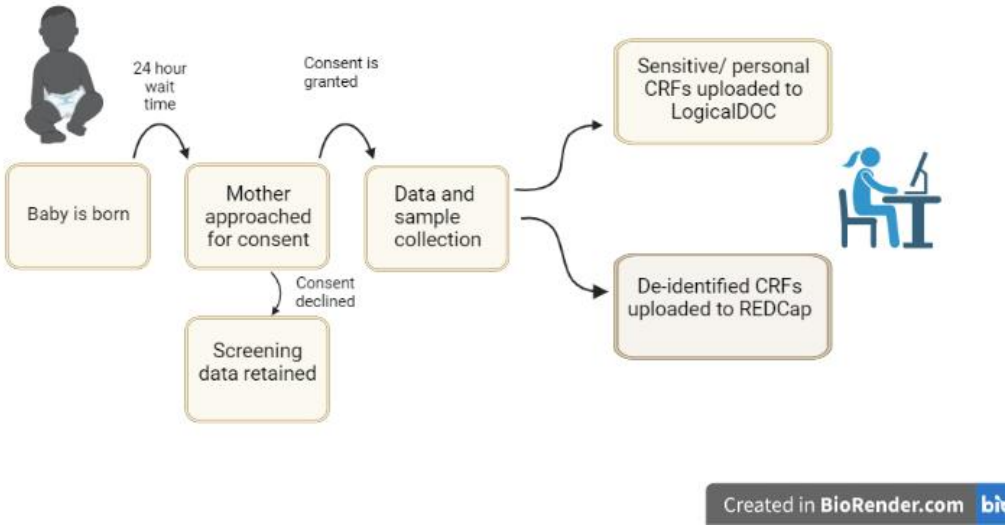

Figure 5: A depiction of the consenting process in the NESHIE study. Twenty-four hours after a neonate fitting the enrolment criteria is born, the mother is approached for consent to be enrolled in the study. If it is not granted, only the screening data is retained; however, if consent is given, data and samples are collected, and the accompanying case report forms are uploaded to the relevant platforms.

3.2.2 Security and management

Data security was ensured in line with the relevant University of Pretoria (UP) data security policies. This included working with UP IT to ensure remote access to the database in a controlled manner. The term "database" here refers specifically to the NESHIE REDCap platform, which was used for data management throughout the study. Access to the database was also password controlled. All data collected in the study was managed according to the NESHIE data management plan. (This plan will not be included in this project owing to size but can be provided on request.)

3.2.3 Database design

A beta-version REDCap database had been established before the commencement of this MSc project. The work described in this chapter is based on this foundation. The database for the NESHIE study was built using the online designer and offline data dictionary functionalities for REDCap version 11.3.3 (Vanderbilt University, Nashville, TN).

As shown in Table 2, the requirements analysis, conceptual/logical design, and physical design were repeated for each iteration of the database. The changeable aspects of the

requirement analysis were limited to adjustments of the CRFs as well as to the design of the REDCap database.

Table 2: The phases of database design in the evolution of the database of NESHIE study

Database design Phases	**Amendment 3**	**Amendments 4/5**	**Amendment 7**
Requirements analysis (CRFs and database)	NESHIE working group	NESHIE working group	NESHIE working group
Conceptual/logical design (Database)	Jeanne van Rensburg	Jeanne van Rensburg and T Kalua	T Kalua
Physical design (Database)	Jeanne van Rensburg	Jeanne van Rensburg and T Kalua	T Kalua

To meet the target enrolment of 2500 neonates over 5 years, it was required that 500 neonates be recruited a year across seven sites. At an estimated 1500 data points per neonate-mother pair, it was anticipated that there would be 750 000 data points that required monitoring annually. Data queries would be raised daily. The expected response time varied depending on the type of query. For queries requiring uploads and minor adjustments, the expected response time was 24-48 hours. However, for significant adjustments to clinical data the expected response time could vary from 48 hours to 14 days. The response times were anticipated to depend on a number of factors, the most prominent of which was expected to include access to hospital records. Other factors likely to affect response times included the electronic capturing of corrected data and uploading of any corrected study-associated documents.

To evaluate the functionality of the REDCap database, a secure test environment was used. The offline data dictionary mode was used as the primary mode of establishing the framework for the database, with the online designer mode used to make adjustments to the post-production database in order to prevent data loss. The Excel CSV format was used to save the data dictionary, which was then uploaded to REDCap. This conveyed all design features and associated changes to the test project.

Due to the size of the database (approximately 1500 variables are collected for a single neonate-mother pair), only select elements of the neonatal clinical data will be presented for the purposes of this dissertation. More specifically, only the screening sheet and

inclusion/ exclusion criteria form, and select sub-sections of the neonatal clinical data sections are included in this chapter.

3.3 Outcomes and findings

3.3.1 Study design

Figure 6 shows that the initial beta-version REDCap NESHIE database was built to reflect the data contained within the Amendment 3 (A3) CRFs. A soft launch of the beta-version was performed on the 19th of May 2019 for a number of reasons. These included establishing data capture and storage capacity, database usability and accuracy relative to the CRF and database troubleshooting. During the months that followed the initial construction & launch of the database, investigators entered data into the REDCap NESHIE database. This process was under continual evaluation and adjustments made according to the feedback received. The initial database was then fully launched following preliminary adjustments.

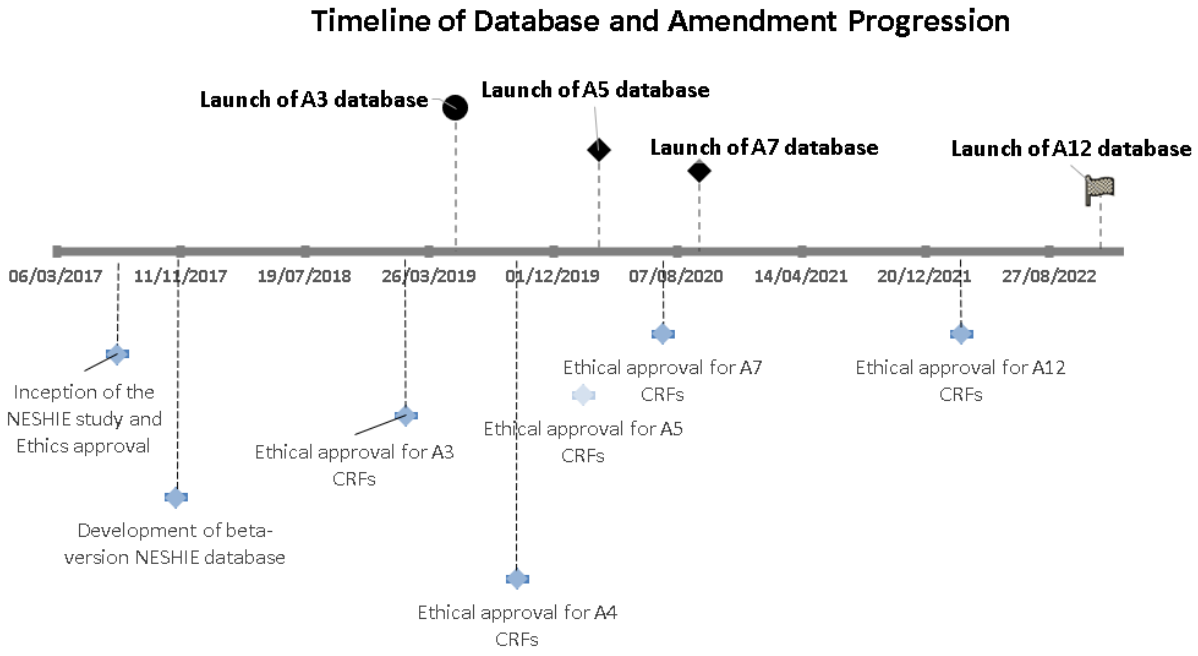

Figure 6: The progression of database and amendments throughout the NESHIE study. After the inception of the study in 2017, the A3 database was developed in 2019 and the first data was collected and uploaded to REDCap.

In the sections that follow, each iteration of the NESHIE database will be presented. The requirements analysis, conceptual, logical and physical design elements will be demonstrated throughout.

3.3.2 A3 REDCap database

3.3.2.2 Requirements analysis

At the onset of the NESHIE study, an informal requirements analysis conducted by the clinical teams determined the permanent needs for the study. These included firstly, that people in charge of collecting and capturing data for the NESHIE study (site appointees) were placed at each site. Secondly, all data requiring capturing needed to be reflected on the CRFs. Thirdly, site appointees were to capture the CRF-related data, both written and electronic, and upload the CRFs to the REDCap database. Fourthly, study sites could only access data generated from their site, in other words, restrictive data access. Fifthly, to make the capturing of data more manageable in both the written and electronic formats, the CRF-linked data needed to be sectioned. This sectioning needed to be reflected in the database. Lastly, the NESHIE study was to monitor patient records, with various export options for data validation. It was understood that a limited number of people would need access to the entire data set for specific purposes. During initial discussions, the only brief given regarding technical specifications was that the A3 database mimic the A3 CRF as closely as possible.

3.3.2.3 Conceptual design

The key points the initial beta-version database sought to address were to:

- Create a data capturing platform capable of housing multiple data points;
- Ensure data storage capability;
- Foster a user-friendly interface for easy data entry;
- Test usability, functionality, and accuracy of the database;
- Ensure logical flow in data capturing; and
- Determine weaknesses in database design that could be adjusted and improved on.

These points were in keeping with the requirements analysis and were also fundamental in constructing the logical schema essential for the physical building of the REDCap database.

3.3.2.4 Logical design

Define the main data

In defining the main data, the data fields within the CRF were sectioned. These sections included the: 1) screening sheet and inclusion/ exclusion criteria form, 2) consent documents and community engagement form(s), 3) neonate clinical data, 4) maternal and obstetric data, and 5) placental pathology data. Within the neonatal clinical data, the following sub-sections were defined: 1) history and delivery sheet, 2) precooling assessment and cooling method, 3) daily general monitoring (Days 1-4), 4) hospital course, discharge and follow-up, 5) Thompson and Sarnat score sheet and 6) follow-up data. Neither the maternal and obstetric data nor the placental pathology data were sub-divided.

Data elements characterising data

According to the conceptual design, specific database design elements were assigned to each data section. The variable name, field type, field label and field choices were established for the relevant sections and sub-sections. It was acknowledged that variable names would be unchangeable after the database was placed into production. This would ensure that there was continuity across the database iterations and also safeguard against data loss. Consequently, the naming of variables had to be clear and concise for the purposes of data analysis. Field annotations were also used as deemed necessary. Furthermore, it was necessary for users to be able to select single data options, multiple data options or provide typed responses. These typed data fields took the form of numeric or written responses. Several of the single data options needed to include “Yes/No” fields. Date and time fields were also a requirement and would be used for auto-calculations of variables such as age and BMI. Where possible, it was necessary to include the units of the numeric data fields.

In further detail, data validation rules were implemented to ensure the quality and integrity of data. For data fields such as the neonate’s date of birth and the date of neurological assessment, a format of YYYY-MM-DD HH: MM was accepted. Range validation was used to make sure that a data point such as the neonate’s temperature during cooling fell within a specified range (e.g. 33.5 – 34.5℃). The consistency validation ensured that the data entered into the database was consistent across fields or instruments. For example, the

neonate's date of birth entered into the *Screening sheet* was matched against the date of first informed consent discussion in the *Neonate consent form* to make sure that the consent was only obtained 24 hours following the birth of the baby. The required field validation was essential in the *Screening sheet* and *Inclusion/Exclusion criteria* instruments to ensure that data fields were completed before the subsequent sub-section, section or instrument became available.

As shown in Table 23 (see Appendix F), there were 33 variables for the screening sheet and inclusion/ exclusion criteria form. While not reflected in Table 23, there were five variables marked as identifier fields. These included: birth_site_province, birth_site_sub_district, birth_site_institution, mat_home_province_yesno, and mat_home_province. Field notes were used for only five variables as follows: 1) "A = Amendment" (neonate_crf_v_no), 2) "only if early gas not available" (apgar5min_yesno), 3) "complete only if clinical seizures were present prior to or at the time of cooling" (clin_seiz_descrip), and 4) "select all that apply" (other_abnorm_signs and excl_crit_descrip). The only field annotation used was the @HIDEBUTTON for the dob_tob_neonate variable. The field validation functionality was also used once for the dob_tob_neonate variable (datetime_dmy). The data validation element was used once for the dob_tob_neonate variable to ensure that data from babies born before the study inception date (01/05/2019) was not captured to the database. Since all completed CRFs and study-associated documents required uploading onto REDCap, the field upload element was activated for each section and sub-section of data. This remained a fixed function throughout all iterations of the database.

The neonatal clinical data sub-sections reflected in this chapter, include: 1) limited aspects of the history and delivery sheet and, 2) precooling assessment and cooling method. Subsequently and as shown in Table 24 (see Appendix F), there were 114 variables in the neonatal history and delivery sheet and 58 variables in the neonatal precooling assessment and cooling method, bringing the total number of variables to 172. Identifier fields were not used since no variables in these sections included identifying fields.

There were 45 variables that used field notes while the @HIDEBUTTON field annotation was used for 13. The text validation type functionality included "integer" (13 variables), "date_dmy" (1 variable), "number" (28 variables), and "datetime_dmy" (12 variables). The

text validation maximum and minimum functionalities were used to specify a range for 39 variables (see Appendix B).

Data stored and used in database

Data storage and usage relies on database structure, field types and data validation as shown in Tables 2 and 23 (see Appendix F). Choosing the correct field types such as numeric, text and dropdown options ensures that data is stored accurately and efficiently. Similarly, data validation, for example a date/time assignment for a text field, is crucial for defining validation rules for the purpose of enforcing data quality.

Data access and processing

User roles and their access levels to the data were determined by the management team and data access permissions were assigned accordingly. This ensured that users could only access and modify the data based on their designated roles in the study. At the study onset, considerations were made regarding various database functionalities, including manual data entry, data capture to CRFs, the ability to generate customised reports, and export data functionalities. These functions were enabled where applicable and in relation to user roles in the study.

Regarding data access and processing, compartmentalisation of the data for each of the seven sites was introduced to satisfy the data access requirements. Data access was restricted according to an individual's role in the study. Conceptually and logically, this resulted in the following roles: study PI, project manager/ database manager, site PIs and appointees, placental pathologists, and placenta pathology QC team. It was also necessary to consider that those responsible for performing quality control would need to be given less restrictive levels of data access, while individuals entering data would need more restrictive levels of data access.

Data constraints for data integrity

Data integrity rules were implemented to ensure that changes related to data were handled correctly and consistently. While not included in Tables 23 and 24 (see Appendix F), this involved creating minimum and maximum ranges for select variables, for example, birth weight. This ensured that captured data fell into generally accepted parameters linked to

that particular variable. For example, a minimum birth weight of 1800g and a maximum of 6000g was implemented.

In addition, creating a data audit trail and logging changes to the data allowed for traceability and identification of erroneous modifications, as well as monitoring of data access. These logs were and continue to be downloaded on the first day of every week.

3.3.2.5 Physical design

Shared design elements (logical and physical design)

To ensure data privacy and security, the database was protected from unauthorised users by defining and refining users and security groups in multiple ways. Users within a given data access group (DAG), described in Table 3, were given view and edit privileges and access to specific records unless otherwise specified.

Table 3: List of DAGs in the NESHIE study

User rights	**DAG**	**Individual(s) included**
Study PI	Access to all data groups	The principal investigator responsible for the entire project
Project manager	Access to all data groups	The person responsible for the day-to-day operations of the project
Data manager	Access to all data groups	The person responsible for overseeing the generated data and implementing initial data quality checks
Site investigators - **Site 1** - **Site 2** - **Site 3**	 groote_schuur mowbray_maternity tygerberg	The site investigators from sites 1-7 including those responsible for capturing data as well as the principal investigators at each site

- **Site 4** - **Site 5** - **Site 6** - **Site 7**	kalafong steve_biko charlotte baragwanath	
Placental pathologists	Site-specific read-only access limited to the placental pathology data, maternal and obstetric and neonatal inclusion and exclusion criteria was provided. DAGs included: groote_schuur mowbray_maternity tygerberg kalafong steve_biko charlotte baragwanath	The experts tasked with site level placental histology reporting
Placenta pathology QC team	Access to all data groups with read-only access limited to the placental pathology data	The experts responsible for conducting quality control on the generated placental histology reports

Data grouping and input control elements

The data grouping and input control elements used in the A3 database were limited to branching logic and custom alignment. While custom alignment was only used on three variables, including maternal_age, gravidity and no_of_fetuses, branching logic was used extensively in a simple format to establish functionality of the database. Branching logic was used to create a more ordered flow of information and followed REDCap syntax requirements. In this process, variable names were placed within square brackets followed by a logical response that would produce a set output. For example, the branching logic [gest_age_yesno] = '0' represents the gestational age requirement for the study (at least 36 weeks; [gest_age_yesno]), while the '0' represents "No". This logic instruction therefore indicates that a variable would only be visible to an end user when the gestational age was selected as being less than 36 weeks. Alternatively, if [gest_age_yesno] = '1', where '1' represents "Yes", then this instruction would make the associated variable visible to an end

user when the gestational age is at least 36 weeks. While Figure 7 shows the end user profile of the branching logic for this example, Table 25 (see Appendix F) shows an example of the branching logic limited to the screening, inclusion and exclusion instruments.

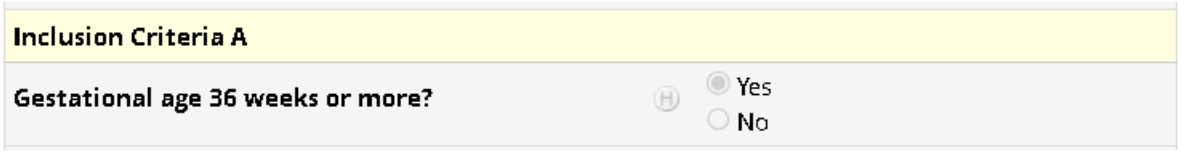

Figure 7: Simple branching logic in the A3 database

Data partitioning elements

While repeatable instruments were not in use in the A3 database, a single data arm was established for the NESHIE study in which approximately 1500 data fields were created to capture the necessary information. As shown in Figure 8, these data fields were organised into eight different events: Neonate data, Maternal data, Daily General Monitoring, Follow up, Hospital course and discharge, Uploads and Withdrawal.

	Event #	Days Offset	Offset Range Min / Max	Event Name	Custom Event Label (optional)	Unique event name (auto-generated)
	1	0	-0/+0	Epidemiological Data		epidemiological_da_arm_1
	2	0	-0/+0	General Daily Monitoring: D1		general_daily_moni_arm_1
	3	0	-0/+0	General Daily Monitoring: D2		general_daily_moni_arm_1b
	4	0	-0/+0	General Daily Monitoring: D3		general_daily_moni_arm_1c
	5	0	-0/+0	General Daily Monitoring: D4		general_daily_moni_arm_1d
	6	0	-0/+0	Patient Follow-up		patient_followup_arm_1
	7	0	-0/+0	Upload: NESHIE study documents		upload_neshie_stud_arm_1
	8	0	-0/+0	Withdrawal		withdrawal_arm_1

Figure 8: The unique events in the REDCap database

A total of twenty-seven data instruments were assigned to their respective data events. These are indicated in Figure 9, where the green tick marks indicate into which event each instrument was assigned. Since this was implemented, any instrument not assigned to an event would not be visible to a user.

Data Collection Instrument	Epidemiological Data (1)	General Daily Monitoring: D1 (2)	General Daily Monitoring: D2 (3)	General Daily Monitoring: D3 (4)	General Daily Monitoring: D4 (5)	Patient Follow-up (6)	Upload: NESHIE study documents (7)	Withdrawal (8)
Neonate Screen Inclusion And Exclusion Criteria	✔							
Initial Informed Consent	✔							
Neonate Biobanking Informed Consent	✔							
Maternal Biobanking Informed Consent	✔							
Paternal Biobanking Informed Consent	✔							
Neonatal History And Delivery Sheet	✔							
Neonatal Precooling Assessment And Cooling Method	✔							
Neonatal Daily General Monitoring Day 1		✔						
Neonatal Daily General Monitoring Day 2			✔					
Neonatal Daily General Monitoring Day 3				✔				
Neonatal Daily General Monitoring Day 4					✔			
Hospital Course Discharge And Followup	✔							
Followup Neurodevelopmental Data						✔		
Thompson HIE Score	✔							
Placental Pathology	✔							
Maternal And Obstetric Data	✔							
Upload Community Engagement Forms								
Upload aEEG							✔	
Upload CTG Trace							✔	
Upload Cranial Ultrasound							✔	
Upload MRI							✔	
Upload Placental Pathology Information							✔	
Upload Neonate Clinical Data							✔	
Upload Neonate INA Data							✔	
Upload Neonate Bayley Evaluations							✔	
Upload Maternal And Obstetric Data							✔	
Withdrawal								✔

Figure 9: The REDCap instruments in the A3 database

Data quality control and security

Double data entry, the data resolution workflow, and the data quality module were not utilised in the A3 database. However, the user rights and access control features were fundamental for meeting the requirements analysis. As shown in Table 4, user rights and permissions were dictated according to user roles with variable degrees of user privileges observed.

Table 4: Basic user rights and priviliges in the NESHIE database

Rights	**Basic privileges e.g. data capturers**	**Intermediate privileges e.g. data manager**	**Admin privileges**
Create records		X	X
Data exports (de-identified and full data sets)		X	X
Data entry: View and edit	X	X	X
Project design and setup		X	X
Add/remove users and create new roles			X
Assign user rights		X	X
Add/edit/organise reports		X	X

Optional "other" elements

The auto-numbering of records and the randomisation modules were disabled for the A3 database. The scheduling module was however activated. Physical and password security measures were enforced, and audit trails were maintained to record database use and access. Additionally, the existence of audit trails potentially dissuaded users from making mistakes.

The database was tested extensively using simulated and test data to ensure its functionality and performance. During testing, the database was able to capture all the required information accurately and efficiently. The validation rules were also effective in ensuring the accuracy and completeness of the data. Once deemed acceptable, the A3 database was placed into production mode (i.e. launched for active research use).

3.3.2.6 LogicalDOC

During the requirement analysis phase, consideration was given to the types of documents to be stored as part of the study. It was required that de-identified CRFs be uploaded to REDCap. However, this did not account for documents containing sensitive data such as informed consent forms, which also required uploading, but needed to be kept separate from REDCap documents to ensure participant confidentiality. It was therefore further required that a separate system be implemented through which this requirement could be fulfilled. This was achieved by implementing the LogicalDOC system which mirrored the system established as part of the REDCap database. The choice to use LogicalDOC instead of a separate REDCap project was driven by its robust document management features, including the ability to clearly segregate sensitive information from other study data. As such, only a brief description of the system will be provided.

Some metadata attributes such as title, author and creation dates were considered for document identification and retrieval. User rights determining the read, write or delete permissions for different documents or folders was a key point of access control. According to the logical design phase, folder structure and hierarchy was used to organise documents logically by determining the parent-child relationships between folders and files in order to maintain consistency and organisation. Seven folders were created to represent each site hospital with sub folders to enable efficient document management.

The physical design phase, relative to this dissertation, involved implementing security settings configurations, plan(s) for regular backups to prevent data loss and editing the chosen architecture to accommodate user load and increased document volume.

3.3.3 A4/A5 REDCap database

3.3.3.2 Requirements analysis and conceptual design

Following the analysis of the A3 database, the database architects utilised feedback from the investigators as well as the NESHIE team to draw up a new list of requirements, which included the requirements of a 1) more refined way of detecting and addressing data capture errors in REDCap, 2) way to track and minimise data errors in real time; and 3) additional levels of data control to ensure data quality. These criteria were used in the

construction and refinement of the CRFs, which informed the A4/A5 database design for the NESHIE study.

The data continued to be collected by the site investigators in the manner first established at the onset of the study. The basic interface including data entry forms, data collection methods and tools, and standard operating procedures also remained the same. The access rights and permissions for different user roles established in A3 were carried over into A4/A5. Roles and responsibilities of individuals or DAGs involved in data collection also remained the same, with the exception of adding new investigators to the DAGs.

Although there were minor adjustments made to the CRF data, principally, the same data collected as part of A3 was also collected in the A4/A5 database. As such, adjustments to the variables will not be discussed in detail here but can be observed in Appendix B. However, it was necessary that data visualisation and layout be reconsidered in line with the requirements analysis. This included several major adjustments to the layout of the CRFs which included 1) the removal of fields containing personal patient information from the *Screening sheet* and its addition to a newly created sheet named *Masterlist details,* 2) editing of the *Exclusion criteria*, 3) dividing some instruments into clearer, succinct sections; and 4) rearranging data points to group similar data together. These changes were incorporated into the Amendment 4 and 5 CRFs which were introduced on September 2019 and January 2020, respectively (A4 and A5; Appendix C and D, respectively). These changes were aimed at creating clearly defined fields for data entry, simplifying data capture through the addition of tick boxes, providing clear headings to delineate the forms for clarity, and consolidating similar data points under similar content umbrellas. Key elements of the conceptual design phase were carried over from the A3 database and will not be repeated here.

3.3.3.3 Logical design

As per Table 5, and as it relates to the screening and Inclusion/ Exclusion, the five variables noted as identifier fields in Amendment 3 appear unchanged in Amendment 4/5. Field notes were used similarly as in A3 with the addition of “excl_crit_a5” to the “select all that apply” field. The field annotation @HIDECHOICE was introduced for two variables (“neonate_crf_v_no” and “recruited_yes”), the @NONEOFTHEABOVE function was used for two variables (“excl_crit” and “excl_crit_a5”), and @HIDDEN was used for 9 variables

("excl_crit_present_yn", "excl_other_consent_descrip", "other_enceph_cause_descrip", "excl_congen_abn_descrip", "excl_no_cooling_reason", "excl_crit_other_descrip", "counselled_yn_a5", "recruit_failure_a5" and "recruited_yes_a5").

While the history and delivery sheet was phased out in A3 (see Table 5 form name "neonatal_history_and_delivery_sheet"), the neonatal clinical details were migrated to the precooling assessment and cooling method section while the maternal and obstetric details were integrated into the maternal and obstetric component of the CRF. For the A4/A5 precooling section of the database, there were 45 variables that used field notes while the @HIDEBUTTON field annotation was used for 13. The text validation functionality included "integer" (13 variables), "date_dmy" (1 variable), "number" (28 variables), and "datetime_dmy" (12 variables). The text validation maximum and minimum functionalities were used to specify a range for 39 variables (see Appendix D).

HTML headings, as shown in Table 5 and Figure 10, were added to the A4/A5 database to 1) clearly present headers and data-associated prompts (e.g. prompt investigators to use the newly introduced documents), 2) provide reminders for the end users (e.g. remind investigators to upload CRFs once the data capturing was complete and where applicable), and 3) continuously orient the end users within the database (e.g. provide explanatory notes and necessary solutions when access to certain database forms was restricted). This was done in order to address the aims set out following the requirements analysis and conceptual design.

The construction of the HTML was dictated by strings of code that have been shown in Table 5. A simplified example of the HTML code adhered to the following format: <h3> text </h3>. Additions such as "<font colour = 'red'>" to prescribe font colour to the text, "div class = 'yellow'" to describe the display colour and "text-align:centre" to place the text in centre alignment are examples of the detail that was added to the HTML code to achieve the specified goal. In this case, the following data point in the sequence would not be displayed, and an HTML banner would appear, indicating further restrictions to the database. As shown in the first line of Table 5, the HTML prompt was established as a descriptive field.

Table 5: Examples of HTML usage in the A5 database

Variable / Field Name	Form Name	Section Header	Field Type	Field Label
please_complete_1	neonatal_screening_sheet		descriptive	<h3> Amendment 4 Documents are now in use! <h3 style="text-align:center;"> Form must be completed in full in order to proceed to Neonatal Inclusion/Exclusion Form. </h3>
cooling_hospital	neonatal_screening_sheet	<font size=3> Neonatal Screening Information	dropdown	Name of hospital providing cooling

However, HTML codes could also be embedded into existing variables in order to enhance certain aspects of the variable, for example, to embolden the text of the section heading (line 2, Table 5). As shown in Figure 10, the banner's location in the *Screening sheet*, the first instrument in the database, was the first item that investigators came across. The red HTML was chosen for its visual impact and effectiveness in drawing attention to the banners. Directly below it, an HTML-associated section header can be observed.

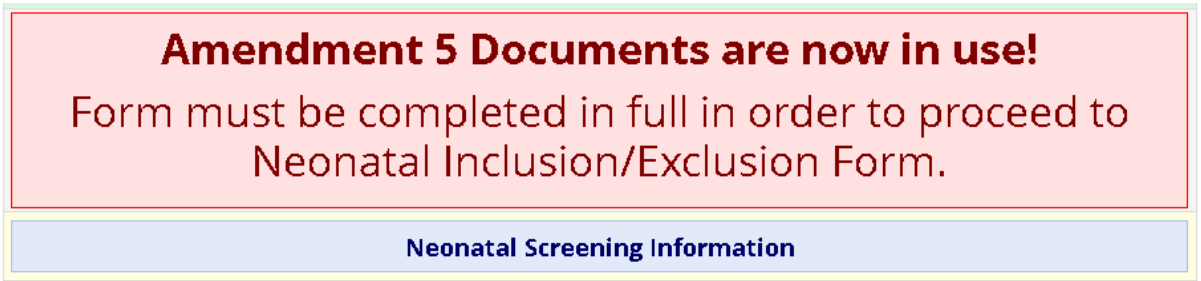

Figure 10: An introductory HTML banner in the NESHIE REDCap database together with a blue heading HTML

In addition to the examples in Table 5, HTMLs were also utilised alongside branching logic prompts. For example, and as shown in Figure 11, when screening conditions have not been

met, a red HTML prompt was displayed. Alternatively, a green HTML prompt appeared to indicate to the user that data capture could continue, and that study associated documentation required uploading.

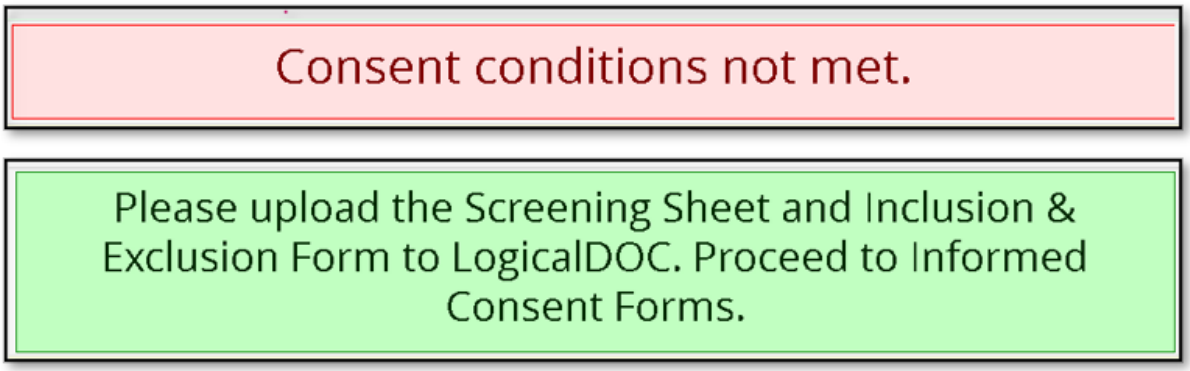

Figure 11: HTML prompts associated with screening conditions. The red banner showed that screening conditions had not been met, whereas the green banner showed that screening conditions had been met and therefore, users could proceed.

Lastly, as an example of orienting users within the database and in an attempt to reduce errors and guarantee that all necessary data is collected, Figure 12 displays an HTML prompt at the beginning of the *Main Informed Consent* form which advised users to complete the instrument accurately in order to fully access other sections of the database. This key design feature incorporated strict branching logic that prevented users from proceeding unless data values were provided for all variables. A red banner indicated that all necessary data points were not captured whereas a green banner indicated that data had been captured completely.

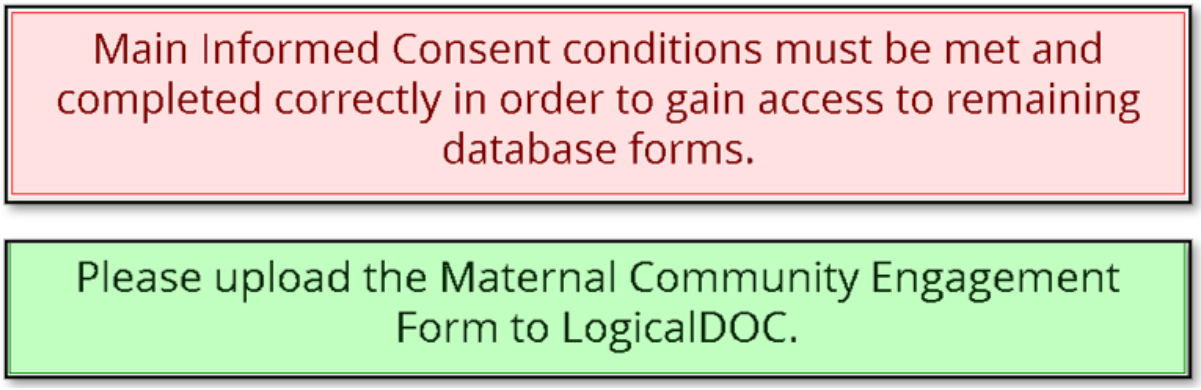

Figure 12: HTML alerts associated with data completeness. The red banner indicated that data had not been captured in full while the green banner indicated that data had been captured completely.

3.3.3.4 Physical design

In addition to fine-tuning the branching logic in the A4/A5 database, the 'matrix of fields' functionality was incorporated into the design of the database. The main driver of incorporating more complex branching logic was to establish a more structured way of capturing data to the REDCap database. While the branching logic for the *Screening sheet, Inclusion/ Exclusion Criteria* and *Consent forms* was strictly applied in the A3 database, it was refined in the A4/A5 database to become more stringent regarding data capture. While a visual example is not shown here, reasons of why it became necessary included 1) ad hoc capturing of clinical data and 2) incomplete capturing of data within data instruments in the A3 database. As shown in Table 26 (see Appendix F), an example of the more complex branching logic is displayed.

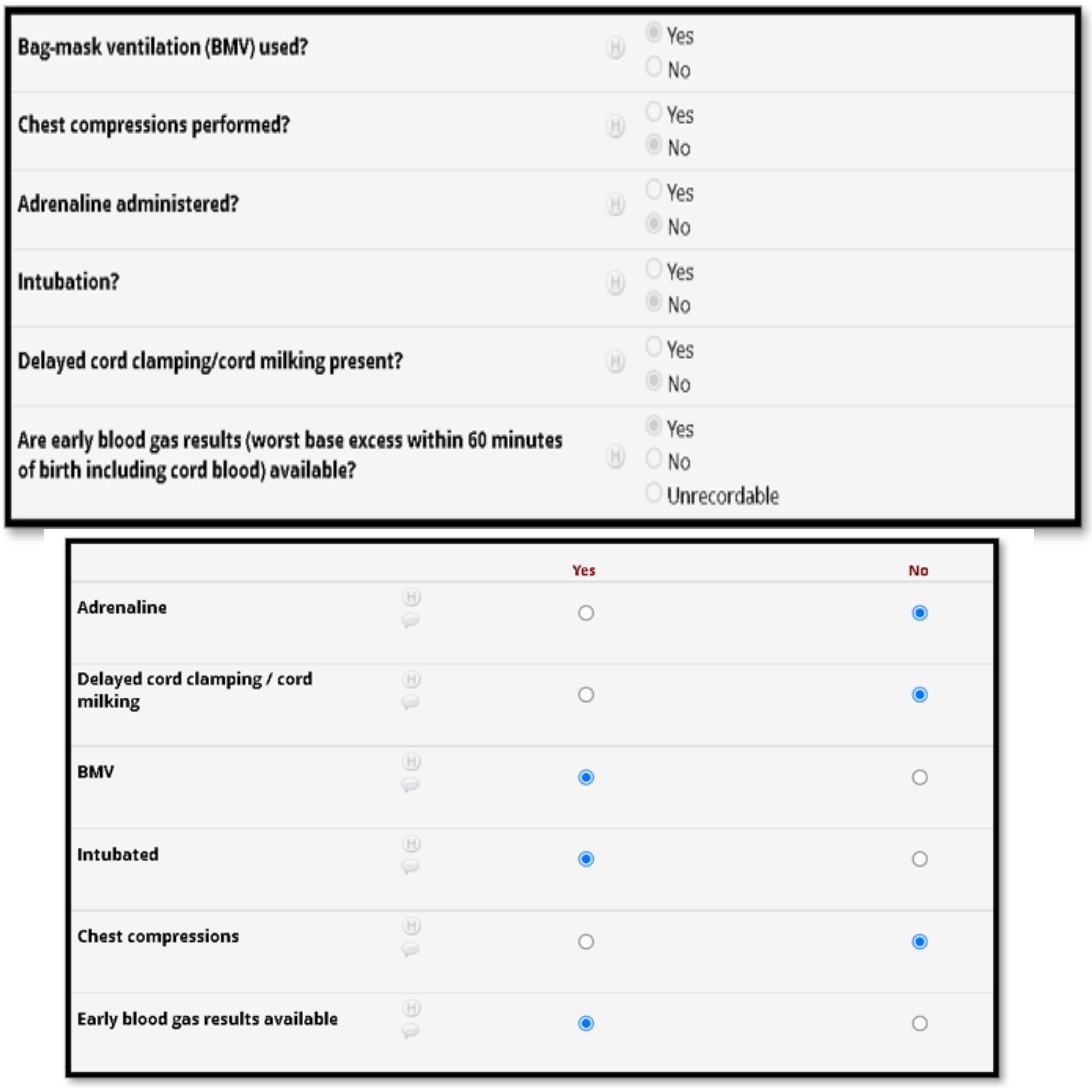

Figure 13: An example of the matrix of fields. The above sections show the arrangement of data in the A3 database, while the bottom section shows the use of matrix of fields in the A5 database.

A matrix of fields was used to group similar data types, creating an organised format for easy data capture. These organised forms were assigned matrix group names. For example, Yes/No fields were tidily grouped to enhance the appearance of the user interface and save the investigator time on data entry as evidenced above in Figure 13. The grouping of answer choices requires that a user select only one option.

Lastly, the access patterns and partitioning of data were unchanged from the previous database, however, clustering of data was adjusted in accordance with the restructuring of the CRFs. This included making adjustments to the events and re-allocating instruments within events as shown in Figure 14 The notable changes included 1) consolidating the four General Daily Monitoring events into one event with four instruments, 2) the addition of the 'Epidemiological data: Maternal' event to house the maternal data, and 3) the addition of the 'Hospital course & Discharge' as an event.

	Event #	Days Offset	Offset Range Min / Max	Event Name	Custom Event Label (optional)	Unique event name (auto-generated)
	1	1	-0/+0	Epidemiological Data: Neonate		epidemiological_da_arm_1
	2	2	-0/+0	Daily General Monitoring		daily_general_moni_arm_1
	3	3	-0/+0	Hospital Course & Discharge		hospital_course__d_arm_1
	4	4	-0/+0	Epidemiological Data: Maternal		epidemiological_da_arm_1b
	5	5	-0/+0	Patient Follow-up		patient_followup_arm_1
	6	6	-0/+0	Upload: NESHIE study documents		upload_neshie_stud_arm_1
	7	7	-0/+0	Withdrawal		withdrawal_arm_1
Add new event		Days - 0 + 0 Convert from other units		Descriptive name for this event	Custom Event Label (optional) Example: [visit_date], [weight] kg	

Figure 14: List of the updated events in the A4/A5 database

3.3.4 A7 database

3.3.4.2 Requirement analysis and conceptual design

After analysis of the A4/A5 database, a list of requirements was drawn up including 1) an advanced way to detect discrepancies and inconsistencies in the captured data, 2) an

improved way to conduct automated checks and real-time validation, and 3) a way to safeguard against the risk of erroneous or inconsistent data.

Data collection in the A4/A5 phase remained consistent with the established methodology. The basic interface, data entry forms, and procedures remained unchanged. Access rights and permissions from A3 and A4/A5 were carried over, while new investigators were added to existing roles and responsibilities. Minor adjustments were made to the layout of the data in the A7 CRFs. Specific details of these adjustments can be observed in Appendix E.

After launching the A7 database in September 2020 (A7; Appendix E), user training sessions were arranged to familiarise investigators with the new environment and adjusted CRFs. Key elements from the conceptual design phase implemented in the A3 database and maintained in the A4/A5 database were carried over into the A7 database and will not be reiterated.

3.3.4.3 Logical and physical design

As summarised in Table 6, the only significant adjustments made from the A4/A5 database to the A7 database was to the HTML prompts. Branching logic, field annotation and user rights & permissions were minimally adjusted while piping and the data quality module were implemented for the first time. All other database elements remained the same.

Table 6: A summary of the design elements incorporated throughout the evolution of the REDCap databases

REDCap Design phase	REDCap element	A3	A4/A5	A7
Logical design	Variable name	Introduced	Maintained	Maintained
	Field type	Introduced	Maintained	Maintained
	Field label	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Field choices	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Field notes	Introduced	Maintained	Maintained
	Field validation	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Data validation	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Field upload	Introduced	Maintained	Maintained
	HTMLs	Not used	Introduced	Significantly adjusted
Physical design	Matrix of fields	Not used	Introduced	Maintained A4/A5 adjustment
	Data instruments	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Data resolution workflow	Not used	Introduced	Maintained
	Auto-numbering of records	Not used	Not used	Not used
	Branching logic	Introduced	Significantly adjusted	Minimally adjusted
	Data events	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Double-data entry	Not used	Not used	Not used
	Randomisation module	Not used	Not used	Not used
	Piping	Not used	Not used	Introduced
	Repeatable instruments	Not used	Introduced	Maintained A4/A5 adjustment
	Data quality module	Not used	Not used	Introduced
	Scheduling module	Introduced	Maintained	Maintained
	Custom alignment	Introduced	Maintained	Minimally adjusted
	Data arms	Not used	Not used	Not used
	User rights & permissions	Introduced	Minimally adjusted	Minimally adjusted
Logical & Physical design	Identifying fields	Introduced	Minimally adjusted	Maintained A4/A5 adjustment
	Field annotation	Introduced	Significantly adjusted	Minimally adjusted
	Data access groups	Introduced	Maintained	Maintained
	Codebook & Data Dictionary	Introduced	Maintained	Maintained

There are 57 variables in the neonatal screening sheet, and inclusion and exclusion form in the A7 data dictionary, of which 32 are displayed in Table 27 (see Appendix F). The elements shown in Table 27 are shown to have undergone changes from A3 and A4/A5 including modifications to HTML prompts and the addition of a 'COVID-19 Status' section. While not shown in Table 27, the field annotation functionality was used in the following ways: the @HIDEBUTTON and @HIDECHOICE for 3 variables respectively and @NONEOFTHEABOVE was used for one variable. The field notes used in the A7 data dictionary were identical to the previous amendments with the exception of the "excl_crit_a5" variable which was removed. In addition to the datetime text validation, the date_dmy functionality was added for the "cov_mat_test_date" and "cov_neo_test_date" variables. Custom Alignment was used for two variables namely "neonate_study_id" and "dob_tob_neonate" (see Appendix E).

Of the 116 variables in the Precool assessment and cooling method section, the 26 that underwent notable changes from A4/A5 are shown in Table 28 (see Appendix F). These changes included refinements to the HTML prompts as well as the introduction of variables. While not reflected in Table 28, matrix group names continued to be used to arrange big clusters of data namely "clinical_dets", "assessment_day_1" and "cool_induc_method_setting". The field annotation was used for 9 variables: the @HIDEBUTTON for 8 variables and @NONEOFTHEABOVE in one instance. Custom alignment was used in only one instance ("cooling_method_used"). Field notes were used for 47 variables. The text validation type functionality included "integer" (6 variables), "number" (39 variables), and "datetime_dmy" (8 variables). The text validation maximum and minimum functionalities were used to specify a range for 23 variables.

To mitigate the risk of data loss due to power outages and unintended deletion, data backup and recovery plans were reinforced. The data can be downloaded from REDCap in partitions or as a whole. Additionally, project administrators and managers can back up and download the entire project data. The REDCap server administrators also implemented a daily filesystem as well as database dump backup to multiple sites.

Within the A7 database, if data required rearranging based on adjustments to the A7 CRF, these changes were implemented. This notably included a necessary restructuring of the informed consent instruments. Based on the conditional access structure created during the

design of the A4/A5 database, the database architects decided to implement a stricter protocol on the basis of the changes made to the informed consent documents by using more complex branching logic. Given these adjustments, it was necessary to amend the designation of instruments to the events established in the A4/A5 database. The A7 database consequently consisted of eight events and twenty-seven instruments as shown in Figure 15.

Data Collection Instrument	Epidemiological Data: Neonate (1)	Daily General Monitoring (2)	Hospital Course & Discharge (3)	Epidemiological Data: Maternal (4)	Patient Follow-up (5)	Upload: NESHIE study document (6)	Withdrawal (7)
Neonatal Screening Sheet	✔						
COVID Medication	✔						
Neonatal Inclusion And Exclusion Criteria	✔						
Informed Consent Process Checklist	✔						
Recontact Form	✔						
Neonatal Informed Consent	✔						
Parental Informed Consent	✔						
Community Engagement	✔						
Neonatal Precooling Assessment and Cooling Method	✔						
Neonatal Sample Collection Sheet	✔						
Neonatal Daily General Monitoring Day 1		✔					
Neonatal Daily General Monitoring Day 2		✔					
Neonatal Daily General Monitoring Day 3		✔					
Neonatal Daily General Monitoring Day 4		✔					
Hospital Course Discharge and Follow-up			✔				
Thompson HIE Score	✔						
Maternal And Obstetric Data				✔			
Followup 20 Week INA					✔		
Followup Bayleys					✔		
Placental Pathology				✔			
Cerebral Ultrasound Report Form	✔						
Upload Checklist						✔	
Upload Screening Sheet						✔	
Upload Sampling Kit Requisition Forms						✔	
Upload aEEG(s)						✔	
Upload Cranial Ultrasound(s)						✔	
Upload MRI						✔	
Upload Neonate Clinical Data						✔	
Upload Maternal and Obstetric Data						✔	
Upload Placental Pathology Information						✔	
Upload CTG Trace(s)						✔	
Upload Neonate Follow-up Data						✔	
Withdrawal							✔

Figure 15: A7 events and instruments

Regarding the significant changes to the HTML prompts, the following notable adjustments were implemented: 1) addition of several new HTMLS, 2) consistency with regard to colour, font and alignment in relation to the information the HTMLs convey, and 3) reorganization of the HTML content to streamline readability and facilitate intuitive navigation. As shown in Figure 16, the use of yellow was intended to draw the user's attention to vital information that required action. Red banners signified caution or indicated a need for action, while green banners signalled that the user could proceed to the next form or instrument. As introduced in the A4/A5 database, blue banners were used to indicate data-associated section headings more clearly.

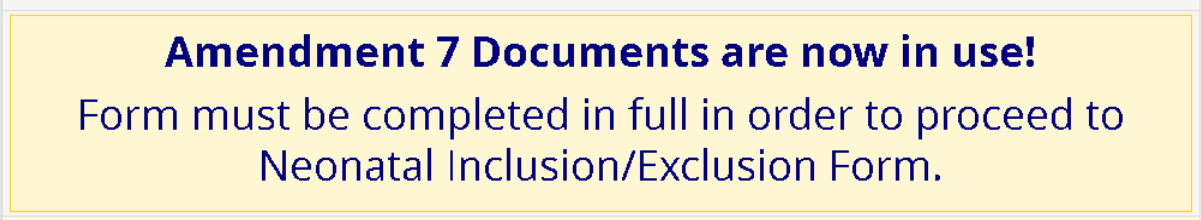

Figure 16: Introductory banner to the A7 database

While the A3 database had four instruments for the purpose of capturing relevant consent-associated data, the A7 database promoted the process of obtaining consent by streamlining it into three instruments because of a restructuring of the study's informed consent forms and associated adjustment of the informed consent process. As shown in Figure 17, an example of where these adjustments were pivotal included the introduction of HTML banners flagging cases in which consent conditions were not met. While strict branching logic conditions had always been in place regarding consent conditions, HTML prompts had not previously been used to signal when the necessary conditions were not met. This resulted in users not easily being able to identify where data may have been inaccurately captured and prevented further action on the participant's data record.

Similarly, access to data fields subsequent to the completion of the consent section was always contingent on correct and complete data entry in previous fields. For example, if consent for the collection of the placenta was not provided, then access to the *Placental Pathology* instrument would be restricted. Where appropriate, access to subsequent data fields would be contingent on the correct and complete data entry in previous fields. The uploading of required documents was only accessible if initial data fields were completed

correctly and fully. As a result, the database effectively controlled access to data entry based on consent conditions. These restrictions were however not previously obvious to the end-user. HTML banners were used in the A7 database to more clearly indicate where definitive “stops” had been built into the database.

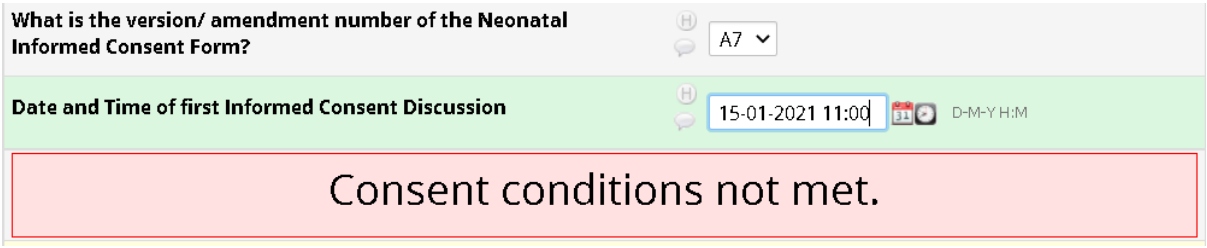

Figure 17: An HTML prompt indicating that consent conditions were not met for inclusion into the study

Similarly, while calculation fields have been in use since the A3 database, they were hidden from user view until the A7 database (Figure 18). This was done for two purposes, firstly, to help confirm the accuracy of captured ages on the CRFs and secondly, to aid in data monitoring. In Figure 18, the date and time the neonate was admitted to the cooling centre is accurate, however, the date and time at which cooling commenced is incorrect yielding an incorrect age calculation leading to a negative value being indicated.

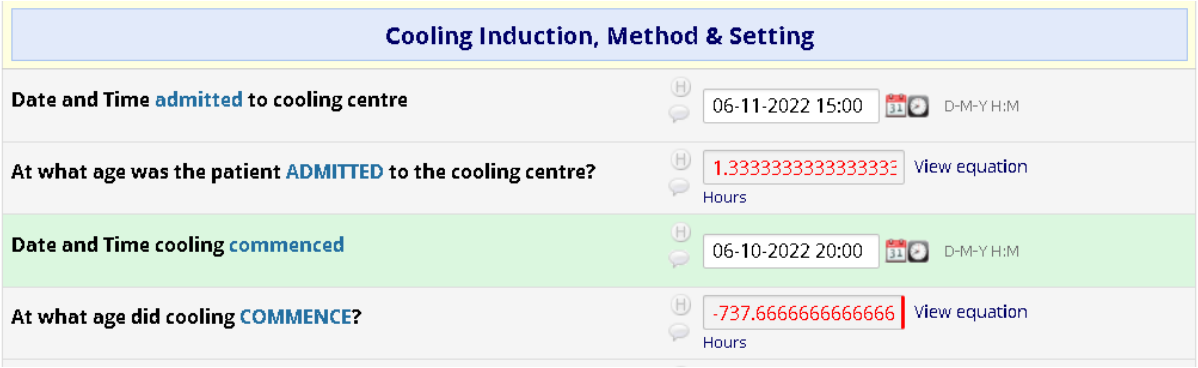

Figure 18: An example of age calculations in the Precool instrument

Lastly, piping was a useful tool used in the refinement of the A7 database. This technique involves employing variable names enclosed in square brackets [] within question or field texts. When a data form is completed, the captured variable data in the database is reflected in the specified form or location in another part of the database. This eliminates the need for participants to enter the same information twice. For example, in the *Parental Informed Consent* and *Community engagement* instruments, piping was used to tailor the header of the form based on whether consent was provided by the mother or father as seen

in Figure 19. Since the data fields in the consent section were interrelated, data could be re-used, and this ensured that related information remained consistent.

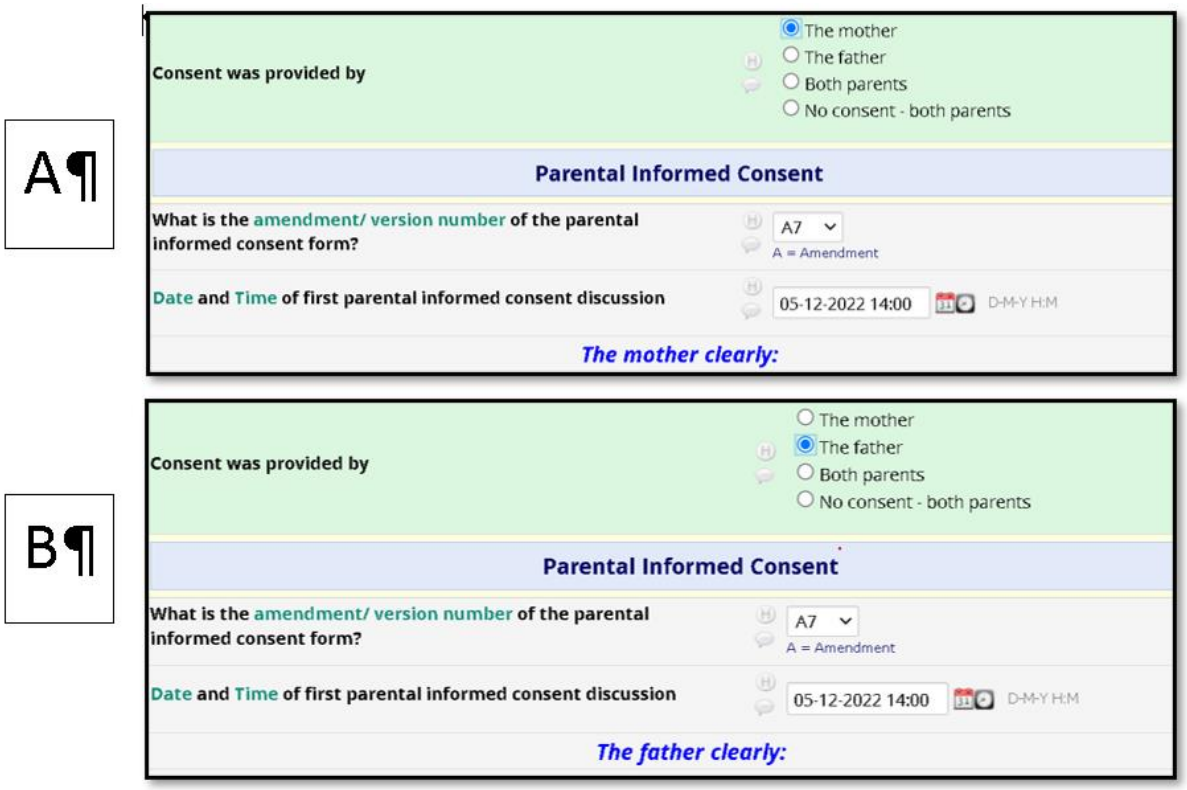

Figure 19: Piping usage in the parental consent instrument in A7

Piping was also used in conjunction with the alerts and notifications function. This allowed the database architects to set up notifications to inform the database manager of activity on the database, such as when a *Screening sheet* was uploaded. Using the piping function, the neonate's study ID was inserted into that field, making it easier to track and manage data.

3.3.5 Comparison of the A3, A5 and A7 databases

The A3 database was constructed to house the clinical data collected in the initial phase of the NESHIE study. While it was able to accommodate the volume of data produced during the NESHIE study, a requirements analysis identified several limitations. This led to the creation of a new logical schema that paved the way for the physical creation of the A4/A5 database and eventually the A7 database.

Upon migrating to the A5 database, investigators noticed a yellow banner that served as an introduction to the new design. The HTML banner served to orient and introduce users to the new database. HTML banners were also used in other ways. The addition and use of

'please upload' banners in multiple instruments provided much needed reminders to users to upload the specified CRFs to the relevant stated fields. Alerts were also incorporated into the HTML design which explained that access to the database hinged on completing the *Main Informed Consent* instrument. If an instrument did not function as expected, investigators could infer that the *Main Informed Consent* may not have been completed fully and thus take corrective measures there. The blue HTML headings and subheadings were useful as section breaks and beacons, allowing investigators to compare them to the physical CRFs in the data capture process to continuously orient themselves.

In the A5 database, branching logic was kept simple, and when used in conjunction with features such as matrix of fields and piping, it enhanced the visual display of the variables within instruments, but it also further enforced a stepwise data entry process. In several forms, once data was entered, access was provided to the rest of the forms and instrument for further data capture. The intention of this was to decrease the incidence of errors through the design of the database as opposed to re-training the investigators and implementing other interventions targeted at the end users. Additionally, the stepwise data entry created by the branching logic was further introduced, or in some instances enhanced, to make end users more aware of each data point being entered. Because the forms in the database displayed one data point at a time, investigators were more likely to pay close attention to each data point as it was entered. The matrix of fields functionality was used as far as was possible throughout the database to consolidate data and make the completion of these data points, mostly 'Yes/No' fields, easy and quick to capture and review.

The transition to the A7 database led to changes in the REDCap database and created a user-friendly interface that engaged the users. This time around, investigators were greeted by a red introductory banner reminding them to use the newly appointed A7 documents. A higher level of restriction was applied to the first two instruments containing screening and inclusion data to ensure adherence to study protocols. The consent instruments also received attention to further ensure study compliance with regard to consent conditions. To further aid this vital consent process, piping and matrix of fields were applied to the consent instruments to provide a segmented flow for stepwise data capture. The stepwise data capture flow was streamlined in subsequent instruments. And lastly, additional HTMLs were

utilised to continuously orient database users as well as to enhance the data capturing process.

Through continuous maintenance and fine-tuning of the A7 database, end-user performance was targeted to maintain timely and accurate data generation. By providing a standardised approach to data collection and analysis across multiple sites, the database helped identify trends and patterns in the data that may otherwise have gone unnoticed. Moreover, the implementation of data validation rules helped to ensure that the data entered into the database was accurate and consistent.

3.4 Discussion and conclusion

Database design is a complex and intricate process that demands substantial dedication and effort. In the realm of database development, it's not merely about skill, but about the application of those skills within a broader context, and an understanding of the data as well as how the data would be used in the project. Crafting an effective database requires meticulous consideration and purposeful integration of each feature. It is also a multifaceted endeavour that necessitates several key factors for success including skills such as a willingness to learn, attention to detail, patience, self-motivation, creativity, and logical thinking. Although these are important, they are only part of the equation. The ability to adapt these skills effectively to the unique requirements of a specific project is key.

Firstly, an often-underestimated skill in database development is the ability to critically evaluate the database throughout its development and evolution. In practice, this is more challenging than it might seem. This involves a continuous process of assessing and refining the database to ensure alignment with its intended purpose and evolving user needs, including aspects like efficiency, scalability, security, and user-friendliness. These aspects require a discerning eye and the willingness to challenge one's own decisions.

Persistence is key in navigating the complexities of database development, as it requires developing unwavering commitment to overcome challenges. However, it’s equally important to maintain a clear overarching goal. While this goal guides the design process, it's essential to periodically re-evaluate whether it still serves the project's best interests. Pursuing a predefined goal can lead to tunnel vision, potentially causing designers to overlook more efficient or innovative solutions that may emerge during development.

3.4.1 Practical implications, advantages, limitations and future directions

The practical implications of the database design demonstrated in this project are important for researchers conducting studies in various fields particularly in clinical research. By utilising the design features highlighted in this dissertation, researchers can improve their data management and potentially reduce errors that may impact the validity and reliability of their findings. Additionally, the use of HTML banners and other design elements made it easier to communicate and manage important details regarding the collection of accurate data. Overall, these findings have important implications for the design and implementation of databases in REDCap and can contribute to improving the quality of research studies.

Advantages and limitations are inherent in any database design, and the A7 database in REDCap is no exception. One of the key advantages of the A7 database design is the incorporation of step-by-step channels for data capture and branching logic, which helped ensure that all necessary data fields were completed in a structured manner. The use of HTML banners was also effective in drawing the user's attention to important information and encouraging compliance with study protocols.

Another advantage of the A7 database was its ability to enforce strict data validation rules. The use of matrix of fields functionality also aided in data capture and entry by simplifying the process and reducing the time required for completion. Furthermore, the use of shorter and split forms made it easier for investigators to enter data accurately and efficiently.

However, there were also limitations to the A7 database design. Strict validation rules, while effective in preventing errors, occasionally made it more difficult for investigators to enter data when they encountered unforeseen circumstances or exceptional cases. Furthermore, while branching logic was effective in streamlining data collection for investigators, it may inadvertently restrict the inclusion of specific data points that clinicians deemed relevant but have not been accommodated within the database's design. Additionally, the use of HTML banners while visually striking, may not have been universally effective, potentially proving distracting for some users. Moreover, the use of colour in banners could potentially create issues for colour-blind users. Further research is however required to assess these points in the context of the NESHIE study.

While REDCap offers numerous advantages, it offers fixed design features which do pose certain design limits. Additionally, the reliance on manual data entry by investigators introduces the possibility of human error, which may impact data accuracy and reliability. It is also important to note that REDCap is limited in the errors it can detect. REDCap has fixed design features, so any modifications were subject to the built-in parameters of the database, making the database design highly dependent on the skills and expertise of the database architect(s). The more skilled the architect, the more discerning the platform that could be developed.

Despite the successful implementation of the A7 database in this study, there is still room for improvement in database design using REDCap.

3.4.2 Data quality challenges and measures to improve data collection and database performance

Data entry errors, including typographical mistakes and inaccurate entries, as well as instances of incomplete and inconsistent data, were encountered during the data collection process. These errors have the potential to compromise data integrity, affect statistical analyses and consequently limit the ability to draw meaningful conclusions. Additionally, technical challenges, including network connectivity issues and server crashes, disrupted data collection for extended periods, affecting data completeness and the NESHIE study's timeline. To address some of these issues, a standardised format for data collection was established and shared with investigators. With the movement to a new amendment, training sessions were conducted to communicate the instructions regarding improvements in the data collection process as well as database orientation. Although data entry errors persist in the database, training sessions and ongoing data monitoring have been incorporated into the NESHIE study's protocol to address this issue. Stricter data validation rules were enforced to help identify and prevent data errors.

Moreover, the implementation of security measures such as access controls, and audit trails afforded a level of data security and privacy. Regular maintenance and testing of the database helped to identify and address technical issues before they could impact the accuracy and completeness of the data in the NESHIE study.

3.5 Conclusion

In conclusion, this project has demonstrated the importance of thoughtful database design in clinical research. The use of REDCap as a platform allowed for the implementation of advanced features to improve data capture, accuracy, and organisation. Implementing these features, created a user-friendly and efficient database, positively impacting data integrity and completeness.

These findings have important implications for future clinical research studies, as the use of advanced features in database design can target data quality by creating management systems and discerning platforms. Moreover, this project highlighted the importance of meticulous planning in the database design phase, as it can greatly impact the success of a study.

This project contributes to the growing body of knowledge on database design in clinical research and provides valuable insights for researchers seeking to improve the quality and efficiency of their data management practices particularly regarding maternal health and neonatal associated studies. By utilising the advanced features available in REDCap, we were able to create a database that was more accurate, organised, and efficient in engaging with end users to encourage good quality data, improving the quality of the NESHIE study's research outcomes.

The use of REDCap's features for improved database design drew the user's attention to vital information requiring action, encouraged compliance with study protocols and facilitated communication between investigators, data capturers, and the database designers. The implementation of consent requirements within the database had important implications for clinical practice. By requiring complete and accurate consent information, the database ensured that patient rights were respected and that study protocols adhered to ethical standards.

Furthermore, the efficient data capture and management facilitated by the database design can lead to more efficient clinical research practices, potentially reducing costs and eventually down the line improving the speed and accuracy of data analysis.

In conclusion, this project demonstrates the potential of REDCap as a tool for developing user-friendly and robust discerning platforms for clinical research. While the database

design iterations were successfully implemented, there is still room for improvement to further enhance its usefulness. Overall, these demonstrations show the capacity of the platform to enhance the quality of clinical research and ultimately impact patient care.

In addition to these database design advancements, this project also delved into improving the user interface to enhance its effectiveness in communicating vital details to users to aid the data capture process. The next chapter will present statistical evidence to determine whether there was a reduction in errors across database iterations, and if so, whether it achieved statistical significance.

3.6 References

1. Coronel C, Morris S, Rob P. Database systems: Design, implementation, and management. Course Technology Press; 2009.

2. Halpin T, Morgan T. Information levels and frameworks. In: Halpin T, Morgan T, editors. Information modelling and relational databases. 2nd ed. San Francisco: Morgan Kaufmann; 2008. p. 27-57.

3. Gunjal B. Database management: Concepts and design. 2003.

4. Badia A, Lemire D. A call to arms. ACM SIGMOD Record. 2011; 40(3):61-9. doi:10.1145/2070736.2070750

5. Pomerantseva V. Coding & designing a clinical database. Applied Clinical Trials. 2012; 21(3):44-50,2-3.

6. Teorey TJ, Lightstone SS, Nadeau T, Jagadish HV. Database modelling and design: Logical design. San Francisco: Elsevier Science & Technology; 2011.

Chapter 4: Analysis I

4.1 Introduction

"You can't be analytical without data, and you can't be really good at analytics without really good data" [1]. In the realm of human research, the imperative need for data quality becomes evident when striving for the execution of accurate and stringent data. In clinical research, in particular, collecting, preparing and depicting good quality data is critical [2]. "Data quality" refers to the accuracy and completeness of information systems and is crucial for organisations to get the most value from their data[3]. Laney (2018) argued that the most significant problem with poor data quality is the risk it poses to organisations that rely on data to make decisions. While data audits or monitoring are associated with good governance, they do not guarantee good quality data [1].

There is currently no "gold standard" approach to data audits or monitoring in clinical research, meaning that there is a wide variety of methodologies in use [2]. The European Medical Agency (EMA) defines central monitoring as document review, data checks and analysis conducted separately from the investigator site to assess the collected data to establish compliance, and distinguish odd data patterns, invalid data or deviations from protocol [4]. These assessments, including reporting error rates, provide objective information about faulty processes and can help prevent future errors by recognising problematic work patterns or behaviours [5].

At present, there is no universally accepted method for quantifying error rates in clinical research. Nevertheless, valuable insights can be obtained from Houston et al.'s (2018) comprehensive literature review of clinical trials. Across nine publications, an average error rate of between 10-20% was consistently reported, with one instance where an error rate as low as 0.45% was observed[6]. A descriptive study by Hong et al. (2013) found an error rate of 2.8% across all fields with individual field rates ranging between 0.5% and 6.4% [7]. While opinions may vary on what constitutes an acceptable error rate for a study, Kirch (Ed. 2008) note that an acceptable database error rate is often 0.1% and these types of errors can be decreased through data validation. However, Kirch (Ed. 2008) makes an important note that database error rates should be defined at the beginning of the study, ensuring alignment with the study's specific objectives [8]. The NESHIE study, with the goal of aligning data quality with the objectives of this project, defines an error rate of 5% as an "acceptable" limit.

With an understanding of the importance of data quality established, this chapter will focus on identifying compliance issues, invalid data, and deviations from protocol within all three iterations of the NESHIE project database. Particular attention will be given to errors that may have occurred during data capture, such as data measurement, transcription, and data entry.

4.2 Materials and Methods

To assess data quality in paper-based clinical trials, database entries were compared against data captured on CRFs, resulting in an estimate of the proportion of database errors that occurred between May 2019 and March 2022. Only data from enrolled participants were considered.

4.2.1 Error Reporting

Following the development of the A7 database and as investigators continued to input data into the database, it became necessary to curate the data and to assess the proportion of errors made per CRF instrument, record, amendment and investigator. These errors were determined at a site and study level. As shown in the previous chapter, refinements were made to the CRFs and their associated REDCap instruments. However, due to the sheer magnitude of available data points (approximately 1500 variables), it was decided for this dissertation to exclusively focus on data captured to the following instruments: Screening, Inclusion and Exclusion Criteria and Precooling.

Owing to the practical challenges of accessing hospital records for multiple study sites, the NESHIE CRFs were used as the source documents and were therefore used as the benchmark in the monitoring and error reporting process. To determine the proportion of errors, the study CRFs were checked against the REDCap entries. Per Table 7, if data was captured to both the CRF and REDCap, and there was concordance between the values, ‘1’ was entered into an Excel spreadsheet. If corresponding values captured on REDCap and CRFs didn’t match, the code “0” was assigned. If information was missing from either REDCap or the CRF, the code “555” was used. In situations where a data point was not available or not done (in the case of laboratory or point of care tests), then the code “999” was supplied, while “888” was used when, for whatever reason, a data point was unknown. Because various data fields were incrementally introduced during the study, certain data points were not present on either the CRF or REDCap, and thus the code “777” was used. It is important to note that this

code was assigned by the database architects on the backend and not by the investigators. Data for each category was presented as absolute counts.

Table 7: Codes used in the error reporting process

Code	**Interpretation**
0	Discordance between data captured on the CRF and REDCap
1	Concordance between data captured on the CRF and REDCap
555	Missing value on either the CRF or REDCap
777	Not applicable
888	Unknown value
999	Test not done (therefore, reading unavailable)

4.2.2 Data segregation

Data from the Neonatal Screening sheet, Inclusion/Exclusion criteria and Precooling instruments were segmented according to the A7 CRF format in the following way:

- Neonatal Screening sheet
    - Screening information
    - Birth site information
- Inclusion/ Exclusion criteria
    - Section A
    - Section B
    - Section C1
    - Section C2
    - Section D
    - Section E
    - Section F
- Precooling instrument
    - Clinical details of baby at birth
    - Blood gas evaluation at birth
    - Clinical examination prior to cooling

- Neurological assessment prior to cooling (not earlier than 30mins or at onset of cooling)
- CFM details
- Baseline aEEG assessment
- aEEG assessment at 6 hours of life
- Baseline laboratory investigation: on admission or as close to cooling as possible
- Cooling induction, method and setting

4.2.3 Statistical analysis

The absolute number of errors was calculated by summing the counts of missing ('555') and incorrect ('0') entries. To calculate the proportion of errors, the absolute number of errors was divided by the total number of variables. This was applied to each CRF instrument, record, amendment, and investigator, and then described at both the site and study level. The investigator assignment was based on the individual who first captured the data to each REDCap instrument.

In order to assess the variability across the aforementioned assessment categories, Chi-square analysis was performed with post-hoc analysis and Bonferroni correction. Significance was set at a p-value or adjusted p-value < 0.05.

Statistical analyses were performed using R Statistical software (v4.1.3; R Core Team 2022). Histograms and bar plots were used to visualise outputs. Graphs were created using the ggplot2 (v3.3.5; Wickham 2016) and lattice (v0.6-30; Sarkar 2008) packages in R Studio. Base R functions, together with packages such as dplyr (v1.0.8; Wickham, Francois and Muller 2022) and tidyverse (v1.3.1; Wickham et al. 2019), were used to "tidy" and structure the data[1].

We hypothesised that the proportion of errors in the A7 amendment group would be significantly lower than in previous amendments. In contrast, the null hypothesis was that there would be no significant differences between the error proportions across the amendment groups.

4.3 Results

Seven spreadsheets were created in Microsoft Excel to represent each study site; an example of which is displayed in Figure 20. The observations were arranged in columns while the variables were in rows to help visualise the raw data. Absolute counts were calculated per investigator, instrument and amendment.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	neonate_study_id	01-N-004	01-N-006	01-N-007	01-N-008	01-N-009	01-N-010	01-N-012	01-N-013	01-N-014	01-N-016	01-N-022	01-N-023	01-N-026	01-N-027	01-N-028	01-N-029	01-N-032	01-N-035	01-N-039	01-N-041	01-N-
2	Investigator	1B	1B	1B	1B	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C	1C
3	Amend Nr	A3	A3	A3	A3	A3	A3	A3	A3	A3	A4	A4	A4	A4	A4	A4	A4	A5	A5	A5	A7	A7
4	Site	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1	Site 1
5	Instrument: Screening																					
6	neonate_crf_v_no	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	
7	cooling_hospital	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	cooling_hospital_other	777	777	777	777	777	777	777	777	777	777	777	777	777	777	777	777	777	777	777	777	
9	dob_tob_neonate	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	
10	in_outborn_status	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
11	sex	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
12	birth_site_province	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
13	birth_site_sub_district	0	1	0	1	1	1	1	1	1	1	0	1	555	1	1	1	1	1	1	1	
14	birth_site_institution	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	
15	birth_site_inst_other	1	1	1	1	777	1	1	777	777	777	0	777	1	777	1	777	1	777	1	1	
16	mat_home_province_yesno	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	
17	mat_home_province	777	777	777	777	777	777	777	777	777	777	777	777	777	777	0	777	777	777	777	1	
18	Instrument: Inclusion/ Exclusion																					
19	gest_age_yesno	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
20	incl_weight_yesno	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
21	bd16_ph7_yesno	1	1	1	1	0	1	1	1	1	1	555	1	1	1	1	1	1	1	1	1	
22	bd10_ph7_pp_yesno	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	555	
23	apgar5min_yesno	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	
24	resus_supp_yesno	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	
25	incl_c_enceph_signs	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
26	thompson_yn	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
27	incl_c_seiz_yesno	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	
28	clin_seiz_descrip	777	777	1	777	777	777	777	555	777	0	1	1	777	777	777	1	1	1	777	1	

Figure 20: An example of the error reporting spreadsheet at each of the sites and instruments. The code ‘0’ represents incorrect entries, ‘1’ indicates correct entries, ‘555’ represents missing entries and ‘777’ is used for non-applicable entries.

These error reports were to be insightful in understanding and describing:

- Investigator errors;
- Instrument errors;
- Errors associated with document amendments; and
- Site errors.

4.3.1 Sample size

Between May 2019 and March 2022, data from 160 enrolled patients was assessed and is described in Table 8. Data was collected as it became available and resulted in the uneven distribution. At the beginning of the study, there was a relatively consistent use of A3 records for at least five of the sites since these documents were in use from the inception of the study. Sites 1 and 3 made use of more A4 documents than the other sites, while sites 3, 4, 6 and 7 utilised more of the A5 documents. With the exceptions of site 6 which had not made the transition to the A7 documents, site 3 which only used a few A7 documents and site 1 which

used the most A7 CRFs, there was a relatively consistent use of A7 documents across the other sites. Sites 4 and 7 had the highest number of enrolled patients. The investigator from site 3 was placed on maternity leave for an extended period and this resulted in a halt in the recruitment at this site which resulted in a limited number of A7 records in Table 8. Sites 5 and 6 had the lowest recruitment numbers in the entire study, which was reflected in the small sample sizes accessible for this investigation.

Table 8: Sample size across all sites

Site	N (A3 records)	N (A4 records)	N (A5 records)	N (A7 records)	Total no. of records
Site 1	9	7	3	11	**30**
Site 2	9	4	2	6	**31**
Site 3	8	7	6	4	**25**
Site 4	6	0	7	9	**22**
Site 5	2	3	3	7	**17**
Site 6	2	2	6	0	**10**
Site 7	7	5	6	7	**25**
Total	43	28	33	44	**160**

Table 9 displays the number of variables available in each instrument and across all the amendments and sites assessed in this investigation. The majority of variables were contributed by the Precool instrument. There were more variables made available in the Screen instrument as the study progressed, while the Inclusion/ Exclusion underwent a decrease in available variables. There were more variables in the A3 Precool instrument than in the subsequent amendments.

Table 9: Number of variables per CRF instrument, amendment and site

Instrument name	Variables	(N_{var}) A3	(N_{var}) A4	(N_{var}) A5	(N_{var}) A7
Screen	12	108	84	36	132
Inclusion/Exclusion	21	189	147	63	231
Precool	112	1008	784	336	1232
Sum variables	145				
Screen	12	108	48	24	72
Inclusion/Exclusion	21	189	84	42	126
Precool	112	1008	448	224	672
Screen	12	96	108	72	48
Inclusion/Exclusion	21	168	189	126	84
Precool	112	896	1008	672	448
Screen	12	72	-	108	108
Inclusion/Exclusion	21	126	-	189	189
Precool	112	672	-	1008	1008
Screen	12	24	36	36	108
Inclusion/Exclusion	21	42	63	63	189
Precool	112	224	336	336	1008
Screen	12	24	24	72	-
Inclusion/Exclusion	21	42	42	126	-
Precool	112	224	224	672	-
Screen	12	108	60	72	108
Inclusion/Exclusion	21	189	105	126	189
Precool	112	1008	560	672	1008

Table 10 provides a breakdown of the number of records managed by each investigator at their respective sites. At sites 1 and 2, Investigator 1A exclusively managed A3 CRFs and REDCap entries, while investigators 1C and 2C had experience with all the amendments. Investigator 1D managed A5 data, and 2D oversaw both A5 and A7. Site 3 had only investigator (3A) who was responsible for all the amendments. Investigator 4A managed the A3 CRFs, while investigator 4B managed data from A5 to A7 CRFs. Investigator 4C only worked on one patient. Investigator 5A only dealt with the early CRFs (A3 and A4), while investigator 5B was responsible for one A4 CRF. Investigator 5C was responsible for data from A4 to A7. At Site 6, investigator 6A dealt only with the A3 CRFs, while investigator 6B worked with the A3, A4 and A5 CRFs. Investigator 6C had experience with four A5 CRFs. Investigator 7A along with 7D only worked with A3, while investigator 7B was responsible for all amendment CRFs. Investigator 7C captured data from A4 to A7.

Table 10: Number of records per investigator and amendment

Site	Investigator	(N_{rec}) A3	(N_{rec}) A4	(N_{rec}) A5	(N_{rec}) A7	Total records
	1A					
Site 1	1B	4	NA	NA	NA	4
	1C	5	7	3	2	4
	1D	NA	NA	NA	9	8
	2A					
Site 2	2B	4	NA	NA	NA	16
	2C	5	10	4	1	20
	2D	NA	NA	2	6	8
Site 3	3A	8	7	6	4	25
Site 4	4A	6	NA	NA	NA	6
	4B	NA	NA	7	8	15
	4C	NA	NA	NA	1	1
Site 5	5A	2	1	NA	NA	3
	5B	NA	1	NA	NA	1
	5C	NA	1	3	7	11
Site 6	6A	1	NA	NA	NA	1
	6B	1	2	2	NA	5
	6C	NA	NA	4	NA	4
Site 7	7A	2	NA	NA	NA	2
	7B	4	3	2	3	12
	7C	NA	1	3	3	7
	7D	1	NA	NA	NA	1

4.3.2 Error Values and Proportions

4.3.2.1 Error values and proportions across Instruments

Table 29 (see Appendix F) illustrates the error proportions within the three instruments across amendments. The largest proportion of errors were typically noted in the Precool instrument; however, there was variability across sites and amendments. For example, in the Precool instrument, sites 3, 4, 5 and 7 had no incorrect entries, while Sites 1, 3, 5, 6 and 7 had no missing entries in A3. In A4, Sites 3, 5 and 6 had no incorrect entries in the Screen instrument, and sites 2,3 5, 6 and 7 had no missing entries. Generally, there was a trend of decreasing error proportions in the Screen instrument across amendments, except for Sites 1, 2 and 7, where the proportions peaked in A4 and then decreased thereafter. The Precool instrument consistently displayed higher error proportions than the other two instruments.

At Site 1, the error proportion due to incorrect entries was highest in the Screen instrument during A4 and A5, with an error proportion due to missing entries of 0% in A3 and A5 but reaching 1.2% in A4. The Inclusion/Exclusion criteria showed low error proportions ranging from 0.6 to 2.7% for both incorrect and missing proportions. The Precool instrument had a maximum of 6% error proportion for missing entries and 3.3% for incorrect entries in A4, but both proportions decreased to below 2.6% in A7.

Site 2 exhibited a similar trend to Site 1, with a high proportion of errors due to incorrect entries in A3 and A4 for the Screen instrument. In the Inclusion/Exclusion criteria instrument, the proportion of errors due to both incorrect and missing entries was near zero across all amendments, unlike at Site 1. The Precool instrument showed a range of 1-4% for incorrect entries and a higher proportion of errors due to missing entries at 2.5%-4.7%. Site 3 had near zero proportions for both incorrect and missing entries for almost all instruments and throughout the amendments, with some exceptions. The Inclusion/Exclusion instrument in A3 and A4 exhibited error proportions due to missing entries of 3% and 3.4%, while 4.2% of errors due to incorrect entries were noted in the Screen instrument in A5. The Precool instrument displayed a slightly higher error proportion for both categories, with proportion ranging between 0% and 3.4%, but the error proportions were near zero in A7.

Site 4 had the lowest error proportion due to incorrect entries in the A3 Screen instrument (0%) and the highest in A4(4.8%), while this proportion had dropped to 0.93% in A7. However, the A3 Screen instrument had the highest proportion of errors due to missing entries, while the proportion was lowest in A7. The Inclusion/Exclusion instrument had error proportions in both categories that remained constant between 0% and 4.1%. Still, the A3 instrument had a notable 14.3% of errors due to missing entries. A4 CRFs were not used by this site. In contrast, the Precool instrument reported a higher incidence of errors, particularly due to missing entries, with proportions ranging from 1 to 10.2%. While the error proportions for both missing and incorrect entries decreased in the Screen and Inclusion/Exclusion instruments, the proportion of errors increased in the transition to A7, reaching 2.11% and 0%, respectively.

Site 5 exhibited a notable error proportion due to incorrect capture in the Screen instrument during A5 reaching 11.1%, compared to 0% in the previous amendments for the same instrument. By A7, this rate had dropped to 5.3%. The A5 instrument also saw the

highest proportion of missing entries compared to other sites. However, the trend showed a lower proportion in A7. The Inclusion/Exclusion instrument had a near zero proportion of errors in all amendments, except for 3.2% of missing entries, with proportions remaining below 1.6% for other amendments. The Precool instrument had a wide range of incorrect errors, ranging from 0.3% in A4 to 14.7% in A3 and 2.4% in A5, with 4.3% in A4.

Site 6 experienced an incorrect error proportion above 4% in the Screen instrument for every amendment, except for A4 (0%), with no missing entries. The A4 Inclusion/Exclusion instrument saw no errors in missing or incorrect categories, but A5 had proportions of 2.4% and 1.6%, respectively. A7 CRFs were not used by this site.

The error proportions are zero across each Screen instrument across amendments in Site 7, except for an incorrect proportion of 1.7% in A4 and 4.4% in A7. The Inclusion/Exclusion criteria instruments showed a higher error proportion, with the highest in A4. The Precool instrument displayed a higher error proportion in A4 (7.3%; missing entries) and A5 (4.1%; incorrect entries).

4.3.3 Error values and Proportions among investigators

Table 11 shows the error proportions for each investigator across all the sites. Investigators 1A and 2A were excluded from the analysis because 1A managed only a single record (counted as a training record) at site 1, and 2A dealt with no records at site 2. Among the investigators who handled the A3 CRFs, investigator 5A had the highest proportion of incorrect entries and investigator 4A made the most frequent missing errors. In A4, investigator 2C had the highest error proportion due to incorrect entries and investigator 5B had the highest proportion due to missing entries. In terms of missing data in A5, 6B had a maximum proportion of 6.25% of incorrect data, while investigator 4B had a missing proportion of 7.78%. In the transition to A7, most error rates remained below 4%, however, investigator 7C achieved a proportion of 4.15% while investigator 1C had a missing proportion of 9.66%.

Site 1 investigators had incorrect proportions ranging between 0.61-3.74% and a higher proportion of missing entries (0.54-9.66%). Investigator 1D had the highest proportion of correct entries with 79.39% in A7, while the other investigators got proportions below 69%.

In comparison, investigators 2B and 2C had a proportion of correct entries below 61% for A3 and A4 respectively.

Table 11: An overview of the values and proportions of error codes among investigators

Sites	Investigator	Amendment	'0'	'1'	'555'	'777'	'888'	'999'	Total variables
Site 1	1B	A3	15(2.59%)	336(57.93%)	21(3.62%)	204(35.17%)	3(0.52%)	1(0.17%)	580
	1C	A3	12(1.66%)	405(55.86%)	35(4.83%)	254(35.03%)	2(0.28%)	17(2.35%)	725
		A4	38(3.74%)	681(67.09%)	50(4.93%)	198(19.51%)	2(0.20%)	46(4.53%)	1015
		A5	4(0.92%)	300(68.97%)	14(3.22%)	68(15.63%)	2(0.46%)	47(10.80%)	435
		A7	10(3.45%)	188(64.83%)	28(9.66%)	37(12.76%)	4(1.38%)	23(7.93%)	290
	1D	A7	8(0.61%)	1036(79.39%)	7(0.54%)	159(12.18%)	11(0.84%)	84(6.44%)	1305
Site 2	2B	A3	9(1.55%)	339(58.45%)	33(5.69%)	195(33.62%)	4(0.69%)	0	580
	2C	A3	59(2.71%)	1315(60.46%)	81(5.69%)	510(23.45%)	18(0.83%)	192(8.83%)	2175
		A4	55(3.79%)	877(60.48%)	54(3.72%)	257(17.72%)	16(1.10%)	191(13.17%)	1450
		A5	14(2.41%)	409(70.52%)	24(4.14%)	104(17.93%)	0	29(5%)	580
		A7	1(0.69%)	110(75.86%)	8(5.52%)	20(13.79%)	0	6(4.14%)	145
	2D	A5	1(0.34%)	217(74.83%)	8(2.76%)	41(14.14%)	2(0.69%)	21(7.24%)	290
		A7	9(1.03%)	657(75.52%)	15(1.72%)	123(14.14%)	1(0.11%)	65(7.47%)	870
Site 3	3A	A3	11(0.95%)	747(64.62%)	16(1.38%)	363(31.4%)	19(1.64%)	0	1156
		A4	30(2.96%)	686(67.58%)	30(2.96%)	185(18.23%)	36(3.55%)	48(4.73%)	1015
		A5	12(1.38%)	575(66.09%)	17(1.95%)	131(15.06%)	11(1.26%)	124(14.25%)	870
		A7	0	441(76.03%)	2(0.34%)	85(14.66%)	26(4.48%)	26(4.48%)	580
Site 4	4A	A3	30(3.45%)	423(48.62%)	76(8.74%)	305(35.06%)	34(3.91%)	2(0.23%)	870
	4B	A5	15(1.48%)	585(57.64%)	79(7.78%)	149(14.68%)	44(4.33%)	143(14.09%)	1015
		A7	19(1.72%)	598(57.07%)	107(7.33%)	127(17.24%)	155(12.33%)	153 (4.31%)	1160
	4C	A7	3(2.07%)	75(51.72%)	3(2.07%)	14(9.66%)	21(14.48%)	29(20%)	145
Site 5	5A	A3	33(11.38%)	162(55.86%)	6(2.07%)	89(30.69%)	0	0	290
		A4	1(0.7%)	125(88.03%)	3(2.11%)	13(9.15%)	0	0	145
	5B	A4	0	113(79.58%)	13(9.15%)	19(11.27%)	0	0	145
	5C	A4	0	97(66.7%)	1(0.69%)	25(17.24%)	8(5.52%)	14(9.66%)	145
		A5	25(5.55%)	318(73.44%)	10(2.31%)	59(13.39%)	6(1.39%)	17(3.93%)	435
		A7	37(2.88%)	905(69.75%)	31(2.26%)	140(10.57%)	35(2.72%)	156(11.82%)	1305
Site 6	6A	A3	3(2.07%)	72(49.66%)	12(8.28%)	46(31.72%)	12(8.28%)	0	145
	6B	A3	4(2.75%)	88(60.69%)	4(2.76%)	47(32.41%)	2(1.38%)	0	145
		A4	7(2.42%)	222(76.82%)	10(3.46%)	40(13.84%)	10(3.46%)	0	289
		A5	18(6.25%)	214(74.31%)	6(2.08%)	42(14.58%)	8(2.78%)	0	288
	6C	A5	12(2.07%)	372(64.14%)	8(1.38%)	67(11.55%)	79(13.62%)	42(7.24%)	580
Site 7	7A	A3	10(3.52%)	208(73.24%)	5(1.76%)	60(21.13%)	1(0.35%)	0	284
	7B	A3	18(3.19%)	410(72.57%)	10(1.77%)	125(22.12%)	2(0.35%)	0	565
		A4	14(2%)	539(77%)	42(6%)	78(11.14%)	27(3.86%)	0	700
		A5	12(2.16%)	453(81.47%)	12(2.16%)	55(9.89%)	4(0.72%)	20(3.6%)	556
		A7	21(3.62%)	439(75.69%)	12(2.07%)	85(14.66%)	4(0.69%)	19(3.28%)	580
	7C	A4	4(2.88%)	109(78.42%)	5(3.6%)	15(10.79%)	6(4.32%)	0	139
		A5	15(3.44%)	332(76.15%)	11(2.52%)	58(13.3%)	2(0.46%)	18(4.13%)	436
		A7	18(4.15%)	339(78.29%)	10(2.31%)	53(12.24%)	4(0.92%)	9(2.08%)	433
	7D	A3	0	133(95.69%)	0	6(4.32%)	0	0	139

Some investigators shared data capture responsibilities for specific participants seen in Table 12. In the case of participant 05-N-010 from Site 5, both investigators 5A and 5C were involved, and there were no recorded incorrect entries in either instrument. However, investigator 5C had an error rate of 11.61% for missing entries, while investigator 5A had no errors in this regard. Investigator 7B captured Screen data without any incorrect or missing

entries, whereas investigator 7C logged 0.89% of incorrect data and 2.68% of missing entries. For participants 07-N-019, 07-N-022 and 07-N-112, investigator 7C handled the Screen data during A5, and there were no data capture errors or missing entries. In A7, however, investigator 7C had a 3.03% incorrect entry rate in the Screen instrument. Investigator 7B exhibited the highest error rate in A5, with 2.7% incorrect entries and 4.46% missing entries in the Screen instrument. The highest missing entry rate for investigator 7B was 2.68% in A5 and 1.79% in A7.

Table 12: Overview of investigators according to instruments per participants from sites 5 and 7

Site	Patient ID	Amendment	Investigator	Instrument	'0' (%)	'1' (%)	'555' (%)	'777' (%)	'888' (%)	'999' (%)	Total variables
Site 5	05-N-010	A4	5A	Screen	0	25(75.76)	0	8(24.24)	0	0	33
			5C	Precool	0	89(79.46)	13(11.61)	10(8.93)	0	0	112
Site 7	07-N-013	A4	7B	Screen	0	24(72.73)	0	9(27.27)	0	0	33
			7C	Precool	1(0.89)	97(86.61)	3(2.68)	10(8.93)	1(0.89)	0	112
	07-N-019	A5	7C	Screen	0	24(72.73)	0	9(27.27)	0	0	33
			7B	Precool	3(2.68)	96(85.71)	3(2.68)	9(27.27)	1(0.89)	0	112
	07-N-022	A5	7C	Screen	0	24(72.73)	0	9(27.27)	0	0	33
			7B	Precool	3(2.70)	98(88.29)	1(0.9)	9(8.11)	0	0	112
	07-N-112	A7	7C	Screen	1(3.03)	22(66.67)	0	10(30.3)	0	0	33
			7B	Precool	5(4.46)	94(83.93)	2(1.79)	10(8.93)	0	1(0.89)	112

4.3.4 Error proportions: Amendments on a site level

Table 13 displays an overview of the proportion of error codes taken from the error reporting spreadsheets discussed in Figure 20. The sum variables of 6124 are observed under A3; the greatest proportion of codes were attributed by '1' indicating that a range of 48-76% of A3 variables were correctly captured to REDCap. The proportion of incorrect captures (code '0') ranged between 0.99% and 11.38%, with Site 5 displaying a notable high proportion of errors in relation to the other sites. Site 4 had the highest proportion of missing entries ('555'; 8.74%).

Table 13: An overview of the values and proportions of error codes reported in this investigation per amendment and site.

A3							
	'0'	**'1'**	**'555'**	**'777'**	**'888'**	**'999'**	**Total variables**
Site 1	27 (2.1%)	741 (56.8%)	56 (4.3%)	458 (35.1%)	5 (0.4%)	18 (1.4%)	1305
Site 2	13 (1.0%)	777 (59.5%)	60 (4.6%)	448 (34.3%)	6 (0.5%)	1 (0.1%)	1305
Site 3	11 (0.9%)	747 (64.4%)	16 (1.4%)	363 (31.3%)	19 (1.6%)	0	1156
Site 4	30 (3.4%)	423 (48.6%)	76 (8.7%)	305 (35.1%)	34 (3.9%)	2 (0.2%)	870
Site 5	33 (11.4%)	162 (55.9%)	6 (2.0%)	89 (30.7%)	0	0	290
Site 6	7 (2.4%)	160 (55.2%)	16 (5.5%)	93 (32.1%)	14 (4.8%)	0	290
Site 7	28 (2.8%)	751 (74.0%)	15 (1.5%)	191 (18.8%)	3 (0.3%)	0	988
						Sum variables	**6204**
A4							
Site 1	38 (3.7%)	681 (67.1%)	50 (4.9%)	198 (19.5%)	2 (0.2%)	46 (4.5%)	1015
Site 2	55 (3.8%)	877 (60.5%)	54 (3.7%)	257 (17.7%)	16 (1.1%)	191 (13.2%)	1450
Site 3	30 (3.0%)	686 (67.6%)	30 (3.0%)	185 (18.2%)	36 (3.5%)	48 (4.7%)	1015
Site 4	-	-	-	--	-	-	-
Site 5	1 (0.2%)	335 (77%)	17 (3.9%)	54 (12.4%)	8 (1.8%)	14 (3.2%)	429
Site 6	7 (2.4%)	222 (76.6%)	10 (3.4%)	40 (13.8%)	10(3.4%)	0	289
Site 7	14 (1.9%)	540 (74.5%)	42 (5.8%)	78 (10.8%)	27 (3.7%)	0	701
						Sum variables	**4899**
A5							
Site 1	4 (0.9%)	300 (69%)	14 (3.2%)	68 (15.6%)	2 (0.5%)	47 (10.8%)	435
Site 2	15 (1.7%)	626 (72%)	32 (3.7%)	145 (16.7%)	2 (0.2%)	50 (5.7%)	870
Site 3	12 (1.4%)	575 (66.1%)	17 (2.0%)	131 (15.1%)	11 (1.3%)	124 (14.3%)	870
Site 4	15 (1.5%)	585 (57.6%)	79 (7.8%)	149 (14.7%)	44 (4.3%)	143 (14.1%)	1015
Site 5	24 (5.5%)	318 (73.1%)	10 (2.3%)	58 (13.39%)	6 (1.4%)	17 (3.9%)	433
Site 6	30 (3.4%)	586 (67.4%)	14 (1.6%)	109 (12.5%)	87 (10.0%)	42 (4.8%)	868
Site 7	29 (3.3%)	661 (76.0%)	39 (4.5%)	81 (9.3%)	4 (0.5%)	20 (2.3%)	834
						Sum variables	**5325**
A7							
Site 1	18 (1.1%)	1224 (76.7%)	35 (2.2%)	196 (12.3%)	15 (0.9%)	107 (6.7%)	1595
Site 2	10 (1.0%)	767 (75.6%)	23 (2.3%)	143 (14.1%)	1 (0.1%)	71 (7.0%)	1015
Site 3	0	441 (76%)	2 (0.3%)	85 (14.7%)	26 (4.5%)	26 (4.5 %)	580
Site 4	22 (1.7%)	673 (51.6%)	110 (8.4%)	141 (10.8%)	176 (13.5%)	182 (13.9%)	1304
Site 5	37 (2.8%)	897 (68.7%)	29 (2.2%)	136 (10.4%)	35 (2.7%)	152 (11.6%)	1286
Site 6	-	-	-	-	-	-	-
Site 7	39 (3.8%)	778 (76.7%)	22 (2.2%)	138 (13.6%)	8 (0.8%)	28 (2.8%)	1013
						Sum variables	**5780**

There were fewer sum variables captured in A4 relative to A3; however, Site 5 had the lowest proportion of incorrect entries using A4 CRFs. Collectively, the ranges for both incorrect and missing entries increased from A3 to A4 across all sites. Site 4 did not utilise A4 CRFs.

A5 had the greatest sum variables of which a range of 0.91-5.54% and 1.61-7.78% of incorrect and missing entries were observed across all sites, respectively. Under A7, Site 3 had no incorrect entries, and the proportion of errors at the other sites was observed below 3.85%. However, missing entries were observed at a minimum of 0.34% at Site 3, and a maximum of

8.44% at Site 4. Site 6 did not use A7 CRFs. Apart from Sites 4 and 5, the remaining sites recorded proportions of correct entries above 75%.

Across all amendments, most sites observed at least 50% of correct entries except for Site 4 under A3. Site 7 had the highest proportion of correct entries across all sites throughout every amendment, while the highest proportion of incorrect entries varied per site and amendment. Site 5 had the highest proportion of incorrect entries under A3 and A5, while Site 4 had the highest proportion of missing entries under A3, A5 and A7.

4.3.5 Visual representations

4.3.5.1 Instruments

This section provides a visual overview of the proportion of errors observed in the Screening, Inclusion/ Exclusion and Precool instruments. The data in Figure 21 indicates a higher proportion of errors in the Precool instruments compared to the other two instruments for each amendment. Additionally, within each instrument, A4 exhibited the highest number of errors compared to any other amendment. The Precool instrument exhibited a consistently high error rate across all amendments persisting above 35%. The highest proportion of errors was noted in A5, although there was a noticeable decrease in A7.

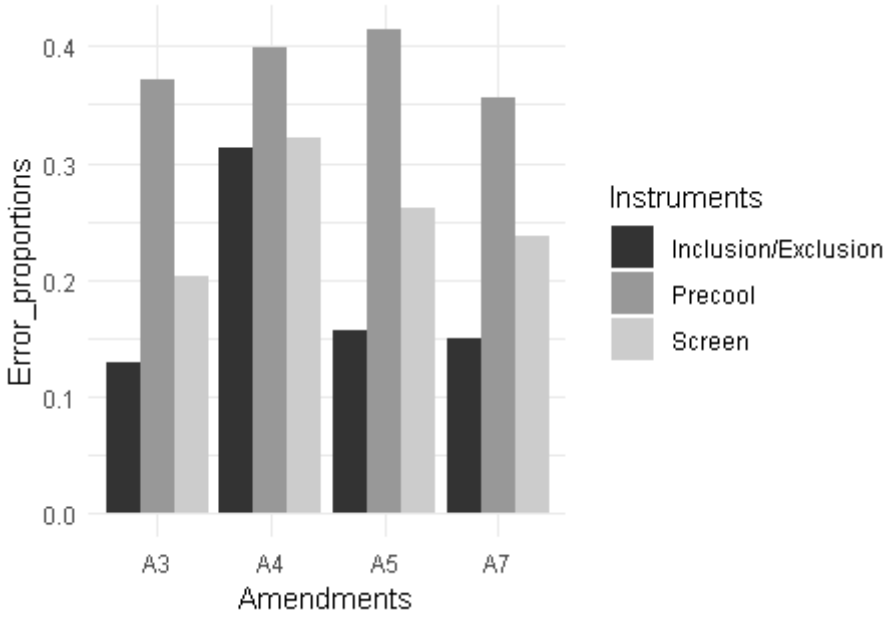

Figure 21 : The error proportions assessed across the Screening sheet, Inclusion/Exclusion criteria and Precool instruments for all sites

4.3.5.2 Amendments

In this section, the proportion of errors observed in amendments 3, 4, 5 and 7 across all sites are visualised. In Figure 22, it is seen that most sites recorded error rates exceeding 5%, with the exceptions of Sites 3 and 7, which maintained error rates below 2.5% and 5% respectively. Sites 4 and 5 reported particularly high error rates, exceeding 10%.

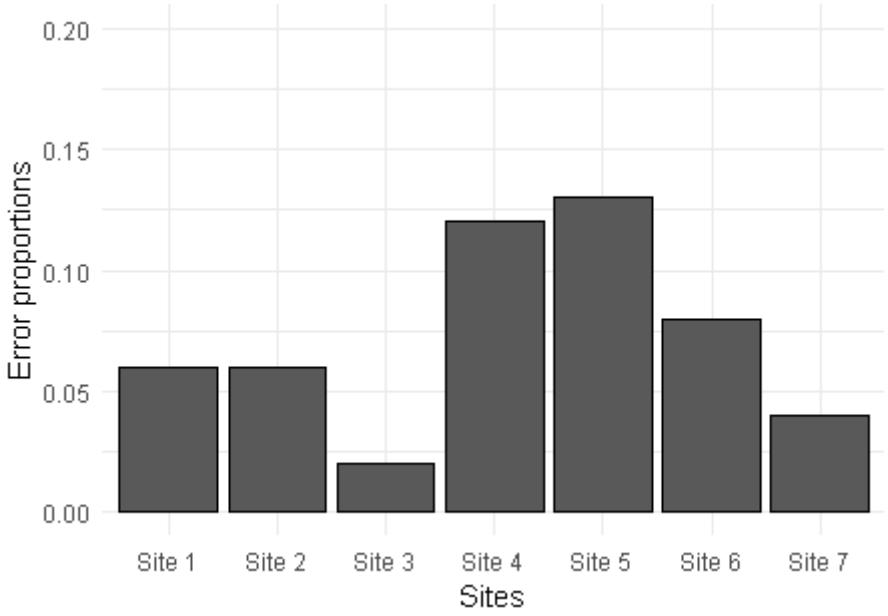

Figure 22: Error proportions in A3 across all sites

With the transition to A4, the error rates at all sites decreased, falling below 10% as seen in Figure 23. However, the majority of sites still maintained error proportions above 5%, with only Site 5 being the exception, recording an error proportion below 5%. Notably, Site 4 did not contribute data for A4 since this site did not use the A4 CRFs.

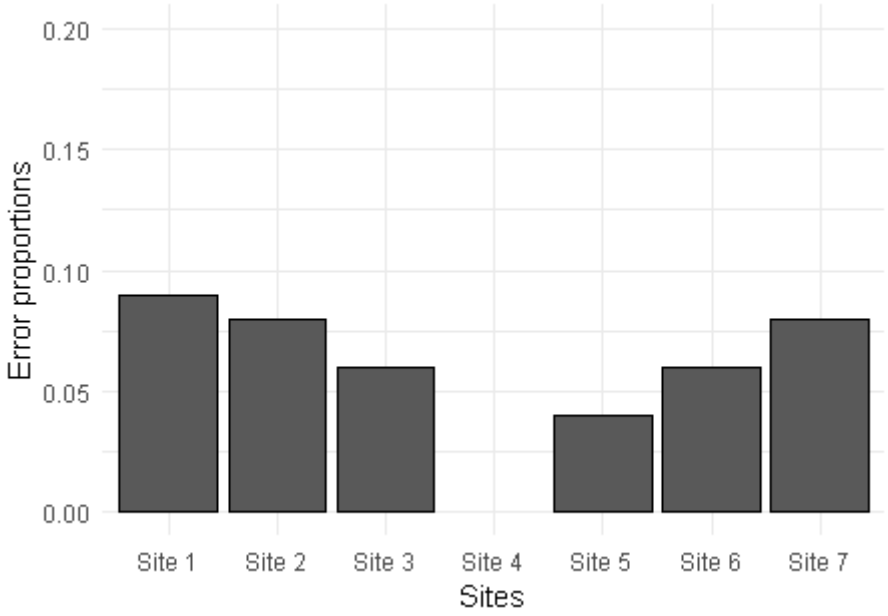

Figure 23: Error proportions seen in A4 across all sites

The data collected from A5 is inconsistent among all the sites in Figure 24, with the error rates consistently below 10%. Sites 1, 2, 3, and 6 had reduced error proportions at or below 5%. In contrast, Sites 4, 5, and 7 exhibited higher error rates above 7.5%.

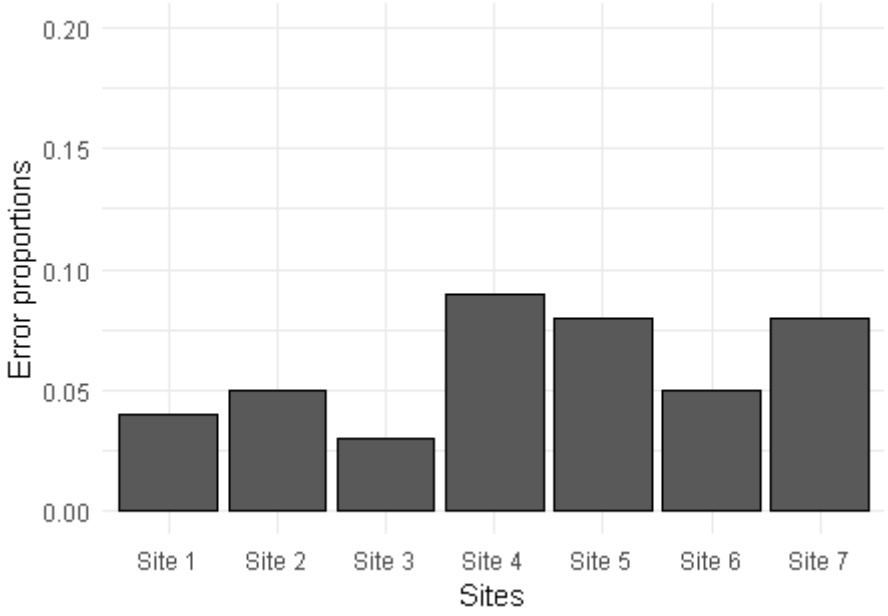

Figure 24: Error proportions in A5 across all sites

The data generated in the A7 group presents low error rates in Figure 25, with Sites 1, 2 and 3 all recording error proportions below 5%, and Site 3 approaching an error proportion of nearly 0%. Site 5's error proportion is slightly above 5%, while Site 4 reported the highest

error proportion, exceeding 10%. Site 7 also recorded an error proportion above 5%. Notably, Site 6 did not provide data for A7 CRFs, as these forms were not utilised.

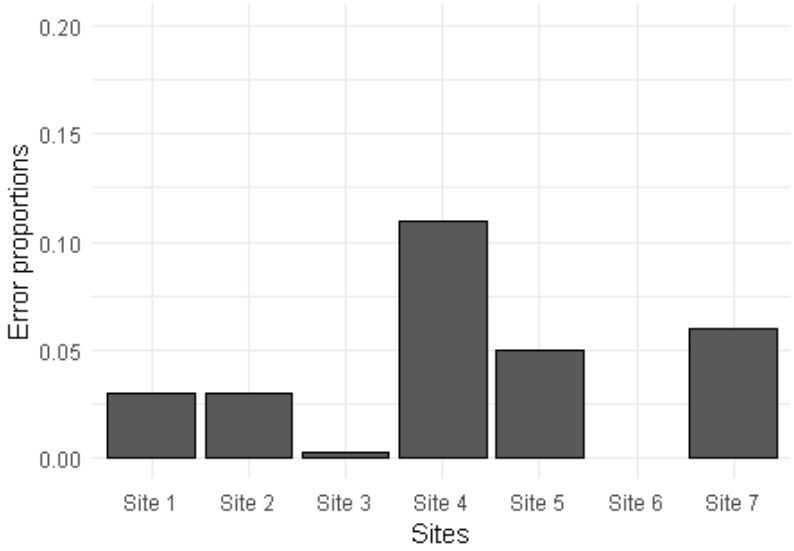

Figure 25: Error proportions in A7 across all sites

In Figure 26, the error proportions exceeded 5% across all amendments, ranging from around 7-8% for amendment groups A3 through A5, with A7 being the exception as it dipped below 5%. Notably, the wide error bars in A3 and A7 indicated the greatest variation in these groups. In contrast, the A4 and A5 groups exhibited narrower error bars, indicating a lesser degree of variation when compared to the other groups.

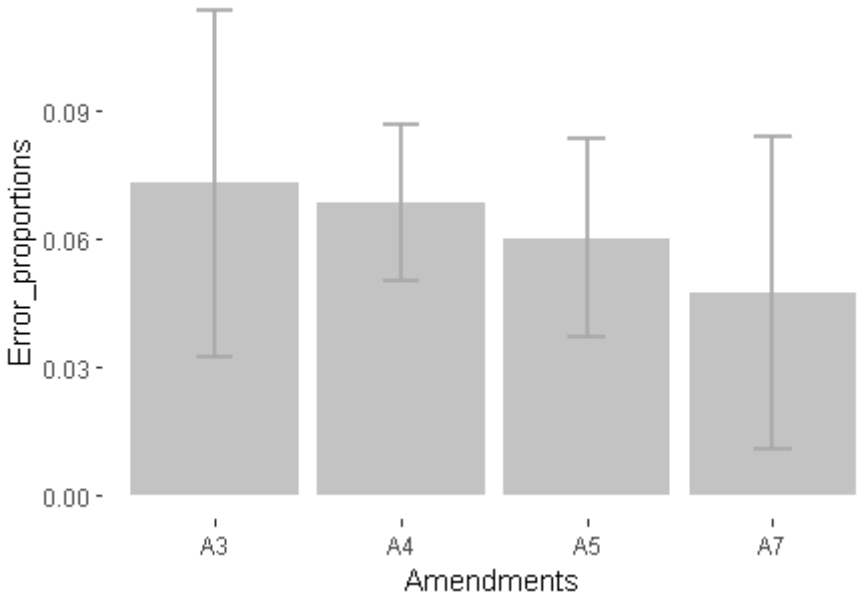

Figure 26: Error proportions across all amendments and all sites with error bars

4.3.5.3 Investigators

This section will present the proportion of errors observed in the investigators across all sites. In Figure 27, most investigators exhibited error proportions (missing and incorrect) exceeding 5%, except for investigator 3A. Investigators 2C, 4A, 5A and 6A had the highest error rates, consistently persisting above 10%, with investigator 2C's error proportion reaching nearly 20%. Investigators 1B and 1C, as well as investigators 7A and 7B, exhibit similar error proportions at their respective sites. Additionally, investigator 2C made more than twice the errors of investigator 2B, while investigator 6B recorded fewer errors than 6A. Investigators 1A and 2A are not included in the Figure 27 below because 1A dealt with a training document, and 2A did not oversee any patient records. Investigators 2D, 5B, 5C and 7D were not part of the study during A3, which is why there are no recorded error proportions for them.

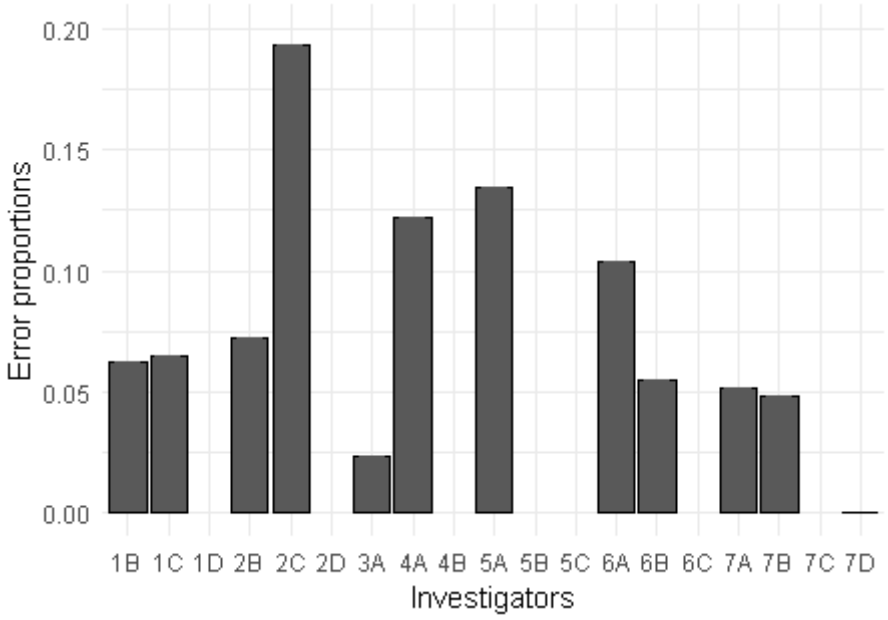

Figure 27: Error Proportions for Investigators in A3

The investigators reduced the proportion of errors below 10% at each site in Figure 28. However, despite being under the 10% mark, investigators 7B, 5B and 1C displayed the highest proportion of errors. Investigators 5A and 5C were exceptions with error proportions below 5%. At site 5, investigators 5B made twice as many errors as 5A, while 5C recorded an error proportion close to 0%. Conversely, at site 7, 7B made more errors than 7C on A4 documents.

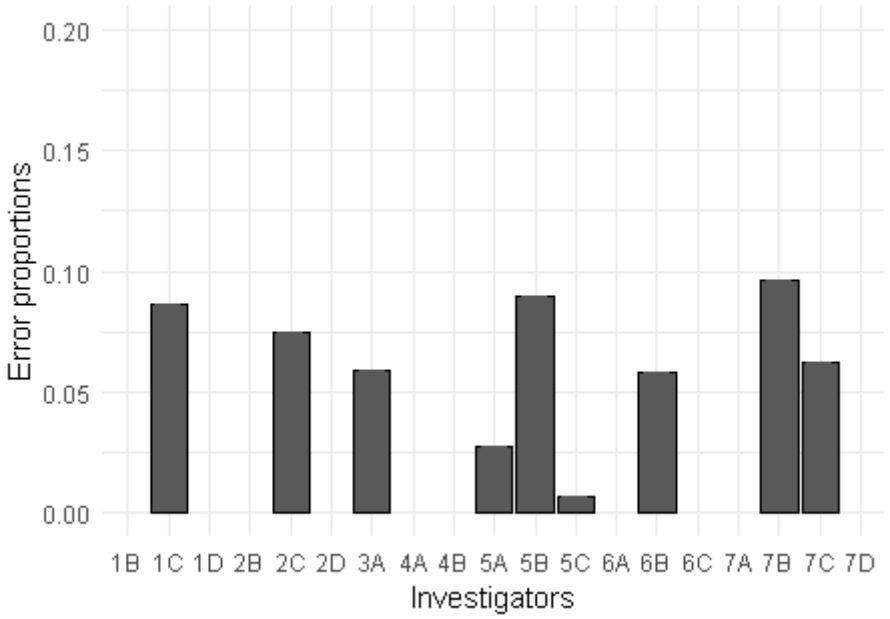

Figure 28: Error proportions for investigators in A4

In Figure 29, akin to A4, the error proportions for A5 remained below 10%. Several investigators were able to reduce their error proportions below 5%, including investigators 1C, 2D, 3A, 6C and 7C. The remaining investigators sustained error proportions within the 5-10% range. Notably, at site 2, investigator 2D made fewer errors than 2C, and at site 6, investigator 6C made fewer errors than 6B, while, at site 7, investigator 7C had fewer errors compared to 7B.

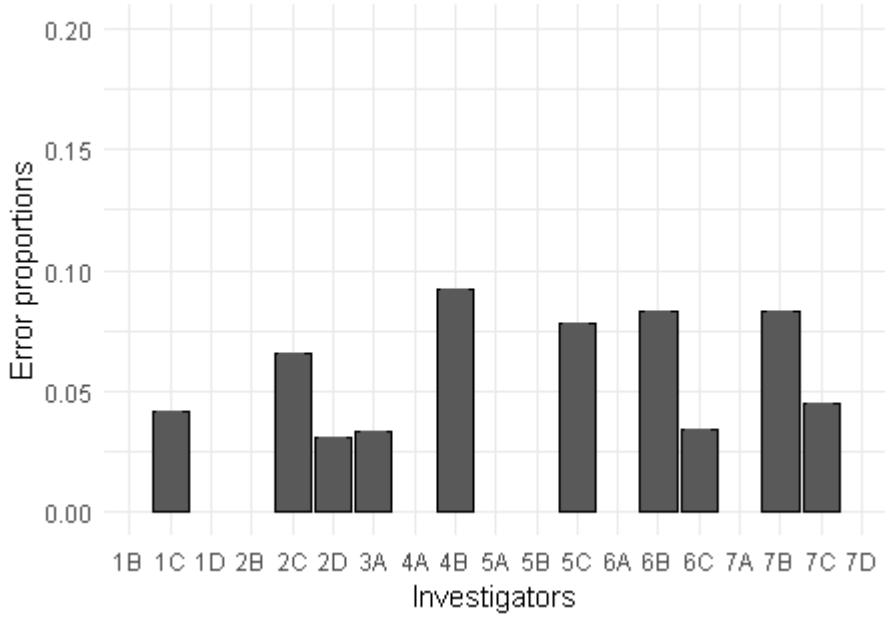

Figure 29: Error proportions for investigators in A5

In Figure 30, investigators 1D, 2D and 3A demonstrated error proportions below 5%, whereas investigators 1C and 4B recorded the highest errors, surpassing 10%. The remaining investigators fell within the error proportion range of 5-10%. At site 1, there is a notable contrast in the errors between investigators 1C and 1D, and at site 2, investigators 2D made fewer errors than 2C. In contrast to previous amendments, investigator 7B made fewer mistakes than 7C on A7.

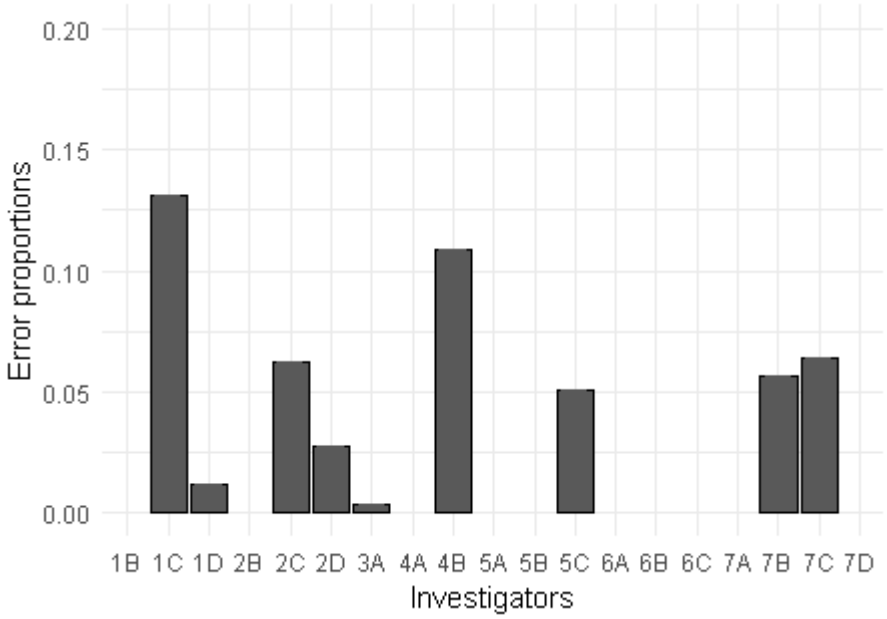

Figure 30: Error proportions for investigators in A7

The architects of the A7 NESHIE REDCap database noted a decrease in observed errors on the database; therefore, this led to the supposition that there would be a lower error rate in A7 than in A3.

4.4 Comparative analyses

Hypothesis: there is a statistically significant difference in the incorrect counts between the A3 and A7 amendment groups.

Table 14 displays the counts of both the correct and incorrect categories across all the sites in the study, as well as with every version of the amendment documents.

Table 14: The incorrect and correct counts across the different sites and different amendments

Site	Amendment	Incorrect	Correct
S1	**A3**	83	741
	A4	88	681
	A5	18	300
	A7	53	1224
	Total	**242**	**2946**
S2	**A3**	73	777
	A4	109	877
	A5	47	626
	A7	33	767
	Total	**262**	**3047**
S3	**A3**	27	747
	A4	60	686
	A5	29	575
	A7	2	441
	Total	**118**	**2449**
S4	**A3**	106	423
	A4	NA	NA
	A5	94	585
	A7	126	598
	Total	**326**	**1606**
S5	**A3**	39	162
	A4	18	335
	A5	34	318
	A7	66	897
	Total	**157**	**1712**
S6	**A3**	23	160
	A4	17	222
	A5	44	586
	A7	NA	NA
	Total	**84**	**968**
S7	**A3**	43	751
	A4	56	540
	A5	68	661
	A7	61	778
	Total	**228**	**2730**

The chi-squared test was performed to evaluate the association between two categorical variables, specifically the amendment groups. The chi-squared test yielded a test statistic (X-squared) of approximately 29.145, indicating the extent of deviation between the observed and expected counts within the contingency table. There were 3 degrees of freedom (df = 3). The p-value was calculated to be approximately 0.000002088 indicating strong evidence against the null hypothesis. Since the p-value was small and well below the significance level of 0.05, we reject the null hypothesis.

Table 15 displays the results of pairwise comparisons conducted to evaluate the presence of statistically significant distinctions among the different levels of amendments.

Table 15: The results of Pairwise comparisons between the amendment groups

Comparison	**Raw p-value**	**Adjusted p-value**
A3 vs. A4	0.969	0.969
A3 vs. A5	0.087	0.130
A4 vs. A5	0.108	0.130
A3 vs. A7	0**	0**
A4 vs. A7	0**	0**
A5 vs. A7	0.004**	0.007

Significance codes: 0.05 – ** 0.1 – *

The comparison "A3 vs. A7" yielded both the raw and adjusted p-values of 0, indicating a highly statistically significant difference between these two groups. Similarly, the "A4 vs. A7" comparison revealed a highly statistically significant difference with both raw and adjusted p-values of 0. "A5 vs. A7" exhibited a statistically significant difference with a raw p-value of 0.004. This significance is maintained even after adjustment (adjusted p-value = 0.007). Conversely, no significant differences were detected in the remaining comparisons.

Pairwise comparisons with Bonferroni correction

Table 16: Results of pairwise comparisons after applying Bonferroni correction

Dimension	Residuals	P-value
A3 vs A4	2.906	0.029**
A4 vs A5	2.568	0.082*
A5 vs A7	-0.041	1
A3 vs A7	-5.015	0**

Significance codes: 0.001 – *** 0.05 – ** 0.1 – *

The “A3 vs A4” comparison reveals a residual of 2.906 and the Bonferroni-adjusted p-value of 0.029which suggests that there is a statistically significant difference between "A3" and "A4" after accounting for multiple comparisons. The comparison "A3 vs. A7" yields the residual value of approximately -5.015, and the Bonferroni-adjusted p-value of 0. This adjusted p-value of 0 indicates a highly significant difference between "A3" and "A7" after Bonferroni correction. The “A4 vs A5” comparison is less than 0.10. Similarly, the “A5 vs A7” comparison suggests that there is no statistically significant difference detected. However, no statistically significant differences were detected for the remaining comparisons.

4.5 Discussion

Data from 160 participants were included in the study encompassing a wide range of data points. Each participant contributed a dataset of 145 variables for this investigation. The collected data included amendment groups (A3, A4, A5 and A7) across all seven sites in the study as well as data from twenty-one investigators and data from three instruments for each amendment group. Furthermore, the total number of correct-incorrect paired data points assessed per site varied, ranging from 1052 records at site 6 to 3309 records at site 3.

It was challenging to use the same number of patients across all sites in the study (see Table 10) since enrolment intake differed at each site. In addition, each site faced its own site-specific challenges that inadvertently affected the data. For example, site 5 only used A3 documents on two of its recruited patients while the data from two patients was selected for use from site 6 even though there were other patients recruited on A3. The recruitment rates were generally low at sites 5 and 6 for A3-A5. At site 3, only data for four patients were recorded under A7, as the investigator was on maternity leave, leading to a halt in recruitment at the site for most of 2021. Although site 7 had the highest recruitment rate in the study,

there was a significant delay between when a patient was recruited and the moment the data became accessible. Therefore, for the purpose of this investigation, the patient data were collected based on what was available at each site.

4.5.1 Analysis of error reports

An error report was constructed for each site to document the characteristics of the recorded entries categorising them as “correct”, “incorrect”, “missing”, “unknown” and “not applicable”. The primary objective of assessing data quality was to ensure the accuracy and completeness of the data. Consequently, particular emphasis was placed on the “incorrect” and “missing” data entries to potentially uncover insights into issues within the data capture process. Subsequently, error proportion rates were calculated for each site within each amendment group, and these error proportions were graphically represented using bar plots.

4.5.1.1 Data Analysis by instruments

The Screening and Inclusion/Exclusion criteria instruments consistently presented lower error proportions compared to the Precool instrument across each amendment. For both the Screening and Inclusion/Exclusion criteria instruments, errors increased from A3 to their peak in A4, declined in A5, and remained stable in A7. Fewer errors were observed in the Screening and Inclusion/Exclusion criteria instruments in A3. This implies the persistence of specific errors within these instruments. In contrast, errors in the Precool instrument peaked above 40% in A5 but decreased to approximately 35% in A7, representing the lowest rate across all amendments. This trend suggests a need for further investigation to uncover the underlying reasons for such patterns. However, given that the Precool instrument contains more text fields than the other instruments, it is plausible that a significant portion of errors may stem from data capturing to REDCap.

4.5.1.2 Data Analysis by Amendments

In the overall bar plot (Figure 21), there is a clear trend showing a steady decrease in mean errors with each transition to a newer amendment. This observable pattern signifies an ongoing, positive continuum of improvement in study processes and the implementation of quality control measures. As the researchers become more familiar with the CRFs and increase their proficiency with the REDCap database, there is a tangible expectation of improved adherence to data quality standards.

Despite the notable reduction in error rates across the amendments, the mean error rate for A7 lingers persistently above 5%, which was initially established as the predefined acceptable error rate for the study. This sustained elevation in error rates within the A7 database prompts a closer look at potential contributing factors. While the overall trend indicates continuous improvement, the specific challenges observed in A7 necessitate focussed attention. There remains the potential for this rate to fall within the acceptable range in subsequent database iterations.

To address this, database architects should fine-tune existing design elements, identify further areas for improvement in both the CRFs and database, and possibly explore other methods for refining the database design to maintain a low error rate.

Although four of the seven sites (sites 1, 2, 3 and 7) exhibited higher mean error rates in the A4 group compared to A3, the overall mean across all sites remained lower than that of A3 (see Figures 22 and 23). Additionally, although the A5 mean was lower than A7 in site 4, a comprehensive view reveals that the A7 mean represented the lowest mean error rate across all sites (see Figures 24 and 25). This observation could certainly be attributed to the implemented design changes in the A7 database. In Figure 26, the A3 and A7 groups exhibit broader error bars associated with their means than A4 and A5. These wider error bars serve as visual indicators of the greater variability that exists within the data, suggesting low precision and higher uncertainty associated with the reported means. This heightened variability suggests a higher level of uncertainty associated with the measured outcomes, emphasising the importance of careful interpretation when considering the central tendencies presented. The extended reach of the error bars also implies that certain individual data points within the dataset may deviate more substantially from the calculated mean proportions. This variability could stem from various sources including methodological challenges, or other unaccounted factors. This increased uncertainty can have implications for the generalisability and reliability of findings. Therefore, even though the A7 group displays the lowest mean error rate of all amendments, the true value could be greater or lower than what is observed.

4.5.1.3 Data Analysis by Investigators

In the examination of the investigator performance within the study, there were intriguing variations in the error rates across the different sites and investigators were observed. Site 4

stood out by recording the greatest number of errors, particularly during A3 and A7. This high error rate observed at Site 4 could be attributed to two noteworthy incidences. Firstly, investigator 4A, responsible for A3 data, exhibited a substantial number of general errors on A3 CRFs. This may be attributed to the novelty of the study, and potentially, a lack of familiarity with the study documents. Moreover, general lack of technological skills on the part of Investigator 4A contributed to a significant prevalence of both incorrect and missing data. Secondly, Investigator 4B, tasked with handling A5 and A7 CRFs, experienced a heightened error rate on A7, primarily due to personal challenges that began impacting the work. This resulted in an increased number of incomplete data points, which contributed to a high prevalence of missing data at the site. These insights highlight the need for support and training, particularly for investigators grappling with technological challenges, and emphasise the use of monitoring and addressing personal challenges that may impact data quality in research studies.

In contrast, the consistent occurrence of errors exceeding 5% at sites 1, 2, 6 and 7 across A3, A4, and A7. These sites maintained a similar number of errors throughout the study and suggests the need for targeted interventions at these sites. Interestingly, investigators introduced later in the study from sites 1, 2, 5 and 6 appeared to commit fewer errors compared to those present from the study's onset. This observation proposes an interesting insight into the learning curve: improved familiarity with study protocols and training methodologies over time may have contributed to enhanced performance among newer investigators.

Inconsistencies were observed in investigators 7B and 7C at Site 7 where their error patterns displayed variations exceeding approximately 5%. These investigators have been involved in the study since its inception and so they had worked on all amendment documents. Investigator 7C made fewer errors than 7B in A4 and A5, while 7B made fewer errors in A7. This suggests that errors did not decrease with the transition to A7 but instead persisted. This calls for a comprehensive exploration into potential site-specific challenges associated with A7 and, by addressing these issues, the overall data quality at this site can be improved.

In contrast, the investigator from Site 3, who has been with the NESHIE study since its inception, exhibited the lowest incidence of mean errors across all sites. This performance

invites a closer look into the practices and techniques employed by this investigator, potentially offering valuable insights that could inform best practices for others.

The data from investigators at sites 1 and 2 provided some interesting observations. Firstly, investigators 1A and 2A had to be excluded from the analysis as they did not work on complete patient data. Lastly, specific investigator behaviours were apparent. It was noted that with each transition to a new document, a new investigator was introduced, although there were occasional overlaps. Investigator 1B's differing error rates at Site 2 compared to Site 1 on the A3 documents, and Investigator 1C's exclusive use of A7 documents at Site 1 present intriguing patterns that merit further explanation into whether these variations were due to document-specific challenges or contextual factors that influenced investigator performance.

4.5.1.4 Results from the comparative analyses

The bar plots constructed for each of the seven sites involved in this investigation depicted a trend that revealed a general decrease in errors with the move from A3 to A4, A5 and finally A7. This lower incidence could be attributed to the improved design of both the CRFs and the database and established the hypothesis that the mean error rate would be lower for the A7 group than that of previous amendments. The chi-square test proved that the difference between the mean errors between the groups was significant with a 95% confidence interval ($p=0.000002083$; $p<0.05$). The chi-square test pairwise comparisons performed on the overall data imputed significance with confidence that supported the hypothesis specifically regarding "A3 vs. A7" as well as "A4 vs. A7) ($p = 0$; $p<0.05$). This shows that there was a highly significant difference between the early amendments and A7. Both the raw and adjusted p-values for the "A5 vs. A7" group were significant ($p= 0.004$ and 0.007 respectively; $p<0.05$).

4.5.1.5 Insights from the descriptive and comparative analyses

Several factors could have played a role in decreasing the incidence of errors described in the comparative analysis. Continuous use of the document-database pair could have furthered mastery and confidence in data entry among investigators. For instance, investigator 3 exhibited an error rate of approximately 2.5% in A3, 6% in A4 and close to 4% in A5 (see Figures 27, 28 and 29). However, with the introduction of A7 documents, the investigator's error rate dropped to nearly 0% (see Figure 30). This progression clearly demonstrates the impact of continued exposure and experience with both the documents and the database.

It is worth noting that not all investigators experienced a similar decrease in error rates. Investigator 1C and investigator 7B were the only other investigators in the study to use all amendments and presented distinct patterns in their error rates over time. Investigator 1C's error rates (at site 1) were approximately 7% and 9% with A3 and A4 respectively (see Figures 27 and 28). However, with the introduction of A5 the rate dropped to 4% only to sharply rise to 13% with A7. Investigator 7B exhibited error rates of approximately 5% and 9% with A3 and A4 respectively (see Figure 30). The error rates of A5 and A7, were lower than A4, at approximately 4% and 6.5%, respectively.

Nonetheless, it is the belief of the database architects that the improved features applied to the REDCap database were instrumental in decreasing the error incidence from the first iteration to the A7 REDCap database.

The overall error rate dropped from roughly 8% to 6% (Figure 26), marking a substantial reduction. Although the A7 error rate remained slightly above the 5% error rate set for the study, it represents a meaningful decrease. Upon closer inspection at the site level, only sites 1, 2 and 3 achieved an error rate of 5% or lower in the A7 group (Figure 30). However, it is important to note that while the error rate decreased in the A7 group for sites 6 and 7 compared to earlier amendments, it exhibited a sharp increase at site 4. It is important to note that the high error rate noted at this site were predominantly due to unavailable data (data not collected) as opposed to data capture errors.

4.5.3 Practical implications, advantages, limitations and future directions

The practical implications from this chapter highlight the critical role of data quality in clinical research. It further underscores the importance of ongoing data monitoring and validation to improve data accuracy, with the hopes of leading to greater decision-making in clinical trials. Variations in error rates across the different sites, investigators and instruments emerged and emphasised the need for tailored support and training to address site-specific challenges effectively. The higher error rates in the earlier amendment documents (A3 and A4) lined up with the need for an enhanced and refined database design to reduce errors in data capture. This insight confirms the conclusion from the previous chapter that data management systems should grow to align with evolving research needs.

This project was considerably limited by the availability of data from specific sites and investigators. For example, the high error rate at site 4 can be primarily ascribed to the contributions of two investigators. Consequently, the findings from this project vary considerably due to site-specific factors and may therefore, limit their generalisability to diverse clinical research settings.

Future research could focus on developing strategies to reduce errors, for instance, targeted training programs for investigators and continuous improvements to the document design. Additionally, developing standardised procedures for error documentation as well as reporting could provide a more systematic approach to error management. And of course, other research teams could investigate changes over extended time periods and implement interventions that could provide insights into the effectiveness of quality improvement efforts.

4.6 Conclusion

In the realm of clinical research, there is a great need for good and reliable data that is accurate and complete. This chapter provided a comprehensive analysis of data quality, shedding light on its critical role in ensuring the integrity and reliability of data to meet research objectives. For this reason, the error reporting was vital so as to identify compliance issues, invalid data, as well as deviations from study protocols. Through an extensive examination of error rates across various sites, amendments, investigators and instruments, valuable insights were gained into the nuances of data accuracy and the factors that influence it. One of the most important outcomes of this project was continuous data monitoring and validation. The error rates varied at the different sites demonstrating the dynamic nature of data quality to enhance data accuracy. Some solutions included tailored support and training initiatives directed at the specific identified challenges encountered at the different sites.

Furthermore, this analysis showed the importance of database design in error reduction. The higher error rates observed in the earlier amendments revealed the need for more refined database systems to enhance data capture precision. These findings underline the relevance of adapting data management systems to the ever-changing research demands.

To address future research endeavours, targeted training programs for investigators, continuous document design enhancements and the consideration of standardised procedures for error documentation and reporting could improve data quality management.

The next chapter will examine the dataset used to identify outliers. This was utilised to distinguish between errors due to data entry and genuine data anomalies.

4.7 References

1. Nagle T, Redman T, Sammon D. Assessing data quality: A managerial call to action. Business Horizons. 2020; 63:325-37.

2. Houston L, Martin A, Yu P, Probst Y. Time-consuming and expensive data quality monitoring procedures persist in clinical trials: A national survey. Contemporary Clinical Trials. 2021; 103:106290. doi:10.1016/j.cct.2021.106290

3. Liu Q, Feng G, Zhao X, Wang W. Minimizing the data quality problem of information systems: A process-based method. Decision Support Systems. 2020; 137:113381. doi:https://doi.org/10.1016/j.dss.2020.113381

4. EMA. Reflection paper on risk based quality management in clinical trials. [Internet]. 2013. [cited Access Year Access Date|]|.

5. Nahm ML, Pieper CF, Cunningham MM. Quantifying data quality for clinical trials using electronic data capture. PLoS ONE. 2008; 3(8):e3049. doi:10.1371/journal.pone.0003049

6. Houston L, Probst Y, Martin A. Assessing data quality and the variability of source data verification auditing methods in clinical research settings. Journal of Biomedical Informatics. 2018; 83:25-32. doi:https://doi.org/10.1016/j.jbi.2018.05.010

7. Hong MKH, Yao HHI, Pedersen JS, Peters JS, Costello AJ, Murphy DG, et al. Error rates in a clinical data repository: Lessons from the transition to electronic data transfer—a descriptive study. BMJ Open. 2013; 3(5):e002406. doi:10.1136/bmjopen-2012-002406

8. Database error rate. In: Kirch W, editor. Encyclopedia of public health. Dordrecht: Springer Netherlands; 2008. p. 196-.

9. Fischetti T Data analysis with r : A comprehensive guide to manipulating, analyzing, and visualizing data in r. [Internet] Birmingham, UK: Packt Publishing; 2018 [cited Access 2018 Access Date]. Available from: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1775099

10. Pearson RK. Exploratory data analysis using r. 2018; doi:10.1201/9781315382111

11. Maindonald J, Braun W. Data analysis and graphics using r: An example-based approach. Maindonald J, Braun J, editors. Cambridge University Press; 2010.

12. Morgenthaler S. Exploratory data analysis. WIREs Computational Statistics. 2009; 1(1):33-44. doi:https://doi.org/10.1002/wics.2

13. Kabacoff RI. R in action: Data analysis and graphics with r. 2nd ed. Shelter Island: Manning Publications; 2015.

14. Fischetti T. Data analysis with r. Birmingham, UNITED KINGDOM: Packt Publishing, Limited; 2015.

Chapter 5: Analysis II

5.1 Introduction

This chapter will address the analysis of a dataset to identify outliers and true errors as well as establish concordance to clarify the accuracy of the data.

Outliers are data points that appear inconsistent and separated from the central body of data[1]. These points can either be errors or genuine values that may confound any data models[1] and inflate error rates[2]. Seeming outliers depend on how the data is presented and can be advantageous in showing the presence of errors or anomalies in the data. Outliers also show trends and patterns specific to groups and subgroups as opposed to analysing points separately[2].

Outliers fall into distinct categories firstly, those that emerge from errors in the data, and secondly, those that emerge from variability of the data. Outliers from data errors are most often caused by human error for instance those typed erroneously in data entry, recording or data collection[2]. These types of errors can be adjusted by reviewing the original documentation and, if possible, recording the correct value [2]. Boxplots can be useful in displaying outliers in one dimension, scatterplots highlight them in two dimensions whilst other outliers will only appear in three or more dimensions[3]. The occurrence of outliers can indicate a divergence from model or data assumptions[3].

Outliers and extreme values are related but hold different concepts. Extreme values are the highest and lowest values in a dataset that are values furthest from the mean or median[2]. Extreme values can be indicative of outliers but may not actually be outliers, while outliers can be described as extreme values that are unexpected or unusual. As a value, an outlier has a low probability of emanating from the same statistical distribution as other observations while an extreme value is an observation that might have a low probability of incidence but cannot be shown to stem from a different distribution than the rest of the data [2].

Outliers can be hard to detect especially when two or more outliers that appear influential overshadow each other [3]. Influential observations or larger outliers will have a significant impact on the results[1]. An influential observation is an observation that has a strong impact on a data model's parameters and in this way, greatly influences these parameters[1].

Whether outliers are influential or not, they should not be ignored as both can affect the validity of results. Analysing the original data can be useful in determining if the outlier is a result of an error in data entry [3].

There are instances in which outliers need to be identified and deleted to improve a dataset's fit to the normality assumption[1]. In other cases, unusual observations may be the most interesting thing about the collected data[1].

The previous chapter dealt with errors due to data entry and therefore, this chapter will utilise these previously confirmed values and further analyse them for the identification of errors which, if present, demonstrate a difference between recorded or reported value and the actual source data. This difference will be referred to as a "true error".

In the context of this study, the reported values appear on REDCap while the source data refers to the hospital records available at the hospital sites, the latter of which was used by investigators on site to transcribe to the study specific CRFs. Distinguishing these true errors is crucial for an organisation to ensure the accuracy and dependability of data to avoid potential problems that could arise from utilising inaccurate data. In essence, outliers might represent genuine values, anomalies, or true errors.

5.2 Materials and Methods

The variables analysed in chapter four were the same variables to be examined in this chapter. While the previous chapter used a range of codes to determine concordance for the variables between REDCap and the CRFs, this chapter focuses on the evaluation of the actual data captured for particular variables.

A comprehensive report was constructed by filtering data from the Screening, Inclusion and Exclusion Criteria and Precooling instruments across all seven study sites. Subsequently, this data was downloaded from REDCap in the CSV format and imported into Microsoft Excel. In the spreadsheet, each variable occupied its own independent column, with study IDs organised in rows. To ensure consistency with the previous chapter, the data was filtered by data access groups (DAGs) to isolate the same study IDs by site. While certain variables, such as birthweight and gestational age, were downloaded as a single entity, others were divided into separate columns, necessitating their consolidation into a single variable.

5.2.1 Data cleaning in R studio

The data was analysed using R Statistical software (v4.1.3; R Core Team 2022). In R, Hmisc (v4.7-0; Harrell and Dupont 2018) was used to provide the five-number summary [1] while Inspectdf (v0.0.12; Rushworth and Wilkins 2019) was used to summarise missingness of data and visualise differences. Ggplot2 (v3.3.5; Wickham 2016) was used to construct graphs and, along with core functions offered in R, Tidyverse (v1.3.1; Wickham et al. 2019), Dplyr (v1.0.8; Wickham, Francois and Muller 2022) and Magrittr (v2.0.2; Bache et al. 2022) were used to sort and arrange the data.

The dataset contained numeric and character variables. The Z-score and Mahalanobis distance are commonly used for identifying outliers in datasets. Z-score describes the deviation of a data point from the mean in terms of standard deviation and is applicable to datasets following a normal distribution [4]. On the other hand, the Mahalanobis distance measures the distance between a data point and distribution, particularly in multivariate data[73].These methods offer quantitative assessments based on statistical measures. However, due to the simplicity and univariate nature of the data, the Mahalanobis distance was deemed unnecessary for outlier detection.

In this project, in addition to Z-scores, visual inspection and descriptive statistics were the favoured techniques related to their interpretation and computational simplicity. The objective was to assess both the skewness and spread of the data while simultaneously pinpointing data points that significantly deviate from the rest such as extreme values and outliers. Outliers were identified through various means, including the assessment of minimum and maximum values, boxplots, bar plots, Z-score plots, and percentiles. Statistical analyses were not performed on the data since the goal of this chapter is to demonstrate how the presence of outliers was indicative of “special cases” and therefore, a substantiation of underlying “true errors” at the source level.

5.2.2 Data analysis in R

5.2.2.1 Numeric variable analysis

The numeric variables included the variables neonate body length, birth weight and Ballard score to name a few. To obtain comprehensive summaries for these numeric variables, the *summary* function from Hmisc package was applied to compute key statistics such as

minimum and maximum values, quartiles, median and mean. Additionally, the *describe* was used to provide the count of missing data points and distinct values.

5.2.2.2 Character variable analysis

The character variables included Thompson scores, Modified Sarnat grades and Apgar scores. The *describe* function within the Hmisc package was applied to these character variables to determine not only the count of unavailable values but also the frequency and proportion of choices within each variable. The *table* function (within the R core functions) was used to calculate proportions of each category per variable to create a new vector. The results were displayed in a barplot illustrating the data by frequency or percentage and highlighting any outlier groups.

It's worth noting that although all 145 variables were evaluated, not all variables were included in the results section, as some do not contribute to the primary goal of identifying outliers. For example, the variable 'Visible seizures' was one of several dichotomous 'Yes/No' data points. With these values, the proportion and frequency of each choice, as well as unavailable data points, were computed. However, beyond these values, there was no further information that could be extracted from these variables.

5.2.2.3 Comparing variables between the Inclusion/ Exclusion criteria and Precool instruments

Further analysis was conducted to compare certain data points in the Inclusion/Exclusion criteria against the Precool instruments. This analysis focused on specific variables, including a minimum Thompson score of 7 and a minimum Apgar score of 7 at 5 minutes in the Inclusion/Exclusion criteria instrument and, the Thompson score and Apgar scores at 1 minute, 5 minutes, and 10 minutes in the Precool instrument.

In the Inclusion/Exclusion instrument, there is a character variable which asks the question whether a minimum Thompson score of 7 was observed (Yes/No), while in in the Precool instrument as well as the Thompson HIE Score instrument, actual Thompson scores between 1-17 (out of a total of 22) are captured.

The Thompson scores in the Precool instrument and the Thompson scores at 6h were transformed using R's base logical operators to facilitate a fair comparison. All Thompson score values including and above 7 were replaced with "TRUE" while values below 7 were replaced with "FALSE". The results of this analysis were placed into a new vector and

compared against the Thompson score character vector from the Inclusion/Exclusion instrument.

##in the Inclusion/Exclusion criteria instrument

 # Creating a logical vector for Thompson score results in the Inclusion/Exclusion criteria instrument

Thompson_screen <- Thompson_yesno == 'Yes'

 # Creating a logical vector for Thompson scores in the Precool instrument

Thompson_precool <- Thompson_scores > 7

These strings of code produced two logical vectors populated with TRUE or FALSE values. If 'Yes' was found in the Thompson_yesno vector, TRUE was returned; if 'No', then FALSE was the returned value. These logical values were aggregated into a new vector named Thompson_screen.

In the Thompson_precool vector, a value less than 7 was assigned 'TRUE', while a score equal to or greater than 7 was assigned 'FALSE'. These logical values were compiled into a new vector named Thompson_precool. The same procedure was repeated to create the Thompson_6h vector. Both vectors were then compared to Thompson_screen using the following code:

Thompson_results <- Thompson_precool == Thompson_screen

Thompson_results

Thompson_results2 <- Thompson_6h == Thompson_screen

Thompson_results2

This code compared the values in the two vectors, returning TRUE when the values matched and FALSE when they did not match. The *summary* and *describe* functions were applied to the newly constructed Thompson_results and Thompson_results2 vectors. These functions provided a proportion and frequency of the results, illustrating the occurrence of each value. The results were then presented in a table and visualised using a barplot.

5.3 Results

The purpose of this analysis was to identify outliers and by doing so, demonstrate true errors brought about by data capture faults at the source level. The data from 159 participants were utilised in this analysis. In this chapter, the primary focus was on individual variables.

Table 17 displays the proportion of true errors calculated from the Z-scores. True errors were identified by points lying outside the range of -2 to 2 Z-scores, indicative of being two standard deviations from the mean. The 'Base excess on admission' variable exhibited the highest proportion of true errors at 9.33%, whereas 'Bicarbonate on admission', 'Laboratory: Platelets on admission' and 'Age when cooling commenced' had the lowest proportion of true errors at 0.33% each, as only one outlier was observed for each of these variables.

Table 17: An Overview of the Proportion of True Errors calculated according to the Z-scores

Variable name	**Variable type**	**Total Variables**	**Number of True errors**	**Proportion of true error (%)**
Bicarbonate on admission	Numeric	300	1	0.33
Laboratory: Platelets on admission	Numeric	300	9	3
Base excess on admission	Numeric	300	28	9.33
Lab Hb on admission	Numeric	300	1	0.33
pH on admission	Numeric	300	6	2
Age at admission to cooling centre	Numeric	300	4	1.33
Age when cooling commenced	Numeric	300	3	1
Age at neurological assessment	Numeric	300	2	0.67

5.3.2 Results from the character and numeric variable analysis

Character variables were summarised and are presented in Table 18. Frequencies for the character variables' responses of "Yes" or "No" were computed to showcase the proportion of each response. It should be noted that these proportions exclude unavailable data points,

such as data points “not applicable”, “not done” and “unknown”, for various reasons. Each variable consisted of a data range between 52 to 161 of available data points.

The data points in Table 18 do not exhibit any errors, but they reveal interesting trends regarding the selected data points. In the Inclusion/ Exclusion criteria instrument, certain data points such as ’Apgar score at 5min less than 7’ and ’Clinical seizures’ showed complete data sets with no unavailable values. However, in the ‘Clinical details of baby at birth’ section of the Precool instrument, every variable had four to five unavailable data points.

Table 18: The results of character variable analysis in the dataset

Variable name	Available data points	Unavailable data points	Proportion of Yes (%)	Proportion of No (%)
Inclusion/Exclusion criteria				
Apgar score at 5min less than 7	161	0	82	18
Clinical seizures	161	0	32.3	67.7
Clinical details of baby at birth				
CPAP	156	5	35.7	64.3
Nasal cannulae	156	5	40.4	59.6
Additional oxygen	157	4	59.2	40.8
Mechanical ventilation	157	4	77.1	22.9
Hypotension or inotropes	157	4	97.5	0.25
Active bleeding	157	4	96.8	0.32
Subaponeurotic haemorrhage	157	4	89.2	10.8
Sinus bradycardia	156	5	97.4	0.26
Arrythmia	156	5	-	100
Neurological assessment				
Visible seizures	157	4	27.4	72.6
Electrical seizures at Precool	127	34	28.3	71.7
Electrical seizures at 6hr	52	109	34.6	65.4
Intubation / IPPV available in cooling unit	158	3	98.1	1.9
Invasive BP monitoring in cooling unit	158	3	58.9	41.1
Inotrope infusions in cooling unit	158	3	96.8	3.2

Table 19 presents the remaining variables from the two instruments that did not fall into either numeric or character analysis groups. Essentially, the proportions provided a summary of the data. Notably, 62.1% of the babies in this project were inborn. Furthermore, most of the babies were male accounting (62.1%), and the majority (63.4%) of Modified Sarnat grades taken before 6 hours of life were classified as moderate. Additionally, the most common method for determining gestational age was 'Dates' (39.2%) followed by the 'Ballard' score (22.3%). There were several combinations of different methods for gestational age with Dates and Ballard being the most common combination (6.08%). Lethargy was the most observed sign of encephalopathy before cooling with a proportion of 71.4% and most babies (82.9%) were placed in the NICU level of care.

Table 19: Results of the analysis of the remaining variables from the Inclusion/Exclusion Criteria and Precool instruments

Variable name	Available data points	Unavailable data points	Proportions of known variables
Inborn/outborn status	161	0	- Inborn – 62.1% - Outborn – 37.9%
Sex	161	0	- Female – 37.9% - Male – 62.1%
Signs of encephalopathy before cooling	161	0	- Lethargy – 71.4% - Stupor – 1.2% - Coma – 3.7% - No signs – 23.6%
Method for Gestational age			- Ballard – 22.3% - Dates – 39.2% - EUS <20 weeks – 23.6% - Footlength – 4.73% - Dates and EUS <20 weeks – 2.7% - Dates and Ballard – 6.08% - Dates and Foot length – 0.68% - Dates, Ballard, EUS <20 weeks and Foot length – 0.68%
Source of bloodgas: At birth	134	27	- Cord (arterial) – 5.2% - Cord (mixed source) – 1.5% - Cord (venous) – 37.3% - Infant (ABG) – 56%
Source of bloodgas: Prior to cooling	95	66	- Cord (arterial) – 3.2% - Cord (mixed source) – 2.1% - Cord (venous) – 2.1% - Infant (ABG) – 92.6%
Modified Sarnat grade (<6h)	145	16	- Mild (1) - 8.3% - Moderate (2) – 63.4% - Severe (3) – 28.3%
Number of channels for aEEG assessment	146	15	- One channel – 0.7% - Two channels – 60.7% - Three channels – 9% - Five channels – 29.7%
Method of aEEG measurement	146	15	- Needle – 16.4% - Skin patch – 83.6%
CFM/ aEEG monitoring status	143	18	- Continuous – 80.4% - Intermittent – 14.7% - Not done – 4.9%
Initiation of aEEG	111	50	- At onset of cooling – 18% - Prior to cooling – 68.5% - Other – 13.5%
Cooling method	157	4	- Automated whole-body cooling – 54.1% - Coolcap method – 0.6% - Servo-controlled gel bag method – 41.4% - Other – 3.8%
Level of care	158	3	- NICU – 82.9% - High care – 17.1%

Table 20 presents the results of the numeric analysis conducted on the 'Clinical details of the baby at birth' section of the Precool instrument. Although certain variables such as the 'Ballard score' and 'foot length' exhibited a great number of unavailable data points, basic statistical measures, including mean, median, range and standard deviation were computed on the available data. The analysis revealed that the mean body length of the babies in this project was approximately 49.68cm (SD =3.62cm, Median = 50cm), the mean gestational age was approximately 39.09 weeks (SD=1.55 weeks, Median= 39 weeks), the mean footlength was 75.38mm (SD= 2.39mm, Median=75mm) and the mean birth weight was approximately 3113g (SD=475.90g, Median=3060g). The birth weight exhibited a large standard deviation owing to the fact that the values in this variable ranged from 1865g to 4529g. Furthermore, the median scores for the Apgar score(s) at 1 minute, 5 minutes and 10 minutes after birth were 3, 5, and 7, respectively.

Table 20**:** Results from the numeric analysis of the 'Clinical details of the baby at birth' section

Variable name	**Total data points**	**Mean**	**Median**	**Standard deviation**	**Lowest value**	**Highest value**
Gestational age (weeks)	300	39.09	39	1.55	33	42
Method used for Gestational age	300	-	-	-	-	-
Ballard score	-	38.35	39	2.87	26	42
Footlength (mm)	-	75.38	75	2.39	73	80
Body length (cm)	300	49.68	50	3.62	35	57
Birth weight (g)	300	3113	3060	475.90	1865	4529
Circumference of head (cm)	300	34.66	35	2.02	29	49
Apgar 1 min	300	3.13	3	-	0	8
Apgar 5 min	300	4.99	5	-	1	9
Apgar 10 min	300	6.42	7	-	2	10

The findings from the numeric analysis of the 'blood gas evaluation performed at birth' section of the Precool instrument are presented in Table 21. The point of care (POC) tests exhibited a notable range of between 29 to 60 unavailable data points due to "unknown" or "not done" tests. Several variables such as low glucose, high glucose, pH, sodium, potassium, bicarbonate, lactate and ionised calcium (iCa), demonstrated small standard deviations, indicating low variability and minimal spread from the mean and thus, less

presence of outliers. The mean and standard deviation are equal for base excess suggesting a symmetrical distribution of data points on either side of the mean for this variable.

Table 21: Results from the numeric analysis of the 'Blood gas evaluation at birth' variables

Variable name	Total data points	Mean	Median	Standard deviation	Lowest value	Highest value
Low glucose	300	4.72	4.35	2.03	0.7	12.4
High glucose	300	7.36	6.85	3.32	1.5	24.5
pH	300	7.05	7.07	0.18	6.4	7.44
Sodium	300	133.3	134	6.16	111	154
Potassium	300	5.15	4.84	1.92	3.2	18.2
Bicarbonate	300	11.87	12.25	4.90	-12	24.10
Lactate	300	10.76	10.5	4.34	0.6	25
Base excess	300	-12.68	-15.5	-12.68	-35.9	29
iCa	300	1.29	1.32	0.19	0.67	1.81
Partial Pressure of Carbon Dioxide (PCO2)	300	23.02	11.11	25.08	1.1	159
Partial Pressure of Oxygen (PaO2))	300	66.79	23.95	81.46	1.10	421.7
Fraction of Inspired Oxygen (FiO2)	300	45.76	25	33.21	20.90	100
Variable name	**Total data points**	**Mean**	**Median**	**Standard deviation**	**Lowest value**	**Highest value**
Haemoglobin (Hb)	300	17.96	16.4	17.2	9.2	188
Nucleated red blood cells (NRBC)	300	13.7	9	14.05	0.5	75
White blood cells (WBC)	300	20.85	20.82	8.13	3.18	39.98
Neutrophils	300	15.2	11.94	12.17	0.37	68
Platelets	300	246.7	236	79.22	3.64	517
Thompson score	300	10.22	10	3.47	3	19
Age at Neurological assessment	300	6.88	1.75	43.88	-4.53	529.42

Variables including PCO_2, PaO_2 and FiO_2 exhibited large standard deviations and interquartile ranges signifying that the data points associated with these variables may have a more diverse range of values.

The standard deviations associated with the laboratory values are moderate to large, indicating a wider spread of values from the mean. The 'age at neurological assessment' variable exhibited a substantial standard deviation. The variable ranged from a maximum of

529.42 hours to a minimum of -4.53, strongly suggesting the presence of outliers within this specific dataset. Neurological assessments are typically made within the first 6 hours of life.

Table 22 presents results for the blood gas evaluations at admission or prior to cooling. Notably, the POC tests exhibited more unavailable values compared to the blood gas evaluations performed at birth.

Several values, including bicarbonate, base excess, Partial Pressure of Carbon Dioxide (PCO_2), Partial Pressure of Oxygen (PaO_2), Fraction of Inspired Oxygen (FiO_2) and sodium displayed substantial standard deviations. These large standard deviations suggest the likely presence of outliers within these datasets. Additionally, the laboratory values also exhibited a considerable number of unavailable data points. White blood cells (WBCs) had a standard deviation of 246.49, the largest of the blood gas evaluations at admission or prior to cooling values.

In the 'Cooling induction, method and setting' section, there were fewer unavailable values than in previous sections, except for the variable 'Target core temp' which had 41 unavailable data points. Both the 'age when neonate was admitted' and 'age when target temperature was reached' variables had negative values for their minimum variable. Conversely, the maximum values for 'age when neonate was admitted', 'age when cooling commenced', 'Temp before cooling begun' and 'Age when Target temp was reached' were considerably high. For variables such as 'age when neonate was admitted' and 'age when cooling commenced', strict cutoffs determine when a neonate is cooled; however, but some exceptions exist.

Table 22: Results from the numeric analysis of the Blood Gas Evaluations at Admission or Prior to Cooling

Variable name	Total data points	Mean	Median	Standard deviation	Lowest value	Highest value
Low glucose	300	4.49	4	1.88	0.7	10.8
High glucose	300	7.83	6.6	2.22	2.4	14.2
pH	300	7.23	7.23	0.14	6.77	7.54
Sodium	300	132.7	133	9.55	74	160
Potassium	300	4.65	4.5	1.24	3.1	13.5
Bicarb	300	15.98	14.5	12.99	3.90	128
Lactate	300	9.47	8.9	4.67	1.6	21
Base excess	300	-9.16	-11.8	10.83	-29.3	21.3
iCa	300	1.24	1.24	0.24	0.23	2.44
Partial Pressure of Carbon Dioxide (PCO2)	300	24.2	15.24	13.8	2.3	63.4
Partial Pressure of Oxygen (PaO2))	300	55.53	15.6	62.7	1.9	235
Fraction of Inspired Oxygen (FiO2)	300	36.69	21	26.4	20.9	100
Variable name	**Total data points**	**Mean**	**Median**	**Standard deviation**	**Lowest value**	**Highest value**
Haemoglobin (Hb)	300	16.62	16.90	2.41	9.6	22
Nucleated red blood cells (NRBC)	300	15.38	8	20.15	1	84
White blood cells (WBC)	300	50.18	20.37	246.49	5.09	2082
Neutrophils	300	13.43	11.77	10.20	2.35	67.91
Platelets	300	240.3	234.5	78.04	36.4	472
Age when neonate was admitted	300	2.28	1.59	1.99	-0.15	13.5
Age when cooling commenced	300	3.98	4.16	1.89	0.25	17
Age when Target temp was reached	300	8.8	7.17	6.36	-11.22	39.53
Temp before cooling begun	300	35.88	36.4	1.25	28.9	37.8
Target core temp	300	33.56	33.5	0.77	33.3	37
Ratio of babies to nurses	300	2.86	2	1.38	1	6

5.3.3 Boxplots

Boxplots provided visualisations of the distribution of data, including the central tendency, spread and potential outliers. The outliers appear as data points that fall outside the "whiskers" of the boxplot. The Z-score plots visually compares each data point to the dataset mean. Points above the mean have positive Z-scores, while those below have negative scores. A threshold line at -2 and +2 delineates regions of importance.

The boxplot in Figure 31 reveals two points that are clear outliers. The minimum Ballard score is 35, with the majority of data points above this threshold. The mean is slightly smaller than the median, indicative of a more positive skewed distribution of the data. Figure 32 illustrates five data points exceeding the threshold of 2, signifying observations lying beyond 2 standard deviations from the mean. The bulk of data points fall within the range of Z-scores of -1 and 1, exhibiting a clustering tendency.

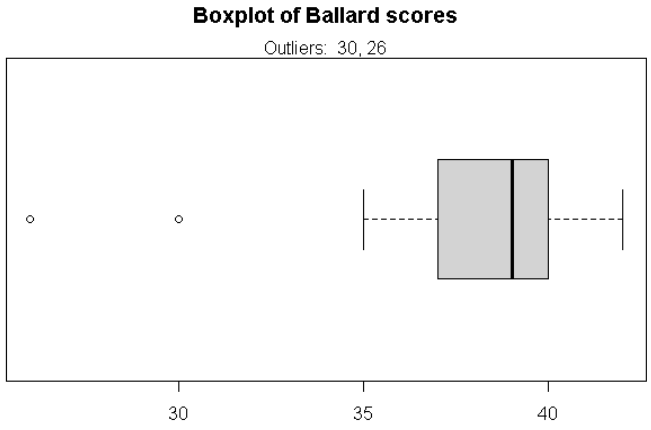

Figure 32: Boxplot displaying the distribution of Ballard scores and outliers

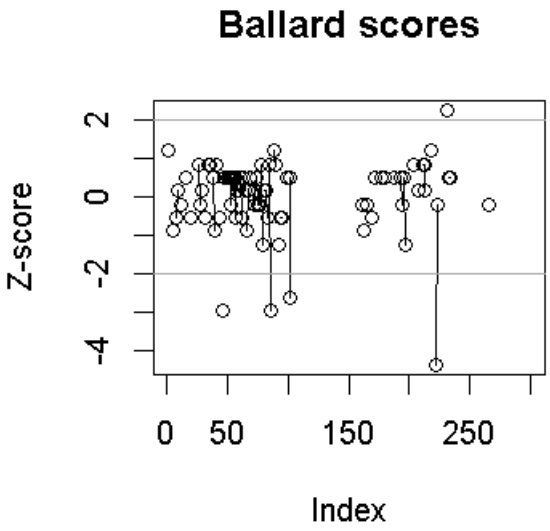

Figure 31: Z-score plot displaying the Ballard scores

In Figure 33, the distribution of birth weights for the babies in the dataset primarily ranged from 2810g to 3402g, with a mean weight of 3113g. The boxplot indicates right-skewness, as evidenced by the clustering of outliers primarily on the right tail end of the plot. The majority of data points fall within the range of Z-scores between -2 and 2, highlighting the variability in birthweights. In contrast to Figure 33, Figure 34 depicts approximately ten data points as extreme values or outliers.

c

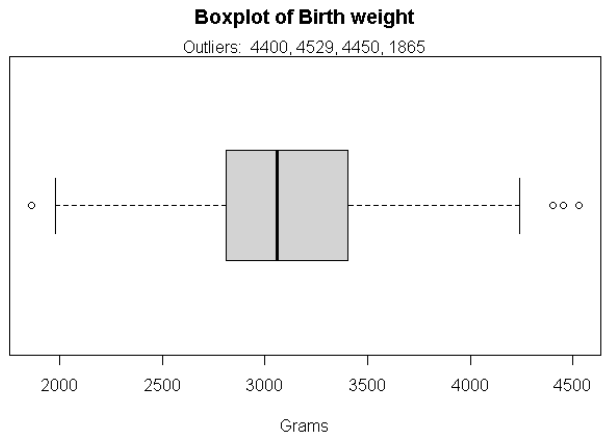

Figure 33: Boxplot showing the distribution of the birthweights of babies in the study

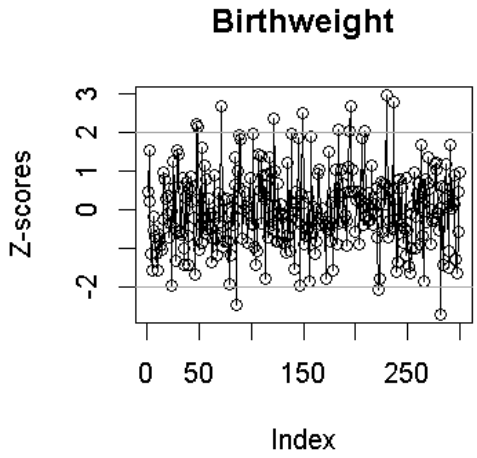

Figure 34: Z-score plot displaying Birthweight

Figure 35 exhibits a pronounced positive skewness, with a significant extreme outlier value of 128, far exceeding the dataset's maximum value. This outlier represents a noteworthy deviation from the bulk of the data points, indicating a unique observation that was a potential anomaly. Furthermore, the Z-score plot in Figure 36 provides confirmation of the extreme outlier noted in Figure 35, positioned approximately twelve standard deviations from the mean, emphasising its significance within the dataset.

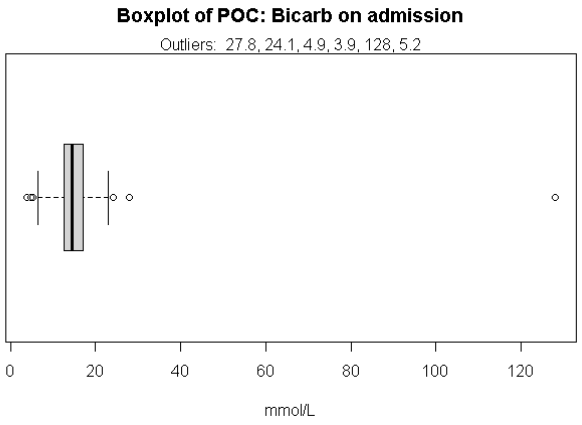

Figure 35: Data distribution of POC Bicarb values measured on admission

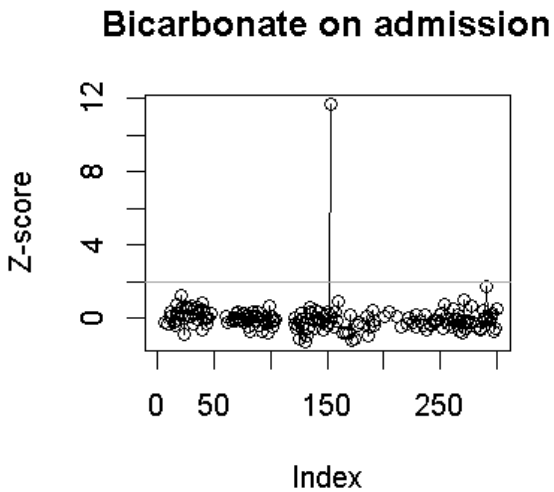

Figure 36: Z-score plot displaying Bicarbonate on admission

In Figure 37, a limited number of outliers are visible, notwithstanding the median and mean exhibiting a near equal distribution. Despite the presence of these outliers, the central tendency of the dataset appears relatively stable. In Figure 38, the distribution of the majority of pH values is concentrated around the '0' mark, indicating close proximity to the mean value. However, several outliers were observed beyond the -2/2 threshold, suggesting deviations from the central tendency of the dataset.

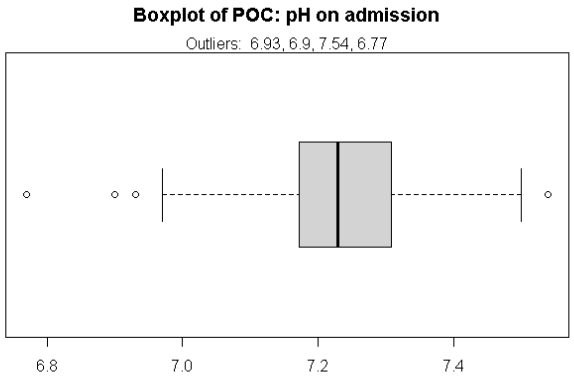

Figure 37: Boxplot showing the distribution of data for pH at admission

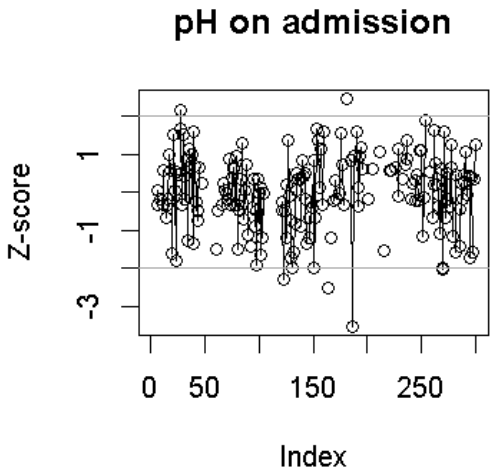

Figure 38: Z-score plot of pH on admission

In Figure 39, a single outlier of 13.5 hours of age for the age at admission to cooling centre variable is evident, representing a value twice the upper limit, while the rest of the data points fall within a range of 0 to 6 hours. In Figure 40, one additional outlier is observed compared to Figure 39. The bulk of the ages lie within the range of Z-scores from 0 to 2.

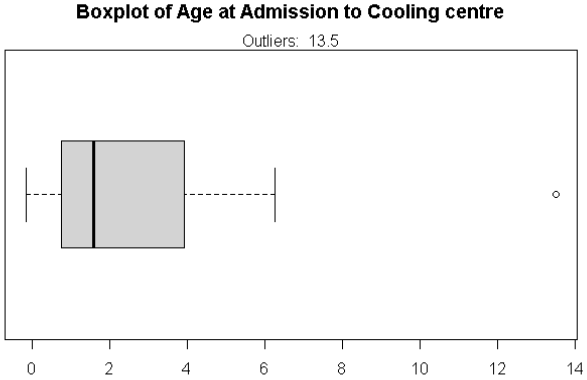

Figure 40: Data distribution of the neonatal age on admission to the cooling centre

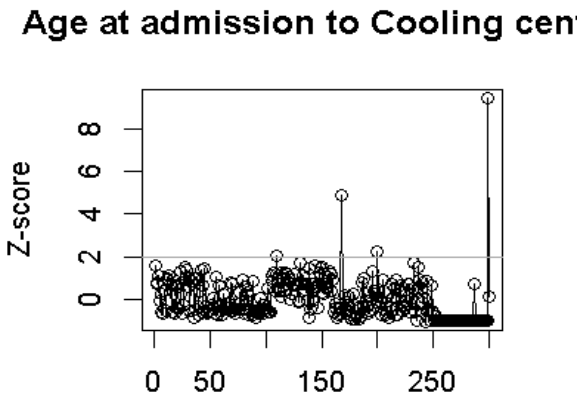

Figure 39: Z-score plot of age on admission to the cooling centre

In Figure 41, outliers are evident in the neonatal age at neurological assessment variable, showcasing values such as -4.53, 9.0 and 12.5 hours, alongside extreme observations like 25.167, 24.25, 122.5 and 529.4167. These observations exert significant influence, causing compression of the data within the box and inducing a rightward skew in the distribution. However, closer examination of the Z-score plot displays only two data points as extreme values, with one of these points extending beyond the Z-score threshold of 15.

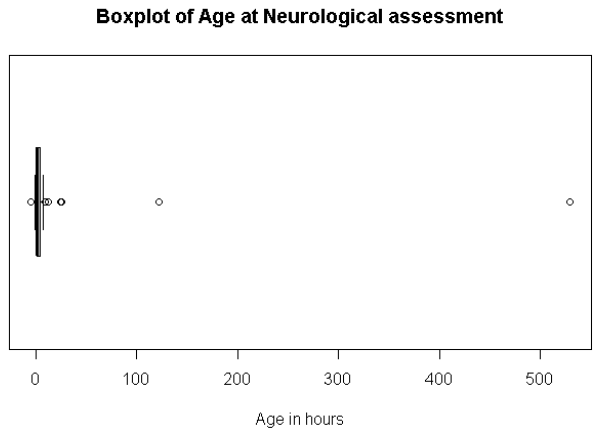

Figure 41: Data distribution of the neonatal age at neurological assessment

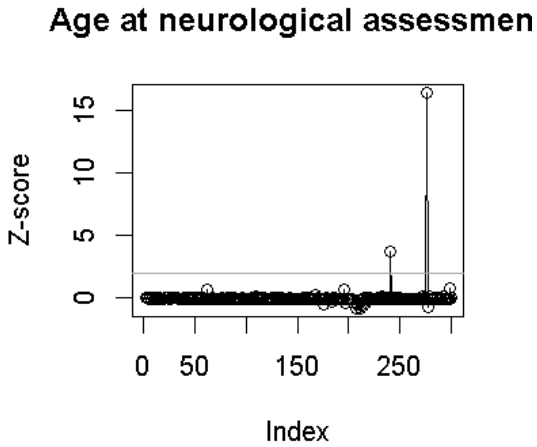

Figure 42: Z-score plot of the neonate age at neurological assessment

In Figure 43, two values stand out from the typical range of observations for the variable neonatal age when cooling commenced at the cooling centre, manifesting as outliers that extend more than 5 hours beyond the maximum value. Such extreme deviations suggest the possibility of errors in the date and/or time recording process. Figure 44 confirms the extremity of these values, depicting their placement beyond 3 Z-scores and extending all the way beyond 10. This indicates that these ages surpass 10 standard deviations from the mean, emphasising their potential for substantially skewing the dataset.

Figure 43: Data distribution of the neonatal age when cooling commenced at the cooling centre

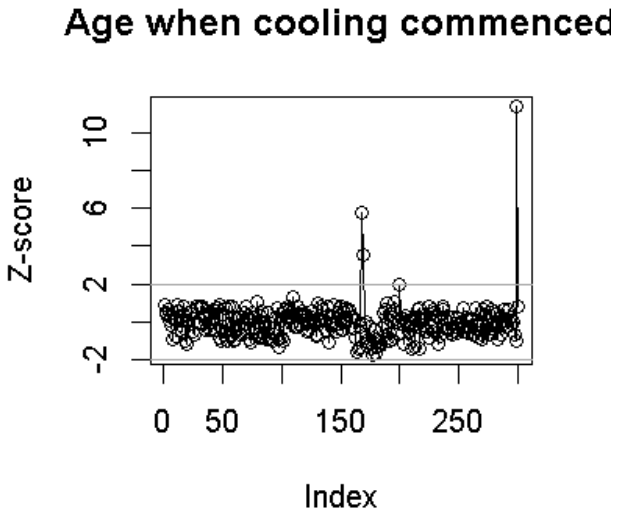

Figure 44: Z-score plot displaying the ages of neonates when cooling commenced

Figures 45 and 46 reveal multiple outliers positioned at both ends of the whiskers in each plot of the laboratory platelet assessments, strongly indicating the presence of substantial variability within the datasets. In Figure 45, a low outlier of 3.64 is observed, while the remaining outliers extend beyond the maximum value for the dataset, indicating significant dispersion in the data distribution. In Figure 46, most values fall within the -2/2 threshold, with seven outliers exceeding the upper threshold and one outlier falling below the threshold.

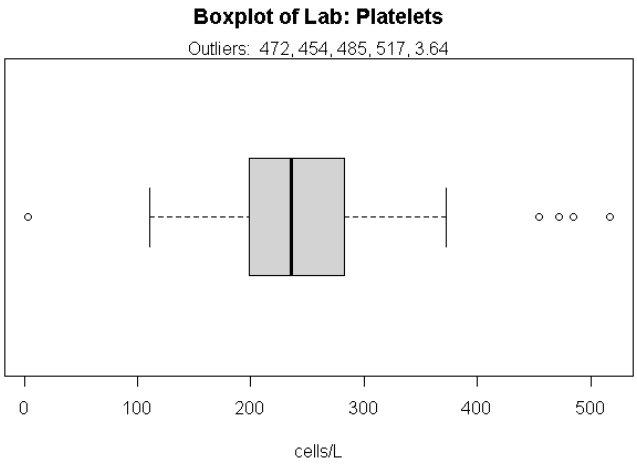

Figure 46: Data distribution of the Lab: Platelets variable

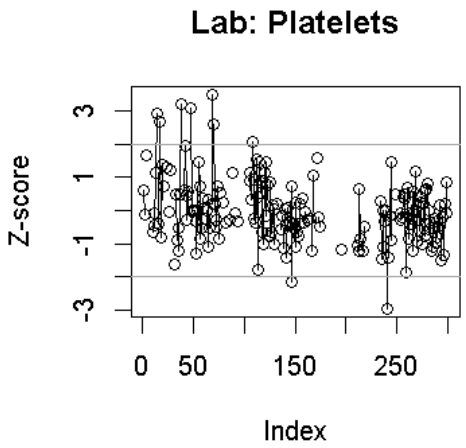

Figure 45: Z-score plot displaying Platelet values

In Figure 47, a single outlier is observed at the minimum end of the boxplot for the POC base excess variable, while the majority of outliers are situated on the right tail, resulting in a rightward skew of the data. In Figure 48, only an upper threshold is evident, around which a cluster of data points from position 250 onwards is observed. This concentration of data points suggests a distinct pattern of outlier distribution.

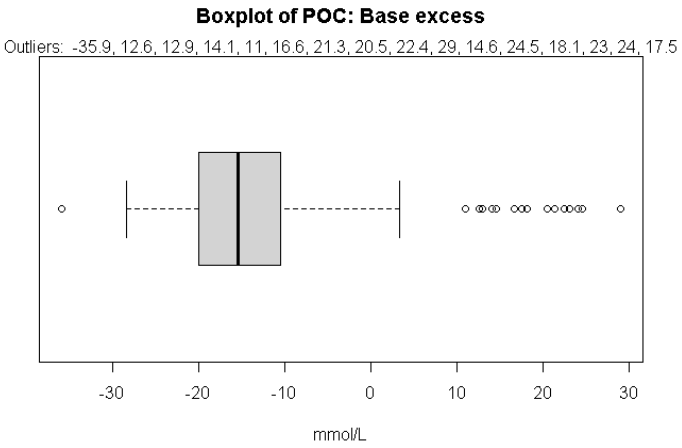

Figure 47: Data distribution of the POC: Base excess variable

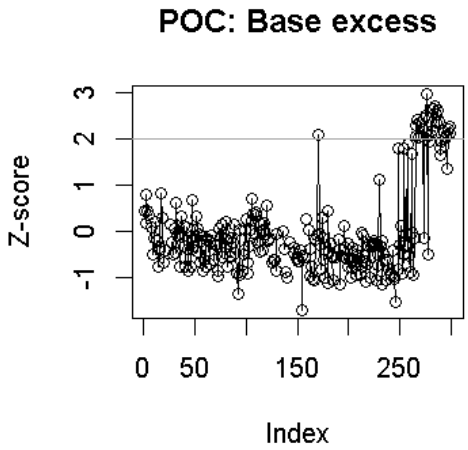

Figure 48: Z-score plot presenting the POC: Base excess variable

The haemoglobin dataset in Figure 49 spans from a minimum value of 9.20 g/dL to a top whisker extending up to 18.18 g/dL. The mean laboratory haemoglobin value of 17.96 g/dL provides a central measure; however, it is overshadowed by a striking outlier of 188 g/dL. This extreme value significantly deviates from the bulk of the data points, as evidenced in Figure 49. Only one outlier is discernible among the laboratory haemoglobin values, while the remainder cluster around the Z-score of 0. This pattern suggests a concentration of values around the mean, with the outlier standing out distinctly due to its significant distance from the mean.

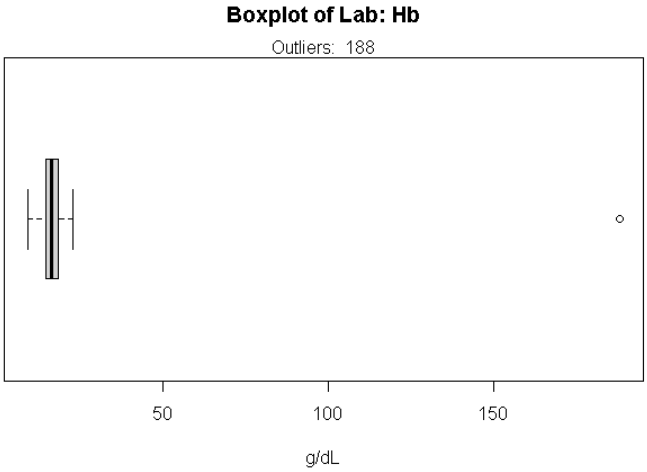

Figure 50: Data distribution and outliers for the Lab: Hb variable

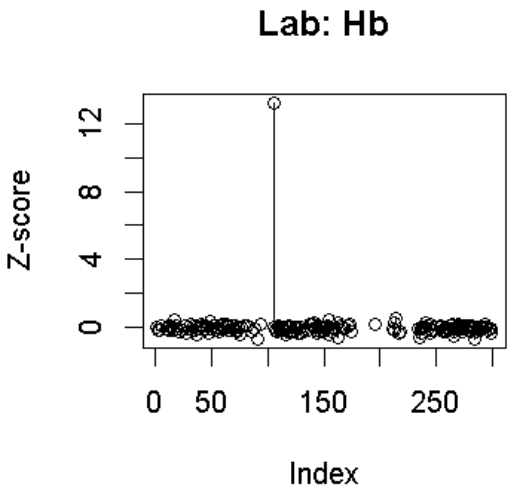

Figure 49: Z-score plot illustrating the variable Laboratory Haemoglobin

5.4 Discussion

The process of identifying outliers in this analysis has proven fruitful in providing valuable insights into the dataset. While both boxplots and Z-score plots were utilised to demonstrate the presence of outliers, an actual true error rate ranging between 0.33% and 9.33% (as shown in Table 16) was observed among the variables containing extreme data points.

In most cases, it is necessary for outliers to be eliminated from a dataset to prevent interference with data analysis; in this project, outliers have offered additional information about the dataset. Numerous extreme values were present in various datasets. While some were found to be inconsequential, other extreme values were significantly distant from the normal range.

For instance, a “low” birthweight of 1800g represents an inconsequential but genuine extreme value since it met the minimum inclusion criteria for the study. However, this data point exhibited a considerable deviation from the median. Conversely, in the neurological assessment of the newborn section, an observed neonatal age of -4.53 hours suggests an anomaly that could indicate a true error.

Analysis of the variables in the Screening and Inclusion/ Exclusion criteria and Precooling instruments revealed proportions and frequencies, shedding light on the most likely data

captures. Additionally, unavailable values were identified across various variables, encompassing unknown data, not applicable data, and instances where tests were not performed. Comparatively, the Screening and Inclusion/ Exclusion criteria instrument had fewer unavailable data points than the Precool instrument. The observation of anomalies in the different instruments could offer insights into areas that might require tighter design control. While most variables exhibited between 3 and 5 unavailable data points, 'Electric seizures at precool' and 'Electrical seizures at 6H' displayed 34 and 109 unavailable data points, respectively. The elevated rate for these variables can be attributed to the delayed introduction of the 'Electrical seizures at 6h' variable during the study, as well as the variable being tested infrequently across all study sites.

The numeric analysis, predominantly within the Precool instrument, provided significant insights. Variables like birthweight and 'age at neurological assessment' exhibited large standard deviations, indicating considerable variability within these datasets. Notably, the 'age at neurological assessment' ranged from -4.53 hours to a maximum value of 529.42 hours. Considering that most babies underwent neurological assessments between 0 and 6 hours, these results clearly indicate the presence of outliers. Similar patterns were observed in other variables recording age in hours, including 'age when neonate was admitted' and 'age when target temperature was reached'. These anomalies likely stem from genuine errors in the recording dates and times for these particular variables, necessitating a thorough investigation of the source records, such as the hospital records on site.

Several data points, including white blood cells and platelets, reached maximum values of 2082 cells/L (reference range: 18±8cells/L) and 517 cells/L (reference range: 100-450 cells/L), respectively[5]. Compared to other data points within each set, these figures stand out as extreme values.

The boxplot analysis distinctly identified extreme values and outliers, highlighting their influence on the symmetry of the boxplots. Boxplots representing Ballard scores and birthweight exhibited both data spread and central tendencies, revealing a few outliers at each tail end (refer to Figures 31 and 33). However, some extreme values were so prominent that they compressed the boxplots, making it more challenging to visualise the data distribution between the tails and other existing outliers (see Figures 35 and 41). The outlier observed in the boxplot of haemoglobin (Figure 49) is a clear representation of a

probable true error, likely originating from a misreading of the hospital records during data entry. It is likely that the intended value was “18.8g/dL” rather than “188g/dL”, a value that aligns with the dataset and falls within the normal range for this variable.

Similarly, the standardisation of data achieved by Z-score plots, which measured the number of standard deviations a data point was from the mean, facilitated a comparison of outliers across different variables. This standardised approach allowed for a more comprehensive understanding of the significance of outliers within the dataset. For instance, certain variables such as ‘Laboratory: Haemoglobin’ and ‘Neurological assessment at neonatal age’, exhibited outliers that deviated beyond the selected threshold of -2/2, thus identifying these variables as true errors. These outliers, with Z-scores of 12 and 15 respectively, were clear anomalies within their respective datasets, indicating a relative severity that warrants prioritised attention for further investigation of hospital records.

Moreover, the Z-score plots were instrumental in identifying potential data heterogeneity, as illustrated in Figure 36 depicting Bicarbonate on admission. The visualisation provided by Z-score plots aided in pinpointing variables with homogenous data distributions. In contrast, the boxplot in Figure 35 displayed only a few outliers. However, upon closer examination with Z-score plots, only one outlier was identified, while the majority of data points appeared homogenous within the provided threshold. A similar trend was observed with neonatal birthweights which provided an example of a genuine extreme value. While the boxplot in Figure 33 revealed one outlier; a closer look using the Z-score plot in Figure 34 identified additional outliers and showcased a more normal, homogenous spread of data.

Examining the discordance observed in the Thompson scores suggested a true error rate of 1.31%. While most entries aligned between the Inclusion/ Exclusion criteria and the Precool instruments at both time points (prior to 6 hours and at 6 hours), the 1.31% hints at a potential data entry error during the transcription process from hospital records to the NESHIE study CRFs or due to user error.

5.4.1 Limitations and Future Directions

Sample size was contingent upon the availability of data points per variable. While variables with Yes/No responses generally exhibited minimal to no unavailable data points, numeric

variables contained numerous unavailable values, leading to a reduced sample size for these data points.

Addressing any noted discordance could be achieved by implementing coding prompts within REDCap. If there is concordance, the end user will be allowed to proceed, however, if discordance is noted, then an HTML prompt could display to alert the user of the issue and advise a corrective course.

5.5 Conclusion

The presence of extreme values or outliers in the datasets explored in this project implied potential errors in the data, which could stem from recording, transcription, or data collection inaccuracies. Among these outliers, some were genuine values, anomalies or true errors. The genuine outliers often represented valid extreme values, as observed in the outliers within the birthweight dataset. However, a majority of outliers appeared to be anomalies, particularly in multiple POC and laboratory values. These anomalies will require confirmation at the hospital level, although some values may indeed be considered true errors. Other instances of true errors included negative neonatal ages and values outside a specified range for the variable, for example, a neonatal age of 529 hours.

The existence of outliers and extreme values across various data points poses a challenge for the broader NESHIE study, especially since true errors can influence the mean and inflate standard deviations. This is of particular significance as it suggests that a dataset with outliers might not be suitable for predicting or estimating values based on the mean.

5.6 References

1. Kabacoff RI. R in action: Data analysis and graphics with R. 2nd ed. Shelter Island: Manning Publications; 2015.

2. Osborne J, Overbay A. The power of outliers (and why researchers should always check for them). Pract Assess Res Eval. 2004; 9:1-12

3. Maindonald J, Braun W. Data analysis and graphics using R: An example-based approach. Maindonald J, Braun J, editors. Cambridge University Press; 2010.

4. Kaliyaperumal S, Kuppusamy M, Arumugam S. Labeling methods for identifying outliers. International Journal of Statistics and Systems. 2015; 10(2):231-8.

5. Bates I. 2 Reference ranges and normal values. In: Bain BJ, Bates I, Laffan MA, editors. Dacie and Lewis Practical Haematology (twelfth edition): Elsevier; 2017. p. 8-17.

Chapter 6: Discussion

6.1 Introduction and summary of findings

This project was conceived with two objectives: first, to establish a robust database capable of capturing and managing extensive data across multiple national sites, and second, to enhance the process of assessing data quality for improved accuracy. Achieving these goals was contingent upon the development of a data collection system meeting stringent management criteria. Specific objectives included defining clear data capture parameters, optimising the interface for investigators to reduce errors, promoting data verification and incorporating statistical analyses for outlier detection.

The project involved quantitative and comparative analyses on both existing and newly developed databases, particularly focusing on the amendment versions, sites, sub-sections of CRF data and investigator-related aspects to discern trends.

To fulfil these objectives, REDCap was selected due to its capacity to handle extensive data, approximately 1500 data points per research subject in this case, and its built-in features for data quality evaluation. A beta version of the REDCap database was already in use before the commencement of this project, forming the foundation for the new design, as described in the methodology chapter. The initial database, named A3, accommodated the vast data generated during the NESHIE study. Recognising the need to ensure data quality and integrity and to mitigate errors, database architects found it necessary to implement design enhancements predominantly through REDCap's built-in capabilities along with the incorporation of advanced features such as piping, HTML and matrices of fields. These features not only enabled the demarcation and grouping of data points, allowing for a systematic presentation of information to investigators, but also retained robust data-capturing capabilities. Furthermore, they facilitated user interaction by offering comprehensive guidance throughout the data capture process. The incorporation of user-friendly control mechanisms ensured that investigators could navigate through the data capture process with ease, enhancing overall efficiency and accuracy.

Moreover, the database life cycle phases served as a foundational framework for managing the database design process. By adhering to this blueprint, essential guidance was provided for the meticulous design of the A4/A5 and A7 databases. This structured approach ensured

that key considerations were addressed at each stage of development, facilitating the creation of databases optimised for functionality, usability, and data integrity.

The methodology chapter provided a comprehensive overview of the database evolution, tracing its development from the initial A3 database to the subsequent iterations of A4/A5 and lastly, A7. Central to this evolution were the evolving requirements and demands of the study, which necessitated the integration of new elements and features into the database architecture. This chapter revealed differences between each iteration, highlighting shifts primarily in aesthetic aspects. Descriptive and comparative analyses, as detailed in the first analysis chapter, shed light on these differences offering valuable insights into the evolution of the database structure. Moreover, the methodology chapter delved into the identification and resolution of compliance issues stemming from data entries, particularly focusing on incomplete and inaccurate data. Descriptive statistics were instrumental in visualising the distribution and trend of errors within bar plots. For example, within the analysis of the CRF instruments, some instruments exhibited a higher error incidence, signifying that these areas required specific attention for enhanced database design. Patterns among the errors made by investigators provided insights into training needs and prevalent trends in errors suggested potential solutions on the database end, such as incorporating reminders for data completeness.

While some errors were common across all the study sites, some were site-specific, and others were specific to particular amendments. Most observed errors stemmed from missing data points from either the CRF or REDCap rather than from data captured incorrectly to REDCap. This finding suggested that one of the solutions to this particular issue could be as simple as incorporating reminders in the database to ensure the completeness of data points during data capture as well as reminders to upload the relevant CRFs. In some cases, the database could be configured to prohibit further steps unless all essential elements were completed, a feature implemented in A5 and solidified throughout A7. For these reasons, the new A7 database not only showcased aesthetic appeal but, more importantly, demonstrated enhanced functionality compared to previous iterations. It boasted an easy-to-navigate interface, simplified troubleshooting and offered corrective measures where and when necessary.

From the implementation of the beta A3 database, data capture errors were noted and promptly addressed. However, their persistence led to the establishment of error reports to ascertain whether user error and insufficiency in the CRF design and/or database design were contributing factors. These error reports categorised errors as “incorrect” and “missing”, while also recording data points that were “unknown” and “not applicable”, alongside “correct” data points. The ‘error’ categories, “missing” and “incorrect”, were assessed to determine the database error rate for each amendment iteration (A3 to A7), addressing the question of whether the implemented database design effectively reduced the error rate.

The hypothesis tested whether the mean error rate in A7 was lower than in the early amendments, particularly A3. The A7 database recorded an error incidence of roughly 6%, slightly higher than the predetermined study threshold of 5%. Nevertheless, this represented a reduction from the initial rate of approximately 9%. Statistical analysis using chi-square demonstrated a significant difference between A3 and A7 ($p = 2.083 \times 10^{-6}$; $p<0.05$) as well as A4 and A7 ($p = 0.00$; $p<0.05$), underscoring the meaningful reduction in errors. This finding was significant, but equally noteworthy were the trends observed in data capture and the occurrence of errors. Error rates were examined at the level of amendments, sites, investigators and instruments to offer a thorough overview of patterns that might reveal deeper underlying issues. Some investigators had high error rates which contributed to the site’s overall high error rate, while in some cases it was noted that certain instruments experienced recurring errors across all study sites which could be indicative of a general misunderstanding of the CRFs.

The final project component aimed to distinguish between errors arising from data entry and true errors stemming from the transcription of hospital records onto study CRFs. Exploration of the dataset revealed the presence of outliers and extreme values in several variables. Some outliers represented extreme values but were clinically valid. These instances were noteworthy as they characterised that a dataset could contain a broad range of values, thus demonstrating the diversity of variables present. However, other variables displayed anomalies, representing true errors that markedly deviated from the dataset. These outliers were identified as instances where values either matched or were consistent between the CRFs and REDCap. This indicated potential errors that likely occurred prior to

the data entry step, underscoring the importance of careful consideration in data interpretation and calls for a thorough investigation into hospital records for validation. Alternatively, implementing predefined ranges or thresholds for each numeric or textual data point in REDCap could be considered. Data managers could then regularly review these parameters to identify any values that fall outside the specified range. This systematic approach ensures data accuracy and integrity, allowing for timely identification and resolution of discrepancies.

6.2 Summary and Interpretation of Findings

The findings of database design revealed that implementing HTML, matrices of field and advanced branching logic proved to be effective and seamlessly integrated into the system. These features allowed for easy replication and customisation to suit the needs of the study.

The structured approach of the database life cycle, put together from existing literature and tailored for this project and the NESHIE study, offers a replicable model for other research groups. Therefore, this approach can be used as a model for other studies. The flexibility of REDCap allowed researchers to create tailored data collection instruments to meet the specific needs of their study. This enables the adaptation of forms to accommodate the complexity and diversity of large-scale research projects. It boasts features such as data validation rules and calculated fields to streamline data entry and ensures data integrity to help researchers maintain consistency and accuracy across a vast amount of data.

The methodology of ascertaining error proportions to determine an amendment error rate is applicable across various clinical research settings. This thorough approach not only enhances confidence in the data but can also speak to the overall quality of the study's findings.

Finding anomalies such as recording an age of 529.4167 hours for a participant's neurological assessment point for a participant. These outliers, including instances surpassing 25 hours of age, suggest discrepancies likely emanating from transcription errors or incorrect data capture protocols.

The implication is that inaccuracies in date and time can significantly impact the integrity of the entire dataset, affecting subsequent data analysis and interpretation. This underscores

the critical importance of implementing robust data validation procedures and ensuring accuracy at every step of the research process.

In terms of instruments, the error rate was higher in the Precool instrument than the Screening and Inclusion/Exclusion criteria instruments. The data in the Screening and Inclusion/Exclusion criteria instruments were mostly composed of Yes/No questions as well as radio questions (see Appendices B-E) while the Precool instrument had a mixture of selection options. A salient difference was that the Precool instrument contained text boxes for recording text responses (numeric and text). It is possible that typographical mistakes, spelling errors and unclear handwriting could have contributed to some of these errors.

Most REDCap projects are listed on the REDCap consortium, but no information is given on the specifics of each project. The reliance on training videos, tutorials and blogs for fine-tuning information underscores the collaborative nature of REDCap and the decentralised nature of knowledge dissemination within the community. The listing of REDCap projects on the consortium provides a centralised repository for researchers to browse existing projects, fostering transparency and facilitating collaboration. However, the lack of specific project details may limit the utility of this resource for researchers seeking in-depth insights or guidance on project design and implementation. In this case, blogs and web pages serve as invaluable platforms for sharing best practices, troubleshooting common issues, and exchanging tips and tricks for maximising the utility of REDCap.

Aesthetics are integral to database design as they directly impact user experience and engagement, capturing the attention of end users and encouraging their continued participation in database activities. Design features were meticulously crafted to ensure visual appeal, accessibility, and comprehensibility, allowing key points to stand. This included strategically placing emphasis on data entry fields, minimising empty fields, and ensuring that every data point had a clear selection or input. However, while aesthetics are crucial, data validation through rigorous statistical analysis remains paramount.

Despite site-specific trends, the study revealed overarching significance with broad implications. Common patterns emerged that transcended specific sites, highlighting consistent challenges and opportunities for improvement. Errors encountered during data analysis, rather than being viewed as setbacks, were embraced as opportunities for learning

and discovery. Through the examination of these mistakes, valuable insights were gained, informing the integration of new features to enhance future databases. The trajectory suggests a continuous decrease in error rates over time as researchers refine their processes and enhance their understanding of data management principles.

Furthermore, an evaluation of the data capture process is imperative to refine data quality evaluation and enhance accuracy. Beyond simple comparison for concordance, a deeper analysis revealed underlying errors within the data itself, originating from discrepancies between hospital records or source documents and the study's CRFs, making them challenging to identify.

Lessons learned from these experiences underscore the importance of open dialogue and collaboration within multidisciplinary teams. For instance, an apparent error regarding neonate length sparked a meaningful conversation with the clinical team, ultimately leading to the inclusion of previously overlooked data points. Similarly, the discovery of other errors prompted further discussions, highlighting the complexities inherent in running a national study with diverse stakeholders. These conversations serve as opportunities to align understandings, foster harmony, and promote standardisation across teams from varying backgrounds. Through such collaborative efforts, the study strives to achieve a unified approach to data management and analysis, ensuring the integrity and reliability of research outcomes.

The presence of other errors has prompted more conversations which are testament to the fact that when running a national study, it can be challenging to merge multiple understandings. These conversations allow many people, all teams from various backgrounds to get on the same page about the meaning of a data point. This is one step to achieve harmony and standardisation.

Comparison with Existing Literature

Most research articles highlight the versatility of REDCap in various fields. Researchers employ REDCap to design data collection forms, manage participant data, and analyse outcomes. Key features such as branching logic, automated validation and export capabilities are frequently reported in the literature. These features empower research teams to overcome challenges and achieve their research goals. However, the way the

NESHIE database, developed within the REDCap framework, deviates in some aspects from conventional usage. The team of Bangdiwala and Boulware [1] harnessed branching logic, piping and action tags uniquely. In contrast, other research teams, exemplified by Hossmann, Haynes, Spoerri, Diatta, and Aboubacar B (2019) found validation rules and data control features more relevant [2]. Martin-Willetta's projects integrated surveys, validation rules and branching logic [3].

While scholarly articles often provide high-level descriptions of feature use, a practical understanding was facilitated by training videos available on the REDCap site and the comprehensive REDCap user guide created by the REDCap consortium. Additionally, insights from blogs and affiliate universities' REDCap help websites were instrumental mastering the intricate code strings required for project implementation.

There is a lot of variance in what is considered to be an acceptable database error rate. Determining an acceptable database rate hinges on contextual factors, data type, and potential implications for study outcomes [4]. Errors can be random errors (due to chance), systematic (due to inaccuracies in data collection), or related to missing data. The general trend is to set the threshold at a low level to ensure reliability and validity of study findings. The main aim for researchers is to achieve error rates that are statistically insignificant or fall below a predetermined significance level. By minimising the likelihood of false-positive or false-negative findings, researchers enhance the robustness of their results.

6.2.1 Implications of Findings

Several unforeseen benefits emerged during this study, as outlined below:

a. The versatility of REDCap features

   The setup and organisation of REDCap projects involved dividing data collection forms into separate instruments. By structuring projects with separate instruments, researchers can modularise data collection, where different aspects of the study such as demographic information and medical history could be collated. This segmentation was used to capture data into distinct instruments, ensuring clarity through the data capture process and additionally to enhance data integrity, allowing for targeted validation rules and quality checks.

b. The remote accessibility of REDCap

REDCap's remote accessibility was advantageous as researchers could access and enter data from various locations, improving efficiency. This meant that there were several benefits to its accessibility, enabling researchers to quickly get up and running with adequate training or technical expertise. In addition, it is flexible, timely and reduces paperwork. But the aspect of a discerning platform for real-time data entry error recognition is unique.

c. REDCap features utilisation in quality control

Through validation rules, automated alerts and quality control, researchers could define rules, flag discrepancies and real-time error recognition which minimises downstream data cleaning efforts. Clinical research is dynamic and evolves rapidly due to scientific advancements, changing protocols, and emerging technologies. A way to establish quality control parameters to ease the workload and burden could be valuable associated with data management and ensure research integrity. The detection of errors in real-time could further enhance this process, enabling researchers to intervene swiftly and mitigate any adverse effects on study results.

6.2.2 Data management practices in low-resource settings

In many African settings, where resources are often scarce and reliance on paper-based systems remains prevalent, the transition to electronic methods presents a formidable challenge. However, this shift holds the potential to significantly benefit researchers by circumventing the costs, resource constraints, security measures, the challenges of maintaining a secure storage system, vulnerability risks of loss, damage or unauthorised access, logistical challenges, accessibility issues, limited training opportunities and the resource-intensive monitoring associated with printing physical CRFs.

Despite all these benefits, effecting change within hospital protocols deeply entrenched in longstanding practices can be resistant to change[2]. Influencing modifications in well-established procedures can be an uphill battle requiring persistent effort and strategic persuasion. Nevertheless, the move toward electronic systems could be a positive step for streamlining processes and revolutionising data management practices within healthcare institutions[2].

Interestingly, one of the site PIs shed light on the potential impact of electronic surveillance on clinical outcomes. Unofficially, the site PI observed a decreased incidence rate of NESHIE

in the neonatology unit since starting the study. Although it was stated in passing conversation, she speculated that this reduction might be attributed to increased surveillance. In other words, involvement in the study could be influencing health practitioners to be more vigilant and focused on their work. While this phenomenon aligns with existing literature, it remains contingent on the specific research intervention and the characteristics of the studied population.

Navigating the transition from paper-based to electronic systems in resource-limited contexts requires strategic planning, adaptability, and a keen understanding of local dynamics. Researchers must carefully weigh the potential benefits against the inherent challenges, all while considering the unique context in which their study operates. Embracing innovation while navigating the intricacies of the local landscape, researchers can harness the transformative power of electronic systems to advance healthcare delivery and research outcomes in resource-constrained settings.

6.3 Limitations and Methodological Considerations

Several limitations were encountered that impacted the data collection process and methodology. These limitations are important to acknowledge and consider when interpreting the study results:

a. Site 6 Incomplete Data

Site 6 presented significant challenges due to predominantly incomplete data, leading to a high incidence of errors. These errors were primarily due to missing and unknown values across all amendments. As a consequence, the reliability of the data collected from this site was compromised and only a few study IDs could be used to perform analysis.

b. Low Participant Recruitment at Sites 5 and 6

Both sites 5 and 6 faced limitations in participant recruitment, with a low number of participants enrolled in the study overall. Additionally, there were specific amendments where participant recruitment was particularly limited, further impacting the overall sample size and potentially affecting the generalisability of findings.

c. Maternity Leave Impact at Site 3

At Site 3, the absence of the investigator due to maternity leave posed a limitation on the review of A7 documents, potentially affecting data completeness and comprehensiveness for this amendment.

d. Incomplete Data at Site 4

Site 4 encountered challenges with incomplete data, particularly in A5 and A7 documents. These gaps contributed to a high incidence of errors, primarily in the form of missing and unknown values. These data integrity issues posed significant limitations to the reliability and validity of the findings derived from this site.

6.3.1 Methodological Considerations, limitations and Future studies

The online design of our study introduced certain vulnerabilities as reliance on internet connectivity left the database design prone to mistakes in the code. Oversight during data collection may have been more prevalent due to the remote nature of the study and could impact data collection.

The study sample comprised a representative sample of 161 participants and included data from 145 variables. While this sample size provided insights, it may not fully capture the diversity of the population or account for all potential variables. Moreover, some data points were occasionally removed or added during the study and with the introduction of new amendments. This manipulation could potentially shift the results, affecting the overall conclusions. Despite the small sample size, the process of compiling and analysing the data proved to be time-consuming, suggesting that handling a larger sample size may not be feasible without additional resources. Furthermore, the limitations of REDCap became apparent during the study, as visualising outliers required the use of external programs rather than within the platform itself. Additionally, drawing trends from the data was only achievable when comparing data points side by side. This limitation highlights the need for improved data visualisation within the platform.

To address these limitations and improve future research methodologies, future studies should aim for a larger and more diverse participant pool for greater representativeness and statistical power. Furthermore, efforts should be made to enhance the functionality of REDCap by developing better tools within the platform to visualise trends and outliers

directly or explore alternative platforms that offer more comprehensive data visualisation capabilities.

6.2.3 Significance of the Study

This research contributes to the field by providing innovative and practical ways to leverage REDCap features for data management and analysis. Specifically, the study explores novel methodologies and considerations in quality control within the context of utilising REDCap, shedding light on best practices and strategies to enhance data integrity and reliability in research endeavours.

This research highlights REDCap's potential beyond basic usage. By applying the database life cycle to its development, this project demonstrates the importance of discerning platforms in clinical research. Moreover, the study distinguishes various types of errors, shedding light on the intricacies of data management. These findings contribute to expanding the understanding of REDCap's capabilities and offer valuable insights for its utilisation in diverse research settings.

6.4 Reflections on Research Process

a. COVID-19 Disruptions

The pandemic disrupted our research timeline significantly. Lockdowns, travel restrictions, and safety protocols halted data collection for an extended period. To mitigate this, our project milestones were adjusted and dynamic approaches to patient recruitment and data collection were implemented.

b. Investigator Rotation

Due to unforeseen circumstances, we had to rotate between different investigators at some sites. Each investigator brought unique perspectives and expertise, but continuity was challenging.

Significant insights were obtained during the course of this research endeavour, and several crucial lessons were learned. The following reflections highlight key observations:

a. Time-Consuming Nature of Research

Research is inherently time-consuming. From study design to data collection, analysis, and interpretation, each step demands meticulous attention and dedication. Acknowledging this reality helps manage expectations and plan effectively.

b. Protocol Implications

These research findings may have implications for researchers. By understanding the challenges faced during data collection and analysis, researchers can make informed decisions regarding research practices in creating standard operating procedures (SOP) and protocols and data management systems.

c. Foot Length discussion

During the research process, an unexpected issue arose concerning the recording of foot length data. Although continuously captured CRFs, this value was inadvertently omitted in the REDCap database. Upon investigation, it was revealed that investigators had misinterpreted the reading as neonate length at birth rather than foot length. Importantly, this error was not isolated to any specific site; it was observed across all sites involved in the study. Recognising the significance of neonate length for diagnostic and prognostic purposes, the NESHIE team engaged in discussions. With input from experts and investigators, a new variable was introduced on the CRFs and in REDCap specifically for neonate length.

6.5 Synthesis and Conclusions

This chapter provided an exploration of the complexities inherent in the research journey by highlighting both strengths and limitations. The following summarises the primary findings:

a. Usefulness and effectiveness of REDCap as an EDC system

The project aimed to establish a robust database for managing extensive data across multiple national sites and enhance data quality assessment processes. REDCap was chosen for its capacity to handle extensive data and built-in features for data quality evaluation, with the initial database serving as a foundation for further development. Design enhancements, such as piping, HTML, and matrices of fields facilitated data demarcation, grouping, and user interaction.

b. The necessity for database life cycle phases

The database life cycle phases provided a structured framework for managing database design, ensuring functionality, usability, and data integrity.

c. The analysis proved fruitful

Comparative analyses focused on existing and newly developed databases, identifying trends across amendments, sites, CRF subsections, and investigator-related aspects. Error analysis revealed persistent issues, leading to the establishment of error reports and continuous refinement of database design to reduce error rates. Statistical analysis demonstrated a significant reduction in error rates from initial iterations to the latest version, underscoring the effectiveness of design enhancements.

d. Outlier analysis for true error detection

Outlier detection highlighted potential errors originating from data entry and transcription discrepancies, emphasizing the importance of careful data interpretation and validation.

This research project yielded numerous findings; however, the important ones that stand out include:

a. Good design elements can impact data quality

Design elements significantly impact user experience and engagement in database activities, capturing the attention of end users and encouraging their participation. Meticulously crafted design features can ensure visual appeal, accessibility, and comprehensibility all while emphasising key data entry fields and minimising empty fields.

b. Errors provided opportunities for improvement

Errors encountered during data analysis are viewed as opportunities for learning and discovery, informing the integration of new features to enhance future databases. Continuous decrease in error rates over time suggests ongoing refinement of processes and enhanced understanding of data management principles by researchers. Some

errors created an open dialogues and collaboration within the team to facilitate understanding and promote standardisation of data capturing and management.

This research project has delved into various facets of database design, data management, and analysis within the context of clinical research. The use of REDCap as an EDC system in clinical research demonstrated its utility, robustness, and effectiveness across multiple national sites. Through meticulous design features and iterative improvements, the study has demonstrated the importance of user-friendly interfaces, rigorous data validation processes, and continuous refinement in enhancing overall data quality and integrity. Additionally, the project has shed light on common challenges and opportunities for improvement in data analysis, emphasising the importance of error identification and adherence to data quality standards.

6.6 References

1. Bangdiwala AS, Boulware DR. Technical procedures and redcap tools for internet-based clinical trials. Contemporary Clinical Trials. 2022; 114:106660. doi:https://doi.org/10.1016/j.cct.2021.106660

2. Mkabile S, Swartz L. 'I waited for it until forever': Community barriers to accessing intellectual disability services for children and their families in Cape Town, South Africa. International Journal of Environmental Research and Public Health. 2020; 17:8504. doi:10.3390/ijerph17228504

3. Hossmann S, Haynes AG, Spoerri A, Diatta ID, Aboubacar B, Egger M, et al. Data management of clinical trials during an outbreak of Ebola virus disease. Vaccine. 2019; 37(48):7183-9. doi:10.1016/j.vaccine.2017.09.094

4. Martin-Willett R, McCormick Z, Newman W, Larsen LD, Ortiz Torres MA, Bidwell LC. The transformation of a gold standard in-person substance use assessment to a web-based, redcap integrated data capture tool. J Biomed Inform. 2019; 94:103186. doi:10.1016/j.jbi.2019.103186

5. Hong MKH, Yao HHI, Pedersen JS, Peters JS, Costello AJ, Murphy DG, et al. Error rates in a clinical data repository: Lessons from the transition to electronic data transfer—a descriptive study. BMJ Open. 2013; 3(5):e002406. doi:10.1136/bmjopen-2012-002406

Appendix A: Ethical Approval Certificate

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

Institution: The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IORG #: IORG0001762 OMB No. 0990-0279 Approved for use through February 28, 2022 and Expires: 03/04/2023.

19 November 2020

Approval Certificate
New Application

Ethics Reference No.: 739/2020
Title: The development and design of a highly discerning platform for data capture in a neonatal encephalopathy study

Dear Ms TC Kalua

The **New Application** as supported by documents received between 2020-10-14 and 2020-11-18 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2020-11-18 as resolved by its quorate meeting.

Please note the following about your ethics approval:
- Ethics Approval is valid for 1 year and needs to be renewed annually by 2021-11-19.
- Please remember to use your protocol number (739/2020) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers
MBChB MMed (Int) MPharmMed PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)

Appendix B: A3 CRF

NESHIE study: Screening sheet (Address and phone number separate)

Surname……..

Infant Folder no…… DOB……Time OB……..

The neonate was: **Inborn / Outborn**

Mother Folder no……..Study number (mother)……

Birth site: Province…….....

Health Sub-district……

Name of institution……

Is the **Province** in which the mother gave birth the same as the province in which the mother permanently resides?

Yes / No

If no, in which **Province** does the mother reside permanently? ……..

Hospital Providing Cooling: ……

INCLUSION/EXCLUSION CRITERIA

Must satisfy all categories in A + B + C to qualify for cooling/recruitment (ring the answer)

A: (All must be Yes to be cooled and to be included in study)

≥36 weeks gestation* ***Y / N***
> 1800g ***Y / N***
Able to start cooling before age 6hours ***Y / N***

* As determined by foetal ultra sound at ≤ 20 weeks; or postnatal Ballard, or Foot length measurements (≥73mm in length)

B: (at least <u>one</u> must be Yes to be cooled and to be included in study)

BD 1st hr ≥ 16 / pH ≤ 7 ***Y / N***

BD 1st hr ≥10 / pH ≤ 7 **and** peripartum/ sentinel event ***Y / N***

5min Apgar < 7(only if early gas not available) ***Y / N***

Required resus/respiratory support at 10min ***Y / N***

C: Which <u>sign</u> of encephalopathy was present before cooling?
(At least one must be "yes" to qualify for cooling and to be included in the study)

lethargy ***Y / N*** stupor ***Y / N*** coma ***Y / N*** seizures ***Y / N*** *Thompson score ≥ 7* ***Y / N***

If clinical seizures present, describe……

Which other abnormal signs were present before cooling?
(If seizures are NOT present, at least one must be "yes" to be cooled or included in the study)
hypotonia ***Y / N*** abnormal reflexes ***Y / N*** abnormal suck ***Y / N***

Which amplitude-integrated EEG (aEEG) appearance was present before or at the onset of cooling?
(tick all that apply - must tick at least one to be cooled and included in the study)
Abnormal background voltage / pattern ***Y / N***
Seizures ***Y / N***

If the background voltage was abnormal, select which trace best describes the abnormality.
(tick one)
discontinuous normal voltage ☐
burst suppression, ☐
low voltage ☐
flat or iso-electric ☐

Exclusion criteria *(ring all applicable): none / refused consent / severe PPHN / Severe hypotension / bleeding that is not responding to treatment / Congenital abnormality / Known chromosomal abnormality / moribund and unlikely to benefit from cooling / asystole / congenital infection/ surgical anomaly/ encephalopathy primarily due to non-hypoxic cause(……..)*

Can it be confirmed that counselling was provided prior to seeking informed consent? Y / N
(Must be "YES" to be included in the study)

RECRUITED **Y / N** If No, why not?……

NESHIE study: History and delivery sheet

Fill in details or circle the applicable answer

MATERNAL DETAILS **RACE** BLACK / WHITE / COLOURED / INDIAN /OTHER……

FIRST LANGUAGE…… Age…….yr Gravidity…….. Parity…….

How many fetuses this pregnancy?……..

Did the mother receive antenatal care before delivery (was she "booked")? **Y / N**

RVD pos+arv / pos-no arv / neg / UK **VDRL** pos+fully treated / pos-not fully treated / neg / UK

Pre-existing Maternal Medical Conditions or treatment (not pregnancy complications) ***Y / N***

If Yes give detail……

PREGNANCY COMPLICATIONS ***Y / N*** If YES: Diabetes ***Y / N*** Illicit drug or alcohol abuse ***Y / N*** Maternal Seizure ***Y / N***

Placenta Praevia ***Y / N*** Pre-eclampsia ***Y / N*** Thyroid disorder ***Y / N***

Bleeding ***Y / N*** Infection ***Y / N*** Anaemia ***Y / N*** Pyrexial illness ***Y / N*** Smoker ***Y / N***

PROM > 18h ***Y / N*** Clinical Chorioamnionitis ***Y / N***

Other ***Y / N*** *detail*……

MODE OF DELIVERY Pre-labour CS / In-labour CS/ SVD cephalic / SVD breech / instrumental

DELIVERY COMPLICATIONS ***Y / N*** If YES: Head entrapment ***Y / N*** Placental abruption ***Y / N***

Prolapse Cord ***Y / N*** Ruptured Uterus ***Y / N*** Shoulder Dystocia ***Y / N***

Mec Stain Liquor ***Y / N*** Non-reassuring CTG or fetal Bradycardia ***Y / N / UK***

Prolong 2nd Stage ***Y / N / UK*** Other Ante Partum Haemorrhage ***Y / N*** Maternal Hypoxia ***Y / N***

Other Sentinal Event(s) ***Y / N*** detail…….

CONGENITAL ABNORMALITIES PRESENT AT BIRTH **Y / N**

If yes, detail……

PLACENTA SENT FOR HISTOLOGY/PATHOLOGY **Y / N** **Date sent: ……..**

If yes, brief results……

If not, why not……

SAMPLES FOR METABOLOMICS/BIOMARKER ISOLATION: Where the following samples collected?

Placenta: Microbiomics ***Y / N*** If no, explain why not:……(N/A)

Placenta: Metabolomics ***Y / N*** If no, explain why not:……(N/A)

Cord blood: Metabolomics ***Y / N*** If no, explain why not:……(N/A)

Neonatal urine ***Y / N*** If no, explain why not:……(N/A)

Blood spot (0-6 hrs) ***Y / N*** If no, explain why not:……

Blood spot (48-72 hrs) ***Y / N*** If no, explain why not:……

SAMPLES FOR NUCLEIC ACID ISOLATION: Where the following samples collected?

Cord blood:DNA/RNA ***Y / N***

If "yes", how was cord blood collected (Bag or Direct line?): ……

If "Bag" collection, what was the final volume of blood in bag: ……

Neonatal Peripheral blood (If "No: Cord Blood") ***Y / N*** If no, explain why not:…….. (N/A)

What is the total volume of neonate blood taken for DNA/RNA testing?.....mL

Maternal Peripheral blood ***Y / N*** If no, explain why not:……

Paternal Peripheral blood ***Y / N*** If no, explain why not:……

PLACENTA FOR METABOLOMICS: TIME COLLECTED…… DATE: …… NO. OF SAMPLES TAKEN……

PLACENTA FOR MICROBIOMICS: TIME COLLECTED…… DATE: …… NO. OF SAMPLES TAKEN……

CORD BLOOD FOR METABOLOMICS: TIME COLLECTED…….. DATE: …… VOLUME…….. ML

NEONATAL URINE FOR METABOLOMICS: TIME COLLECTED…….. DATE: …… VOLUME…….. ML

BLOOD SPOT 1 (0-6 HRS): TIME COLLECTED…….. DATE: …… NO SPOTS COLLECTED……..

BLOOD SPOT 2 (48-72 HRS): TIME COLLECTED…….. DATE: …… NO SPOTS COLLECTED……..

NEONATE BLOOD 1– DNA/RNA SCREEN: TIME COLLECTED…… DATE: …… VOLUME……..ML

AGE AT TIME OF SAMPLE COLLECTION FOR METABOLOMICS: (CALCULATE)

MOTHER (PLACENTA): …… HRS:MINS

NEONATE (CORD BLOOD):…… HRS:MINS NEONATE (URINE):…… HRS:MINS

NEONATE (BLOOD SPOT1):…… HRS:MINS NEONATE (BLOOD SPOT 2): …… HRS:MINS

NEONATE (PERIPHERAL):…… HRS:MINS

CLINICAL DETAILS OF BABY AT BIRTH Gestation at birth ……**completed wks**

Indicate which method(s) used for GA: Ballard / Footlength / Sonar at ≤ 20 weeks / Dates **/** EUS

Ballard score (if done)…….. Footlength measure (if done)……cm

Sex **M / F** Birth weight……..gm COH:……cm

Apgar Score 1min: …… 5min …… 10min ……. , BMV **Y/N,** Chest Compressions ***Y / N*** Adrenaline ***Y / N***

Intubated ***Y / N*** Delayed cord clamping/cord milking ***Y / N***

Early Blood gas results *(worst base excess within 60 minutes of birth including cord blood)* available? ***Y / N*** Cord / infant

Result: pH……Base Excess …… Lactate……..

NESHIE study: Pre-cooling assessment and cooling method

CLINICAL EXAMINATION PRIOR TO COOLING

CPAP Oxygen ***Y / N*** *Nasal Canula Oxygen* ***Y / N***

Mechanical ventilation ***Y / N*** *Hypotension or Inotropes* ***Y / N***

Active bleeding treated with blood prod. ***Y / N*** *(Hypotension= MBP persistently < 40mmHg) Subarachnoid haemorrhage (SAH)* ***Y / N*** *Sinus Bradycardia < 80* ***Y / N*** *Other Arrhythmia*

Y / N FiO₂ at time of blood gas below……. KPa

NEUROLOGICAL ASSESSMENT PRIOR TO COOLING (NOT EARLIER THAN 30 MINS)

AGE AT ASSESSMENT…….. *Visible Seizures* ***Y / N***

Thompson HIE score……(Age ……..) Modified Sarnat HIE Grade……

CFM FINDINGS (WORST GRADE PRIOR TO COOLING): AGE OF ASSESSMENT…….

Voltage: [Normal] **or** [Moderately abnorm] **or** [Suppressed]

Pattern: *FT / BS / CLV / DNV / CNV*

Electrical seizures: ***Y / N***

Model used: …… **How were measurement taken:** Needle / Skin patch

No. of channels used: ……

BASELINE LABORATORY INVESTIGATION: ON ADMISSION OR AS CLOSE TO COOLING AS POSSIBLE (WHICHEVER EARLIEST)

Date taken: …… Time at which taken: ……

Lowest blood glucose…….mmol/l Highest blood glucose…….mmol/l

POC: Na (lab)…….mmol/l K (lab)…….mmol/l

pH ……. PCO2……Kpa Bicarb ……mmol/l Lactate……mmol/l Base Excess……iCa(poc) ……mmol/l

Lab: Hb……g/dl NRBC……. Total WBC……Neut……Platelets……

Blood culture result for culture taken on this day ***positive / no growth***

Organism……

COOLING INDUCTION, METHOD AND SETTING

Date admitted to cooling centre…… Time admitted to cooling centre ……

Date *commenced* cooling …… Time cooling commenced……

Calculate Age (hrs:mins) : 1) Admitted to cooling centre……

2) *Commenced* cooling ……

Temperature before cooling started……°C

Tick cooling method used: *Automated whole body cooling (state make……)* ☐

Servo-controlled gel bag method ☐

Coolcap cooling ☐

Manual Techotherm ☐

Other (describe……..) ☐

Target core Temp of 33.5°C attained: Date:…….. Time:……..

Age when Target core Temp of 33.5 °C attained: ……hrs……min

Is the centre where the baby is cooled able to offer intubation/IPPV if needed in this baby? ***Y / N***

Is the centre where the baby is cooled able to offer invasive blood pressure monitoring? ***Y / N***

Are inotrope infusions available at the cooling centre ***Y / N***

What level of care is the baby nursed in? **NICU / High Care**

Was the patient moved to NICU at any point in time if care was initiated in High Care? ***Y / N***

What is the nurse: baby ratio (average)?……

DAILY GENERAL MONITORING DATA after cooling commencement

Do Thompson HIE score and Modified Sarnat HIE score at 6 hours and daily using last page of this booklet

DAY 1 (from start of cooling)

Time	Hours from start cool	Target Temp °C	Core Temp °C	Time	Hours from start cool	Target Temp °C	Core Temp °C	
	0				12			*Clinical **or** EEG Seizures* ***Y / N***
	1				13			*Sepsis (proven on culture this day)* ***Y / N***
	2				14			*Organism……*
	3				15			*CPAP Oxygen* ***Y / N***
	4				16			*FiO_2 Oxygen* ***Y / N***
	5				17			*Nasal Canula Oxygen* ***Y / N***
	6				18			*Mechanical ventilation* ***Y / N***
	7				19			*Hypotension or Inotropes* ***Y / N***
	8				20			*(Hypotension= MBP persistently < 40mmHg)*
	9				21			*Sinus Bradycardia < 80* ***Y / N***
	10				22			*Arrhythmia (other than SB < 80)* ***Y / N***
	11				23			*Bleeding / SAH* ***Y / N***
								Blood spots taken
								FiO_2 at time of bloods ……. KPa
								Investigations done at 24 h (+/- 6h)
								***Poc**: Ph…….CO2…….Kpa BE……*
								Bicarb…….iCa……..Lact…….
								Na……..mmol/l K……..mmol/l Hb…….
								***Lab:** Na……K ……. mmol/l*
								Ca……Mg ……..mmol/l
								Urea……mmol/l Creatinine…….micmol/l
								Lowest/Highest glucose……../……..

Sedation / anticonvulsants given during 1st 24 hours (ring ALL that apply):
None / Morphine / Phenobarb / Midazolam / Lignocaine / Phenytoin / Keppra /
Other ……
aEEG@age 6h Pattern: *FT / BS / CLV / DNV / CNV* **Volt:** *Normal / Modert / Suppressed*
aEEG@age 24h Pattern: *FT / BS / CLV / DNV / CNV* **Volt:** *Normal / Modert / Suppressed*

Fluid / Feed / TPN / HR
Fluid ***Y/N*** *Feed* ***Y/N*** *TPN* ***Y/N***
Fluid/kg…… *Feed/kg……* *TPN/kg……*
Min HR……..bpm Time taken (hr)…… *Max HR……..bpm Time taken (hr) ……*
*Any additional notes: **Y/N** (describe if "Yes")*
……
……

Cranial Ultrasound (D1) : ***Y/N*** If "Yes", Age taken:……*(Hrs)* *Observations:……*
……
……

Additional D1 comments/observations: ***Y/N*** *(describe if "Yes")*
……
……

DAY 2

Time	Hours from start cool	Target Temp °C	Core Temp °C	Time	Hours from start cool	Target Temp °C	Core Temp °C	
	24				36			*Clinical **or** EEG Seizures* ***Y / N***
	25				37			*Sepsis (proven on culture this day)* ***Y / N***
	26				38			*Organism……*
	27				39			*CPAP Oxygen* ***Y / N***
	28				40			*FiO_2 Oxygen* ***Y / N***
	29				41			*Nasal Canula Oxygen* ***Y / N***
	30				42			*Mechanical ventilation* ***Y / N***
	31				43			*Hypotension or Inotropes* ***Y / N***
	32				44			*(Hypotension= MBP persistently < 40mmHg)*
	33				45			*Sinus Bradycardia < 80* ***Y / N***
	34				46			*Arrhythmia (other than SB < 80)* ***Y / N***
	35				47			*Bleeding / SAH* ***Y / N***
								Blood spots taken ***Y / N***
								FiO_2 at time of bloods ……. KPa
								Investigations done at 36 - 48 hours
								***Poc**: Ph…….CO2…….Kpa BE……*
								Bicarb…….iCa……..Lact…….
								Na……..mmol/l K……..mmol/l Hb……
								Lowest/Highest glucose……./…….
								***Lab:** CRP:……mg/dl*

Sedation / anticonvulsants given during 2nd 24 hours (ring ALL that apply):
None / Morphine / Phenobarb / Midazolam / Lignocaine / Phenytoin / Keppra /
Other ……
aEEG@age 48h Pattern: *FT / BS / CLV / DNV / CNV* **Volt:** *Normal / Modert / Suppressed*

Fluid / Feed / TPN / HR
Fluid ***Y/N*** *Feed* ***Y/N*** *TPN* ***Y/N***
Fluid/kg…… *Feed/kg……* *TPN/kg……*
Min HR……..bpm Time taken …….. Max HR……..bpm Time taken ……..
*Any additional notes: **Y/N** (describe if "Yes")*
……
……

Additional D2 comments/observations: ***Y/N*** *(describe if "Yes")* ……..
……
……
……

DAY 3

Time	Hours from start cool	Target Temp °C	Core Temp °C	Time	Hours from start cool	Target Temp °C	Core Temp °C	*Clinical **or** EEG Seizures* ***Y / N***
	48				60			*Sepsis (proven on culture this day)* ***Y / N***
	49				61			*Organism…….*
	50				62			*CPAP Oxygen* ***Y / N***
	51				63			*FiO_2 Oxygen* ***Y / N***
	52				64			*FiO_2 at time of bloods ……. KPa*
	53				65			*Nasal Canula Oxygen* ***Y / N***
								Mechanical ventilation ***Y / N***
								Hypotension or Inotropes ***Y / N***
								(Hypotension= MBP persistently < 40mmHg)
								Sinus Bradycardia < 80 ***Y / N***
								Arrhythmia (other than SB < 80) ***Y / N***
								Bleeding / SAH ***Y / N***
								Blood spots taken ***Y / N***
	54				66			*Investigations done at 72h **(+/- 6h)***
	55				67			
	56				68			***Poc**: Ph…….CO2…….Kpa BE……..*
	57				69			*Bicarb…….iCa……..Lact…….mmol/l*
	58				70			*Na……..mmol/l K……..mmol/l Hb……*
	59				71			*Lowest/Highest glucose……../……*

Sedation / anticonvulsants given during Day 3 (ring ALL that apply):
None / Morphine / Phenobarb / Midazolam / Lignocaine / Phenytoin / Keppra / Other ……
aEEG@72h Pattern: *FT / BS / CLV / DNV / CNV* **Volt:** *Normal / Modert / SevSuppres*

Fluid / Feed / TPN / HR
Fluid ***Y/N*** *Feed* ***Y/N*** *TPN* ***Y/N***
Fluid/kg…… *Feed/kg……* *TPN/kg……*
Min HR……..bpm Time taken (hr)…… *Max HR……..bpm Time taken (hr) ……*
Any additional notes: ***Y/N*** *(describe if "Yes")*
……
……
……

Cranial Ultrasound (D3) – ***Y/N*** If "Yes", Age taken:……*(Hrs)* *Observations:……*
……
……

Additional D3 comments/observations: ***Y/N*** *(describe if "Yes")* ……
……
……

DAY 4

Time	Hours from start cool	Target Temp °C	Core Temp °C	Time	Hours from start cool	Target Temp °C	Core Temp °C	*Clinical **or** EEG Seizures* ***Y / N*** *Sepsis (proven on culture this day)* ***Y / N*** *Organism……* *CPAP Oxygen* ***Y / N*** *FiO_2 Oxygen* ***Y / N*** *FiO_2 at time of bloods ……. KPa Nasal Canula Oxygen* ***Y / N*** *Mechanical ventilation* ***Y / N*** *Hypotension or Inotropes* ***Y / N*** *(Hypotension= MBP persistently < 40mmHg)* *Sinus Bradycardia < 80* ***Y / N*** *Arrhythmia (other than SB < 80)* ***Y / N*** *Bleeding / SAH* ***Y / N***
	72				84			
	73				85			
	74				86			
	75				87			
	76				88			
	77				89			
	78				90			*Investigations done at 96h **(+/- 6h)*** *Lowest/Highest glucose……/……..*
	79				91			
	80				92			
	81				93			
	82				94			
	83				95			

Sedation / anticonvulsants given during Day 4 (ring ALL that apply):
None / Morphine / Phenobarb / Midazolam / Lignocaine / Phenytoin / Keppra /
Other ……
aEEG after rewarm Pattern: *FT / BS / CLV / DNV / CNV* **Volt:** *Normal / Modert / SevSuppres*

Fluid / Feed / TPN
Fluid ***Y/N*** *Feed* ***Y/N*** *TPN* ***Y/N***
Fluid/kg…… *Feed/kg……* *TPN/kg……*
Min HR……..bpm Time taken (hr)…… *Max HR……..bpm Time taken (hr) ……*
Any additional notes: ***Y/N*** *(describe if "Yes")*
……
……
……

Cranial Ultrasound (D4) – ***Y/N*** If "Yes", Age taken:……*(Hrs)* *Observations:……..*
……
……

Additional D4 comments/observations: ***Y/N*** *(describe if "Yes")* ……..
……
……

HOSPITAL COURSE, DISCHARGE AND FOLLOW-UP

DIAGNOSES DURING ENTIRE ADMISSION

Anaemia	**Y / N**	Meconium aspiration	**Y / N**
Active bleeding treated with blood prod.	**Y / N**	Nasal Canula Oxygen	**Y / N**
CPAP Oxygen	**Y / N**	Necrotising enterocolitis	**Y / N**
FiO2 Oxygen	**Y / N**	Other Arrhythmia	**Y / N**
Hypoglycaemia	**Y / N**	Pneumonia	**Y / N**
Hypokalaemia	**Y / N**	Pulmonary airleak	**Y / N**
Hyponatraemia	**Y / N**	Pulmonary Haemorrhage	**Y / N**
Hypotension or Inotropes	**Y / N**	Pulmonary Hypertension	**Y / N**
Late onset sepsis (> 72h)	**Y / N**	Renal Failure treated with dialysis	**Y / N**
Major cerebral anomaly	**Y / N**	Sinus Bradycardia < 80	**Y / N**
Mechanical ventilation	**Y / N**	Subarachnoid haemorrhage (SAH)	**Y / N**
None of these	**Y / N**		

OTHER DIAGNOSES OR DIAGNOSTIC INVESTIGATIONS Y / N DESCRIBE IF 'YES'……

……

……

WAS THERE ANY CONDITION OR EVENT LIKELY TO BE DUE TO COOLING TREATMENT OR REWARMING? Y / N

DESCRIBE IF "YES"

……

FULL SUCKING /CUP FEEDING ESTABLISHED BY DISCHARGE ? Y / N If Yes, age established (days)……

WAS COOLING STOPPED <72 HOURS? Y / N IF COOLING STOPPED < 72 HOURS, EXPLAIN WHY:……..

……

……

Discharged home **Y / N** Age ……..days

Transferred to other hospital **Y / N** Age ……..days **Name of hospital**……..

Died **Y / N** Age ……..days……hours

PM performed? **Y / N** **Result**……..

Fluids ml/kg D5: Total…….Enteral …….. D6: Total…….Enteral …….. D7: Total……Enteral …….

MRI scan performed or booked Y / N **Date booked……**

MRI Results……..

……

Head ultrasound: Was a head ultrasound performed on: **D5 Y / N** **D10 or D/T Y / N**

Head ultrasound: Was a head ultrasound performed on: **D5 Y / N** **D10 or D/T Y / N**

Age at D5 scan Days…… hours ……

D5 scan observations:**……**

……..

Age at final scan in Days…… Day 10 or Discharge/transfer scan observations:

……..

……..

Thompson HIE Score

	0	1	2	3	Pre-Cool < 6h	6h	24h	48h	72h	96h	D5	D6	D7	D8	D9	sD10
Date																
Limb Tone	Norm	HypEr-tonic	HypO-tonic	Flaccid												
LOC	Norm	Hyper-alert or staring	Lethargic or Obtunded	Coma or Stuporose												
Visible Fits	None	Infrequent < 3/day	Frequent > 2/day													
Posture	Norm/Other	Fisting and / or Cycling	Strong distil flexion	Decerebrate												
Moro	Norm	Partial	Absent													
Grasp	Norm	Poor	Absent													
Suck	Norm	Poor	Absent and/or bites													
Respiration	Norm	Hyper-ventilation	Transient apnoea	Apnoea *requiring* IPPV												
Fontanel	Norm	Full	Tense													
TOTAL																
Modified Sarnat HIE categories Mod or Sev or Mild or Norm																

Modified Sarnat HIE categories

Category		Moderate Encephalopathy	Severe Encephalopathy
Level of Consciousness		Lethargic	Obtunded/Stuporous/Coma
Spontaneous activity		Decreased spontaneous activity	No spontaneous activity
Muscle tone		Hypotonia	Flaccid (profound hypotonia)
Posture		Distil flexion or extensor posture	Decerebrate
Suck or Moro		Weak suck or partial moro	Absent
Autonomic system	*Pupils*	Miosis (fixed pinpoint)	Fixed dilated, slow, absent, unequal or deviated
	Heart Rate	Bradycardia (< 100 beats per min)	-
	Respiration	Periodic or shallow breathing	Apnoea req ippv

"Moderate" grade is defined by abnormalities in three or more categories – the majority of the abnormalities under moderate.
"Severe" grade is defined by abnormalities in three or more categories – the majority of the abnormalities under severe.
NOTE: if equal abnormalities under moderate and severe, the grade is defined by the level of consciousness.
"Mild" grade is defined by abnormalities in less than three categories, and/or the isolated presence of one or more of hypertonia, exaggerated reflexes, or hyperalertness.
"Normal" is defined by no abnormalities

MATERNAL AND OBSTETRIC DATA COLLECTION SHEET

NB: Complete all fields in the data sheet. Empty fields will be captured as 'unknown'.

Hospital_____ **Age at time of delivery** ____y

Maternal ethnicity*_____ **Paternal ethnicity***_____

Para (before this pregnancy)___ **Gravida** ___

Education: Tertiary completed: Y/N Grade (if no tertiary)___

Dwelling: House/flat/informal settlement/lodger from rural area outside metro/other

Describe if dwelling is "other" .....

No. previous miscarriages (<22w)___ **No. previous stillbirths**(≥22w)___

No. previous newborns admitted___ Birthweights and reasons.....

No. previous newborns with HIE/'Asphyxia'___ Long-term problems.....

Date of last menstrual period_____ sure/unsure **Booked for antenatal clinic** Yes / No

Date booked for antenatal clinic*_____ **Number of antenatal care visits attended**_____

*if more than one visit booked, please complete table 1 on last page of this data collection form

Gestational age at first ultrasound scan_____wk, date.....

Estimated date of delivery: by dates_____, by first ultrasound_____

Weight at booking (or postnatal weight - specify)......_____kg **Height**_____cm

HIV: P/N/U, last CD4 count_____ **RPR** P/N/U **Lowest antenatal Hb**_____ g/dL

Conditions during pregnancy before onset of labour (tick if applicable):

Gestational or chronic hypertension___ Diabetes___ Epilepsy ___ Thyroid disease___ TB___

Pre-eclampsia___ APH ___ PROM____ None ____

Other.....

Treated for infections during pregnancy (tick): Urinary___ Vaginal/vulval___ Other___ None___

Other.....

Medications (not supplements) used in pregnancy.....

Recreation/habits in pregnancy (tick all that apply):Alcohol___ Cigarettes___ Drugs___ Other___ None ___

Describe habits (if possible) …… …… ……

……

Onset of labour (admission or first noted while admitted): Date_____ Time_____

Nature of labour: spontaneous/induced/ caesarean section (not in labour)

Are there other notes on nature of labour? Yes / No **Describe (if yes):** ……

……

Method of induction (select all that apply)**:** catheter bulb/PGE2/misoprostol/oxytocin/AROM/other

Other ……

Fetal movements at onset of labour: normal/reduced/not noted

CTG at or <24 hours before onset of labour (attach): No CTG/Reassuring/non-reassuring/abnormal†

If no CTG: enter baseline FHR here____ / Unknown Decelerations noted: Y/N/Unknown

Time CTG started ___:___ (24-hour time) (Ensure start time legible on CTG trace)

In labour: yes/no/unknown… if yes, cervical dilatation at time of CTG_____cm/unknown

Reason for CTG done: Routine/main risk factor (specifiy):……..

Baseline FHR _____bpm **Baseline variability:** <5/5-25/>25

Fetal condition during labour:

In first 30 mins of CTG in labour: no. of contractions____ no. of decelerations____

Decelerations, first 30 minutes of CTG (select all that apply)**:** none / early / variable / severe variable / late / prolonged / unclassified

Highest temp in labour_____degrees

Highest maternal heart rate in last four hours of labour: <100/101-110/111-120/>120 bpm

Analgesia in labour: opiate/epidural/None/Other

Other……

Duration of oxytocin augmentation (if used)_____h (completed hours)

Duration of active phase of first stage (1st time found cervix ≥4 cm to 1st time 10cm)_____h____min

Duration of active phase of first stage (1st time found cervix ≥5 cm to 1st time 10cm)_____h___min

Duration of second stage_____min (from 1st time found cervix 10 cm to birth)

Duration of ruptured membranes_____h **Liquor**: clear/mec+/mec++/mec+++ **Offensive**: Yes/No

---

Last 2 hours of CTG in labour (attach): No CTG/Reassuring/non-reassuring/abnormal†

If no CTG: enter final FHR here____ / Unknown Decelerations noted: Y/N/Unknown

Time CTG started ____:____ (24-hour time) (Ensure start time legible on CTG trace)

Reason for CTG done: Routine/main risk factor (specifiy):……..

Final FHR_____ bpm **Final variability :** <5/5-25/>25

In last 30 mins of CTG in labour: no. of contractions_____ no. of decelerations____

Decelarations last 30 minutes of CTG (select all that apply)**:** none/early/variable/severe variable/late/prolonged/unclassified

Sentinel event: none/cord prolapse/placental abruption/uterine rupture/shoulder dystocia/difficult breech/maternal collapse/fetomaternal haemorrhage

Brief description……..

Umbilical cord abnormalities: true knot/cord around neck twice or more/single artery other……..

Placenta sent for histology___

Delivery: date_____, Time ____:____ (24-hour time) **Presentation**: cephalic/breech

Final mode of delivery: spontaneous vaginal/vacuum/forceps/caesarean section **Failed vacuum or forceps Yes / No**

Reason for vacuum or forceps……

Reason for caesarean section……

Difficulties with delivery……

---

*Ethnicity: Xhosa, Zulu, South Sotho, Swazi, Pedi, Tsonga, Venda, Ndebele, Khoisan, Nama, mixed South African African, non-South African African, Coloured, European, Indian, other (specify)
†Use page 39-47 of the NICE guidelines

Table 1: Antenatal clinic visit booking log

Visit Number:	Date booked for visit	Visit Number:	Date booked for visit
1		13	
2		14	
3		15	
4		16	
5		17	
6		18	
7		19	
8		20	
9		21	
10		22	
11		23	
12		24	

Study ID:

Appendix C: A4 CRF

NESHIE STUDY: SCREENING SHEET (To be kept separate from Neonatal and Maternal Case Report Forms)

NEONATAL SCREENING INFORMATION								
Hospital providing cooling								
Surname								
Infant hospital folder no								
DOB				**TOB**				
The neonate was:	Inborn		Outborn		**Sex**	M	F	Intersex

MATERNAL STUDY INFORMATION	
Maternal hospital folder no	
Maternal Study ID	

BIRTH SITE INFORMATION			
Province			
Health Sub-district			
Name of Institution			
Is the <u>Province</u> in which the mother gave birth the same as the province in which the mother permanently resides?		Yes	No
If no, in which <u>Province</u> does the mother reside permanently?			

Study ID:

NESHIE STUDY: INCLUSION/EXCLUSION CRITERIA

COMPLETE NESHIE SCREENING FORM if category A, B & C conditions are met.

FOR ENROLMENT INTO THE NESHIE STUDY: All inclusion criteria should be met with no exclusion criteria present

A: (All must be Yes to be cooled and to be included in study)		
≥ 36 weeks gestation*	**Yes**	**No**
≥ 1800g	**Yes**	**No**

* As determined by foetal ultra sound at ≤ 20 weeks; or postnatal Ballard, or Foot length measurements (≥73mm in length)

B: (at least <u>one</u> must be Yes to be cooled and to be included in study)		
BD 1st hr ≥ 16 / pH ≤ 7	**Yes**	**No**
BD 1st hr ≥ 10 / pH 1st hr ≤ 7.15 and peripartum/ sentinel event	**Yes**	N/A
5min Apgar < 7 (only if early gas not available)	**Yes**	N/A
Required resus/assisted ventilation at 10min	**Yes**	**No**

C1: Which <u>sign</u> of encephalopathy was present before cooling? (At least one must be "Yes" to qualify for enrolment into the study)			**C2: Which other abnormal signs were present before cooling** (If clinical seizures are **NOT** present, at least one of these must be "Yes" to be enrolled in the study)		
Lethargy	**Yes**	**No**	Hypotonia	**Yes**	**No**
Stupor	**Yes**	**No**	Abnormal reflexes	**Yes**	**No**
Coma	**Yes**	**No**	Abnormal suck	**Yes**	**No**
Thompson score ≥ 7	**Yes**	**No**			
Seizures	**Yes**	**No**	*If "**<u>yes</u>**" to **<u>clinical seizures</u>**, describe: .....*		
.....					

D: Exclusion criteria (Tick all that apply. The presence of any exclusion criteria disqualifies the patient from enrolment.)					
None		aEEG not performed		Normal aEEG (no aEEG seizures & normal background voltage / pattern)	
Neonate not cooled ≤ 6 hours of life		Severe PPHN		Hypotension or bleeding that is not responding to treatment	
Asystole		Suspected chromosomal abnormality		Moribund and unlikely to benefit from cooling	
Congenital infection		Surgical anomaly		Neonate died prior to obtaining consent	
Consent refused		Consent not obtained for other reason	State:		
Encephalopathy primarily due to non-hypoxic cause	State:				
Congenital abnormality	State:				
Neonate not cooled	State:				
Other	State:				

E: Can it be confirmed that counselling was provided prior to seeking informed consent? (Must be "YES" to be included in the study)	Yes	No
F: Was the patient recruited?	Yes	No
If "no", why not?		

Study ID:

NESHIE STUDY: PRE-COOLING ASSESSMENT AND COOLING METHOD

CLINICAL DETAILS OF BABY AT BIRTH *Check relevant block and provide details as requested and as applicable*									
Method used for GA	Dates		Ballard		Ballard score		**Gestation at birth**		wks
	EUS at ≤20 weeks		Footlength		Length (cm)		**Birth weight**		g
Apgar scores:	1 min		5 min		10 min		**COH**		cm

Adrenaline	Yes	No	**BMV**	Yes	No	**Chest compressions**	Yes	No
Delayed cord clamping / Cord milking	Yes	No	**Intubated**	Yes	No	**Early blood gas results available**[1]	Yes	No

[1] Worst base excess within 60 minutes of birth including cord blood

BLOOD GAS EVALUATION: AT BIRTH *Check relevant block and provide details as requested and as applicable*			**Done and available?**	Yes	No	Not done
Date taken		**Source: Blood gas**	Cord (venous)		Cord (arterial)	
Time taken			Cord (mixed source)		Infant (ABG)	

Lowest blood glucose		mmol/L	**Highest blood glucose**		mmol/L	**pH** (POC)		
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
PCO_2.....(POC)	kPa	mmHg	**PaO_2**.....	kPa	mmHg	**FiO_2**		%
Lab: Hb		g/dL	**Lab: NRBC**		%	**Lab: WBC**		cells/L
Lab: Neutrophils		cells/L	**Lab: Platelets**		cells/L			

Blood culture result for culture taken on this day	Positive	No growth	Not performed
Organism(s) if positive:			

CLINICAL EXAMINATION PRIOR TO COOLING *Check relevant block and provide details as requested and as applicable*								
CPAP	Yes	No	**Nasal Cannulae**	Yes	No	**Additional oxygen provided? (E.g. FiO_2>21%)**	Yes	No
Mechanical ventilation	Yes	No	**Hypotension or Inotropes**[1]	Yes	No	**Active bleeding treated with blood product**	Yes	No
Subaponeurotic haemorrhage	Yes	No	**Sinus bradycardia <80**	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No

[1] Hypotension = MBP persistently < 40mmHg

Study ID:

NEUROLOGICAL ASSESSMENT PRIOR TO COOLING (NOT EARLIER THAN 30 MINS) OR AT ONSET OF COOLING									
Date taken		**Time taken**		**Age at neurological assessment**[1]	hr	min			
Thompson HIE score		**Modified Sarnat HIE grade**[2]		**Visible seizures**	Yes	No			
CFM FINDINGS (WORST GRADE PRIOR TO COOLING OR AT ONSET OF COOLING)									
Age at assessment		hours	**Voltage:**	Normal		Moderately abnormal		Suppressed	
Pattern	FT		BS		CLV		DNV		CNV
						Electrical seizures	Yes	No	
Model used:						**No. of channels used**			
How were measurements taken?				Needle		Skin patch			
Initiation of aEEG	Prior to cooling		At onset of cooling		**Time aEEG initiated (24H format)**				
Indicate the CFM /aEEG monitoring status				Continuous		Intermittent			

[1] Same as Thomson score age, [2] worst grade prior to cooling,

BASELINE LABORATORY INVESTIGATION: ON ADMISSION OR AS CLOSE TO COOLING AS POSSIBLE *COMPLETE ONLY IF DIFFERENT FROM BLOOD GAS AT BIRTH Check relevant block and provide details as requested and as applicable.*								
Date taken			**Source: Blood gas**	Cord (venous)		Cord (arterial)		
Time taken				Cord (mixed source)		Infant (ABG)		
Lowest blood glucose		mmol/L	**Highest blood glucose**		mmol/L	**pH** (POC)		
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
PCO_2..................(POC)	kPa	mmHg	**PaO_2**...........................	kPa	mmHg	**FiO_2**		%
Lab: Hb		g/dL	**Lab: NRBC**		%	**Lab: WBC**		cells/L
Lab: Neutrophils		cells/L	**Lab: Platelets**		cells/L			
Blood culture result for culture taken on this day						Positive	No growth	Not performed
Organism(s) if positive:								

COOLING INDUCTION, METHOD AND SETTING							
Date *admitted* to cooling centre		**Time *admitted* to cooling centre**	hr	min			
Date cooling *commenced*		**Time cooling *commenced***	hr	min			
Cooling method used:	Automated whole body cooling		State make:				
	Servo-controlled gel bag method		Coolcap cooling		Manual Techotherm		Other
Describe if "Other"							
Temperature before cooling started	°C	**Target core temperature**	°C	**Intubation/IPPV available in cooling unit?**	Yes	No	
Target core Temp =attained: Date		**Target core Temp attained: Time**		**Invasive BP monitoring available in cooling unit?**	Yes	No	
What is the average number of babies per nurse?				**Inotrope infusions available in cooling unit?**	Yes	No	
What level of care is the baby nursed in?	NICU	High Care	**If "High Care" selected, was the patient moved to NICU at any point in time?**	N/A	Yes	No	

Study ID:

NESHIE STUDY: SAMPLE COLLECTION SHEET

Fill in details or check the applicable answer

PLACENTA SENT FOR HISTOLOGY/PATHOLOGY	Yes	Date sent:		No	
If yes, brief result: …… If no, why not: ……					
Placenta: Pathomicrobiome sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Placenta: Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Cord blood (Venous): DNA & RNA sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Cord blood (Venous): Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Blood spot: <60 min sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
[1] Blood spot: 1-6 hour sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Blood spot: 48-72 sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[1] Only applicable if blood spots cannot be obtained within the first hour of neonatal life

Neonatal urine: Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
[2] Peripheral blood: Neonatal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Peripheral blood: Maternal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Peripheral blood: Paternal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[2] Only applicable if neonatal venous cord blood cannot be obtained within the first hour of neonatal life

Study ID:

NESHIE STUDY: DAILY GENERAL MONITORING DATA *after cooling commencement*

COOLING PERIOD							DAY 1
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	0				**12**		
	1				**13**		
	2				**14**		
	3				**15**		
	4				**16**		
	5				**17**		
	6				**18**		
	7				**19**		
	8				**20**		
	9				**21**		
	10				**22**		
	11				**23**		

CLINICAL PRESENTATION DURING THE 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock

INVESTIGATIONS DONE AT 24H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
Lab: Sodium (Na)		mmol/L	**Lab: Potassium (K)**		mmol/L	**Lab: Calcium (Ca)**		mmol/L
Lab: Magnesium (Mg)		mmol/L	**Lab: Urea**		mmol/L	**Lab: Creatinine**		mmol/L

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 1ST 24 HOURS OF LIFE										
aEEG @ age 6H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Suppressed		
aEEG @ age 24H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Suppressed		

FLUID / FEED / TPN DURING 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D1)					
Was a CUS done on D1?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D1 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D1 comments/observations			
Additional D1 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

COOLING PERIOD							DAY 2
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	24				**36**		
	25				**37**		
	26				**38**		
	27				**39**		
	28				**40**		
	29				**41**		
	30				**42**		
	31				**43**		
	32				**44**		
	33				**45**		
	34				**46**		
	35				**47**		

CLINICAL PRESENTATION DURING THE 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG/EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time HR has been moved from Fluid/Feed/TPN/HR block, FiO_2 Oxygen replaced with "Additional oxygen provided (E.g. FiO2 > 21%)"

INVESTIGATIONS DONE AT 36 - 48H								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
Lab: CRP		mg/dL						

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 2ND DAY OF LIFE										
aEEG @ age 48H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal		Moderately abnormal		Suppressed				

FLUID / FEED / TPNDURING 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Have the D2 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D2 comments/observations			
Additional D2 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

COOLING PERIOD							DAY 3
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	48				**60**		
	49				**61**		
	50				**62**		
	51				**63**		
	52				**64**		
	53				**65**		
	54				**66**		
	55				**67**		
	56				**68**		
	57				**69**		
	58				**70**		
	59				**71**		

CLINICAL PRESENTATION DURING THE 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes**[1]	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken**[2]		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken**[2]		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time

INVESTIGATIONS DONE AT 72H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 3RD DAY OF LIFE										
aEEG @ age 72H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Severely Suppressed		

FLUID / FEED / TPN DURING 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D3)					
Was a CUS done on D3?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D3 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D3 comments/observations			
Additional D3 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

REWARMING PERIOD							DAY 4
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	72				**84**		
	73				**85**		
	74				**86**		
	75				**87**		
	76				**88**		
	77				**89**		
	78				**90**		
	79				**91**		
	80				**92**		
	81				**93**		
	82				**94**		
	83				**95**		

CLINICAL PRESENTATION DURING THE 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time

INVESTIGATIONS DONE AT 96H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**				

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 4TH DAY OF LIFE										
aEEG after rewarm	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Severely Suppressed	

FLUID / FEED / TPN DURING 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D4)					
Was a CUS done on D4?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D4 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D4 comments/observations			
Additional D4 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

NESHIE STUDY: HOSPITAL COURSE AND DISCHARGE / TRANSFER

DIAGNOSES & SUPPORTIVE MEASURES DURING ENTIRE ADMISSION								
Check relevant block and provide details as requested and as applicable. Refer to Annexure D.e for DEFINITIONS OF TERMS								
Anaemia	Yes	No	**Hypoglycaemia**	Yes	No	**Active bleeding treated with blood product**	Yes	No
Hyponatraemia	Yes	No	**Hypokalaemia**	Yes	No	**Renal failure treated with dialysis**	Yes	No
Pulmonary hypertension	Yes	No	**Pulmonary airleak**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Pneumonia	Yes	No	**Pulmonary haemorrhage**	Yes	No	**Mechanical ventilation**	Yes	No
Major cerebral anomaly	Yes	No	**Subaponeurotic haemorrhage**	Yes	No	**Nasal Cannulae**	Yes	No
Sinus bradycardia < 80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Additional oxygen provided (E.g. FiO_2>21%)**	Yes	No
Necrotising enterocolitis	Yes	No	**Late onset sepsis (> 72h)**	Yes	No	**CPAP**	Yes	No
Meconium aspiration	Yes	No	**Other**	Yes	No	**None of these**	Yes	

[1]Hypotension = MBP persistently <40mmHg

OTHER DIAGNOSES OR DIAGNOSTIC INVESTIGATIONS	Yes	No
Describe if "Other" is "Yes":.....		
.....		
.....		
.....		
.....		

WAS THERE ANY CONDITION OR EVENT LIKELY TO BE DUE TO COOLING TREATMENT OR REWARMING	Yes	No
If yes, detail		
.....		

WAS COOLING STOPPED <72 HOURS	Yes	No
If yes, detail why		
.....		

FLUIDS FOLLOWING COOLING AND REWARMINNG [1]												
Check relevant block and provide details as requested and as applicable												
Fluids: D5	Yes	No	N/A	**Fluids: D6**	Yes	No	N/A	**Fluids: D7**	Yes	No	N/A	
D5: Total				**D6: Total**				**D7: Total**				
D5: Enteral				**D6: Enteral**				**D7: Enteral**				

[1]Unit of measure = mL/kg/day

Study ID:

DISCHARGE / TRANSFER / DEATH STATUS							
Discharge home? (If "Yes", indicate Age in days on which this occurred)				Yes	No	N/A	days
Transferred to other hospital? (If "Yes", indicate Age in days on which this occurred)				Yes	No	N/A	days
Name of hospital to which neonate was transferred							
Death? (If "Yes", indicate Age in days on which this occurred)	Yes	No	N/A	days			hours
If "Yes" to "Death", was a PM performed?	Yes	No	N/A	Result:			
.....							

FULL SUCKING / CUP FEEDING ESTABLISHED BY DISCHARGE	Yes	No	N/A
If yes, age established?			days

CLINICAL DETAILS OF BABY AT DISCHARGE OR TRANSFER *Check relevant block and provide details as requested and as applicable*			
Weight at discharge / transfer	g	**Circumference of Head at discharge / transfer**	cm
Have the D5 – D10, Discharge / Transfer Thompson & Modified Sarnat scores been completed?	Yes	No	

MRI INFORMATION						
Check relevant block and provide details as requested and as applicable. ***MRI files to be stored onto NESHIE External Hard Drive***						
MRI scan performed?	Yes	No	Booked, not performed	N/A	**Date performed**	
MRI Results (if performed)						
.....						
.....						

WAS A CUS PERFORMED BETWEEN DAY 5 AND DAY 14, DISCHARGE OR TRANSFER?	Yes	No	N/A
If "Yes", indicate on which day(s) a CUS was performed			
Upload completed CUS report (Annexure D. d) for each applicable day / time period onto REDCap.			

Study ID:

NESHIE STUDY: THOMPSON HIE AND MODIFIED SARNAT SCORES

	0	1	2	3	Pre-Cool < 6h	6h	24h	48h	72h	96h	D5	D6	D7	D8	D9	D10	Discharge OR Transfer
				Date													
Modified Sarnat HIE categories Mod or Sev or Mild or Norm																	
Limb Tone	Norm	HypEr-tonic	HypO-tonic	Flaccid													
LOC	Norm	Hyper-alert or staring	Lethargic or Obtunded	Coma or Stuporose													
Visible Fits	None	Infrequent < 3/day	Frequent > 2/day														
Posture	Norm/ Other	Fisting and / or Cycling	Strong distil flexion	Decerebrate													
Moro	Norm	Partial	Absent														
Grasp	Norm	Poor	Absent														
Suck	Norm	Poor	Absent and/or bites														
Respiration	Norm	Hyper-ventilation	Transient apnoea	Apnoea *requiring* IPPV													
Fontanel	Norm	Full	Tense														
TOTAL																	

Modified Sarnat HIE categories

Category	Moderate Encephalopathy	Severe Encephalopathy
Level of Consciousness	Lethargic	Obtunded/Stuporous/Coma
Spontaneous activity	Decreased spontaneous activity	No spontaneous activity
Muscle tone	Hypotonia	Flaccid (profound hypotonia)
Posture	Distil flexion or extensor posture	Decerebrate
Suck or Moro	Weak suck or partial moro	Absent
Autonomic system *Pupils*	Miosis (fixed pinpoint)	Fixed dilated, slow, absent, unequal or deviated
Heart Rate	Bradycardia (< 100 beats per min)	-
Respiration	Periodic or shallow breathing	Apnoea req ippv

"Moderate" *grade is defined by abnormalities in three or more categories – the majority of the abnormalities under moderate.*
"Severe" *grade is defined by abnormalities in three or more categories – the majority of the abnormalities under severe.*
NOTE: *if equal abnormalities under moderate and severe, the grade is defined by the level of consciousness.*
"Mild" *grade is defined by abnormalities in less than three categories, and/or the isolated presence of one or more of hypertonia, exaggerated reflexes, or hyperalertness.*
"Normal" *is defined by no abnormalities*

FORM CHECKED BY ……

FORM COMPLETED BY ……

DATA CAPTURED BY ……

Study ID:

NESHIE STUDY: MATERNAL & OBSTETRIC DATA

NOTE: IF A TEST / EVALUATION WAS NOT DONE, PLEASE INDICATE THIS USING "ND". IF A TEST / EVALUATION WAS DONE BUT DATA / RESULTS ARE NOT KNOWN, PLEASE INDICATE THIS USING "UK"

MATERNAL & PATERNAL DETAILS

Check relevant block and provide details as requested and as applicable

Maternal Race	Black		Coloured		Indian		White		**Non-South African**	Yes	No
	Other		State "Other":.....								
Paternal Race	Black		Coloured		Indian		White		**Non-South African**	Yes	No
	Other		State "Other":.....								

MATERNAL DEMOGRAPHIC DETAILS

Check relevant block and provide details as requested and as applicable

First language				**Education: Tertiary completed?**		Yes	No	UK	**Highest grade (if no tertiary)**			
Dwelling	House		Flat		Informal settlement		Lodger from rural area outside of metro				Other / UK	
Describe "Dwelling" if "Other / UK"												

MATERNAL PREGNANCY DETAILS

Check relevant block and provide details as requested and as applicable

Confidence in date of last menstrual period	Sure		Unsure		**Did the mother receive antenatal care before deliver (was she "booked")?**	Yes [1]		No			
Date of last menstrual period					**Date of 1st ultrasound scan**						
Gestational age by dates		weeks			**Gestational age at 1st ultrasound scan**					weeks	
Estimated date of delivery by: Dates					**Estimated date of delivery by: 1st ultrasound**						
Was weight measured at 1st booking?	Yes	No	Unknown		***If "Yes"*** **Weight** kg					**Height** cm	

Hospital where birth took place							**Age at time of delivery**			years	
Gravidity		**Parity**		**How many foetuses this pregnancy?**			**No. previous miscarriages (<22w)**		**No. previous stillbirths (≥22w)**		
[1,2 & 3] Please complete Tables 1-3 as appropriate				**No. previous newborns admitted?** [2]			**No. previous newborns with NESHIE / 'Asphyxia'?** [3]				
Postnatal measurements:	Not done			**M/U arm circ** [4] cm			**Weight** kg			**Height** cm	

RVD / HIV status	Positive + ARV		Positive No ARV		Negative		Unknown	
Venereal disease status	Positive Fully treated		Positive Not fully treated		Negative		Unknown	
Venereal disease test(s) used	VDRL		RPR		TPHA			
If HIV positive, last CD4 count	cells/µL		CD4 unknown		**Lowest antenatal Hb**			g/dL

[1] **Please complete Table 1 if antenatal care was provided,**
[2] **Please complete Table 2 if previous newborns were admitted,**
[3] **Please complete Table 3 if there had previously been NESHIE / 'Asphyxia' cases**
[4] **Measurements taken on RIGHT arm**

Study ID:

COMPLETE THE INFORMATION IN TABLE 1 ONLY IF: ANTENATAL CARE WAS RECEIVED (MOTHER WAS "BOOKED")

COMPLETE THE INFORMATION IN TABLE 2 ONLY IF: PREVIOUS NEWBORNS HAD BEEN ADMITTED

COMPLETE THE INFORMATION IN TABLE 3 ONLY IF: THERE WERE PREVIOUS NESHIE / 'ASPHYXIA' CASES

Table 1: Antenatal clinic visit booking log

Visit Number:	Date booked for visit	Attended			Visit Number:	Date booked for visit	Attended		
1		Yes	No	Unknown	13		Yes	No	Unknown
2		Yes	No	Unknown	14		Yes	No	Unknown
3		Yes	No	Unknown	15		Yes	No	Unknown
4		Yes	No	Unknown	16		Yes	No	Unknown
5		Yes	No	Unknown	17		Yes	No	Unknown
6		Yes	No	Unknown	18		Yes	No	Unknown
7		Yes	No	Unknown	19		Yes	No	Unknown
8		Yes	No	Unknown	20		Yes	No	Unknown
9		Yes	No	Unknown	21		Yes	No	Unknown
10		Yes	No	Unknown	22		Yes	No	Unknown
11		Yes	No	Unknown	23		Yes	No	Unknown
12		Yes	No	Unknown	24		Yes	No	Unknown

Table 2: Birthweight and reason for previous admissions

Admission number	Birthweight (grams)	Reason for admission	Did the newborn demise?		
1			Yes	No	Unknown
2			Yes	No	Unknown
3			Yes	No	Unknown
4			Yes	No	Unknown
5			Yes	No	Unknown

Table 3: Long-term problems for previous NESHIE / 'Asphyxia' newborns

Admission number	Long-term problem(s)	Did the newborn demise?		
1		Yes	No	Unknown
2		Yes	No	Unknown
3		Yes	No	Unknown
4		Yes	No	Unknown
5		Yes	No	Unknown

Study ID:

MATERNAL MEDICAL CONDITIONS OR TREATMENT PRESENT PRIOR TO PREGNANCY? (NOT PREGNANCY COMPLICATIONS)									Yes	No	UK
Check relevant block and provide details as requested and as applicable below if "Pregnancy complications" is "Yes"											
Anaemia	Yes	No	UK	**Cardiac disease**	Yes	No	UK	**Diabetes**	Yes	No	UK
Epilepsy / Seizure	Yes	No	UK	**Hypertension**	Yes	No	UK	**Thyroid disease: Hypothyroidism**	Yes	No	UK
Thyroid disease: Hyperthyroidism	Yes	No	UK	**Other**	Yes	No		**Detail if "Other" is "Yes"**			
.....											
.....											
Treatment for pre-existing conditions?	Yes	No	UK	**Detail if "Yes"**							
.....											
.....											

COMPLICATIONS DURING PREGNANCY BEFORE ONSET OF LABOUR / PREGNANCY COMPLICATIONS?									Yes	No	UK
Check relevant block and provide details as requested and as applicable below if "Pregnancy complications" is "Yes"											
Anaemia (gestational)	Yes	No	UK	**APH / Bleeding**	Yes	No	UK	**Clinical Chorioamnionitis**	Yes	No	UK
Diabetes (gestational)	Yes	No	UK	**Hypertension (gestational)**	Yes	No	UK	**Intra-uterine growth restriction**	Yes	No	UK
Placenta Praevia	Yes	No	UK	**Pre-eclampsia / eclampsia / HELLP**	Yes	No	UK	**PROM > 18h**	Yes	No	UK
Pyrexia	Yes	No	UK	**TB (Active) during pregnancy**	Yes	No	UK	**On treatment if "TB"= "Yes"?**	Yes	No	UK
Other	Yes	No		**Detail if "Other" is "Yes"**							
.....											
.....											
Infection during pregnancy?	Yes	No	UK	**Selection option only if "Infection" is "Yes"**	Urinary		Vaginal/Vulval		Other		
Detail if "Infection" is "Other"											

RECREATION / HABITS DURING PREGNANCY?				Yes	No	UK
Check relevant block and provide details as requested and as applicable if "Yes" to any recreational habits						
Alcohol	Yes	No	UK	Description:		
Cigarettes / Smoker	Yes	No	UK	Description:		
Illicit drugs	Yes	No	UK	Description:		
Other	Yes	No	UK	Description:		

MEDICATIONS (NOT SUPPLEMENTS) USED IN PREGNANCY?	Yes	No	UK
Describe in full if "Yes"			
.....			

Study ID:

CTG DATA BEFORE ONSET OF LABOUR CONTRACTIONS (<24 HOURS BEFORE ONSET OF LABOUR)						
Scan last 30 minutes of CTG before onset of labour if available						
CTG <24 hours before onset of labour	Available		Not done		Done, not available	
Reason for CTG done (main risk factor):						
.....						
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE / DONE <24 HOURS BEFORE ONSET OF LABOUR						
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

ONSET OF LABOUR						
Check relevant block and provide details as requested and as applicable						
Onset of labour: Admission or first noted while admitted	**Date**		**Time**			
Nature of labour	Spontaneous		Induced		Caesarean Section (not in labour)	
Are there other notes on the nature of labour?	Yes	No	Describe if "Yes"			
.....						
Method of induction (Check all that are applicable)	Catheter bulb		PGE2		Misoprostol	
	Oxytocin		AROM		Other	
Describe if "induction" is "Other"						
Fetal movements at onset of labour	Normal		Reduced		Not noted	

FETAL CONDITION AFTER ONSET OF LABOUR CONTRACTIONS						
Scan all available CTG trace after onset of labour						
CTG in labour earlier than the last 2 hours before delivery	Available		Not done		Done, not available	
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE / DONE AFTER ONSET OF LABOUR						
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

MATERNAL CONDITION AFTER ONSET OF LABOUR CONTRACTIONS								
Highest maternal heart rate in last four hours of labour	<100 bpm		101-110 bpm		111-120 bpm		> 120 bpm	
Analgesia in labour (Check all that are applicable)	None		Opiate		Epidural		Other	
Describe if "Analgesia" is "Other"								
Duration of oxytocin augmentation (if used; completed hours)	N/A	hours	**Highest temperature in labour**				°C	
Duration of active phase of first stage	1st time found cervix ≥ **4 cm** to 1st time 10 cm		h min		1st time found cervix ≥ **5 cm** to 1st time 10 cm		h min	
Duration of second stage	1st time found cervix 10 cm to birth		min		**Duration of ruptured membranes**		hours	
Liquor	Clear		Mec+		Mec++		Mec+++		**Offensive liquor**	Yes	No

Study ID:

LAST 2 HOURS OF FHR/CTG IN LABOUR						
Scan all CTG tracing in the last 2 hours before delivery						
CTG from last 2 hours before delivery	Available		Not done		Done, not available	

COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE FOR >30 MINS AND DONE IN LAST 2 HOURS OF LABOUR							
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

DELIVERY COMPLICATIONS / SENTINEL EVENTS?									Yes	No	UK
Complete block below only if "Delivery Complications" is "Yes"											
Fetomaternal haemorrhage	Yes	No	UK	**Other Ante Partum Haemorrhage**	Yes	No	UK	**Placental abruption**	Yes	No	UK
Maternal Hypoxia	Yes	No	UK	**Maternal collapse**	Yes	No	UK	**Sudden onset bradycardia**	Yes	No	UK
Prolapse Cord	Yes	No	UK	**Ruptured Uterus**	Yes	No	UK	**Prolonged 2nd Stage (>2 hours)**	Yes	No	UK
Shoulder Dystocia	Yes	No	UK	**Difficult breech**	Yes	No	UK	**Other Sentinel Event(s)**	Yes	No	UK
Detail if "Other Sentinel events" is "YES":											

UMBILICAL CORD ABNORMALITIES *(Check relevant block)*	None		True knot		Cord around neck twice or more		Single artery	
	Other		Abnormal insertion		Hypercoiling		Unknown	
	Detail "Other" or "Abnormal insertion".....							

PLACENTAL ABNORMALITIES *(Check relevant block)*	None		Abnormal		Unknown		
	Detail if "Abnormal"						

ROUTE OF DELIVERY *(Check relevant block)*	Pre-labour CS		In-labour CS		Vaginal	
PRESENTATION	Cephalic		Breech			
MODE OF VAGINAL DELIVERY *(Check relevant block)*	Spontaneous		Vacuum		**FAILED VACUUM OR FORCEPS?**	Yes
	Caesarean section		Forceps			No
Reason for vacuum or forceps						
Reason for caesarean section						

DIFFICULTIES WITH DELIVERY	Yes	No	UK
If yes, detail			

Study ID:

Appendix D: A5 CRF

NESHIE STUDY: SCREENING SHEET

NEONATAL SCREENING INFORMATION								
Hospital providing cooling								
DOB				**TOB**				
The neonate was:	Inborn		Outborn		**Sex**	M	F	Intersex

BIRTH SITE INFORMATION			
Province			
Health Sub-district			
Name of Institution			
Is the <u>Province</u> in which the mother gave birth the same as the province in which the mother permanently resides?		Yes	No
If no, in which <u>Province</u> does the mother reside permanently?			

Study ID:

NESHIE STUDY: INCLUSION/EXCLUSION CRITERIA

COMPLETE NESHIE SCREENING FORM if category A, B & C conditions are met.

FOR ENROLMENT INTO THE NESHIE STUDY: All inclusion criteria should be met with no exclusion criteria present

A: (All must be Yes to be cooled and to be included in study)		
≥ 36 weeks gestation*	**Yes**	**No**
≥ 1800g	**Yes**	**No**

As determined by foetal ultra sound at ≤ 20 weeks; or postnatal Ballard, or Foot length measurements (≥73mm in length)

B: (at least <u>one</u> must be Yes to be cooled and to be included in study)			
BD 1st hr ≥ 16 / pH ≤ 7	**Yes**	**No**	N/A
BD 1st hr ≥ 10 / pH 1st hr ≤ 7.15 and peripartum/ sentinel event	**Yes**	**No**	N/A
5min Apgar < 7 (only if early gas not available)	**Yes**	**No**	
Required resus/assisted ventilation at 10min	**Yes**	**No**	

C1: Which <u>sign</u> of encephalopathy was present before cooling? (At least one must be "Yes" to qualify for enrolment into the study)			**C2: Which other abnormal signs were present before cooling** (If clinical seizures are **NOT** present, at least one of these must be "Yes" to be enrolled in the study)		
Lethargy	**Yes**	**No**	Hypotonia	**Yes**	**No**
Stupor	**Yes**	**No**	Abnormal reflexes	**Yes**	**No**
Coma	**Yes**	**No**	Abnormal suck	**Yes**	**No**
Thompson score ≥ 7	**Yes**	**No**			
Seizures	**Yes**	**No**	*If **<u>"yes"</u>** to **<u>seizures</u>**, describe: .....*		
.....					

D: Exclusion criteria (Tick all that apply. The presence of any exclusion criteria disqualifies the patient from enrolment.)					
None		aEEG not performed		Neonate not cooled ≤ 6 hours of life	
Moribund and unlikely to benefit from cooling		Asystole		Hypotension or bleeding that is not responding to treatment	
Severe PPHN		Suspected chromosomal abnormality		Congenital infection	
Surgical anomaly		Consent refused		Neonate died prior to obtaining consent	
Consent not obtained for other reason	State:				
Encephalopathy primarily due to non-hypoxic cause	State:				
Congenital abnormality	State:				
Neonate not cooled	State:				
Other	State:				

E: Can it be confirmed that counselling was provided prior to seeking informed consent? (Must be "YES" to be included in the study)	Yes	No
F: Was the patient recruited?	Yes	No
If "no", why not?		

Study ID:

NESHIE STUDY: PRE-COOLING ASSESSMENT AND COOLING METHOD

CLINICAL DETAILS OF BABY AT BIRTH *Check relevant block and provide details as requested and as applicable*									
Method used for GA	Dates		Ballard		Ballard score		**Gestation at birth**		wks
	EUS at ≤20 weeks		Footlength		Length (cm)		**Birth weight**		g
Apgar scores:	1 min		5 min		10 min		**COH**		cm

Adrenaline	Yes	No	**BMV**	Yes	No	**Chest compressions**	Yes	No
Delayed cord clamping / Cord milking	Yes	No	**Intubated**	Yes	No	**Early blood gas results available**[1]	Yes	No

[1] Worst base excess within 60 minutes of birth including cord blood

BLOOD GAS EVALUATION: AT BIRTH *Check relevant block and provide details as requested and as applicable*			**Done and available?**	Yes	No	Not done
Date taken		**Source: Blood gas**	Cord (venous)		Cord (arterial)	
Time taken			Cord (mixed source)		Infant (ABG)	

Lowest blood glucose		mmol/L	**Highest blood glucose**		mmol/L	**pH** (POC)		
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
PCO_2.....(POC)	kPa	mmHg	**PaO_2**.....	kPa	mmHg	**FiO_2**		%
Lab: Hb		g/dL	**Lab: NRBC**		%	**Lab: WBC**		cells/L
Lab: Neutrophils		cells/L	**Lab: Platelets**		cells/L			

Blood culture result for culture taken on this day	Positive	No growth	Not performed
Organism(s) if positive:			

CLINICAL EXAMINATION PRIOR TO COOLING								
Check relevant block and provide details as requested and as applicable								
CPAP	Yes	No	**Nasal Cannulae**	Yes	No	**Additional oxygen provided? (E.g. FiO_2>21%)**	Yes	No
Mechanical ventilation	Yes	No	**Hypotension or Inotropes**[1]	Yes	No	**Active bleeding treated with blood product**	Yes	No
Subaponeurotic haemorrhage	Yes	No	**Sinus bradycardia <80**	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No

[1] Hypotension = MBP persistently < 40mmHg

Study ID:

NEUROLOGICAL ASSESSMENT PRIOR TO COOLING (NOT EARLIER THAN 30 MINS) OR AT ONSET OF COOLING						
Date taken		**Time taken**		**Age at neurological assessment[1]**	hr	min
Thompson HIE score		**Modified Sarnat HIE grade[2]**		**Visible seizures**	Yes	No

[1] Same as Thomson score age, [2] worst grade prior to cooling

CFM DETAILS						
Model used:					**No. of channels used**	
How were measurements taken?			Needle		Skin patch	
Indicate the CFM /aEEG monitoring status			Continuous		Intermittent	
Initiation of aEEG	Prior to cooling		At onset of cooling		Other	
Describe if aEEG initiation is "Other":						

BASELINE aEGG ASSESSMENT												
							Time of assessment		:			
Is the baseline aEEG assessment different to the aEEG evaluation at 6H							Yes		No			
Electrical seizures present?							Yes		No			
Assessment status				Normal			Abnormal		Not done			
Voltage:				Normal			Moderately abnormal		Suppressed			
Pattern:	Normal		FT		BS		CLV		DNV		CNV	

aEGG ASSESSMENT: 6 HOURS OF LIFE* **Complete this table only if 6H aEEG is different from baseline assessment*							**Time of assessment**		:			
Electrical seizures present?							Yes		No			
Assessment status				Normal			Abnormal		Not done			
Voltage:				Normal			Moderately abnormal		Suppressed			
Pattern:	Normal		FT		BS		CLV		DNV		CNV	

Study ID:

BASELINE LABORATORY INVESTIGATION: ON ADMISSION OR AS CLOSE TO COOLING AS POSSIBLE *COMPLETE ONLY IF DIFFERENT FROM BLOOD GAS AT BIRTH Check relevant block and provide details as requested and as applicable.*								
Date taken		**Source: Blood gas**	Cord (venous)		Cord (arterial)			
Time taken			Cord (mixed source)		Infant (ABG)			

Lowest blood glucose		mmol/L	**Highest blood glucose**		mmol/L	**pH** (POC)		
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
PCO_2.....(POC)	kPa	mmHg	**PaO_2**.....	kPa	mmHg	**FiO_2**		%
Lab: Hb		g/dL	**Lab: NRBC**		%	**Lab: WBC**		cells/L
Lab: Neutrophils		cells/L	**Lab: Platelets**		cells/L			

Blood culture result for culture taken on this day	Positive	No growth	Not performed
Organism(s) if positive:			

COOLING INDUCTION, METHOD AND SETTING										
Date *admitted* to cooling centre			**Time *admitted* to cooling centre**			hr	min			
Date cooling *commenced*			**Time cooling *commenced***			hr	min			
Cooling method used:	Automated whole body cooling		State make:							
	Servo-controlled gel bag method		Coolcap cooling		Manual Techotherm		Other			
Describe if "Other"										
Temperature before cooling started		°C	**Target core temperature**		°C	**Intubation/IPPV available in cooling unit?**		Yes	No	
Target core Temp =attained: Date			**Target core Temp attained: Time**			**Invasive BP monitoring available in cooling unit?**		Yes	No	
What is the average number of babies per nurse?						**Inotrope infusions available in cooling unit?**		Yes	No	
What level of care is the baby nursed in?	NICU	High Care	**If "High Care" selected, was the patient moved to NICU at any point in time?**			N/A		Yes	No	

Study ID:

NESHIE STUDY: SAMPLE COLLECTION SHEET

Fill in details or check the applicable answer

PLACENTA SENT FOR HISTOLOGY/PATHOLOGY	Yes	Date sent:	No
If yes, brief result:			
If no, why not:			

Placenta: Pathomicrobiome sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Placenta: Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

Cord blood (Venous): DNA & RNA sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Cord blood (Venous): Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

Blood spot: <60 min sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
[1] Blood spot: 1-6 hour sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Blood spot: 48-72 sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[1] Only applicable if blood spots cannot be obtained within the first hour of neonatal life

Neonatal urine: Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[2] Peripheral blood: Neonatal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Peripheral blood: Maternal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Peripheral blood: Paternal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[2] Only applicable if neonatal venous cord blood cannot be obtained within the first hour of neonatal life

Study ID:

NESHIE STUDY: DAILY GENERAL MONITORING DATA *after cooling commencement*

COOLING PERIOD							DAY 1
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	0				**12**		
	1				**13**		
	2				**14**		
	3				**15**		
	4				**16**		
	5				**17**		
	6				**18**		
	7				**19**		
	8				**20**		
	9				**21**		
	10				**22**		
	11				**23**		

CLINICAL PRESENTATION DURING THE 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock

INVESTIGATIONS DONE AT 24H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
Lab: Sodium (Na)		mmol/L	**Lab: Potassium (K)**		mmol/L	**Lab: Calcium (Ca)**		mmol/L
Lab: Magnesium (Mg)		mmol/L	**Lab: Urea**		mmol/L	**Lab: Creatinine**		mmol/L

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 1ST 24 HOURS OF LIFE										
aEEG @ age 24H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Suppressed		

FLUID / FEED / TPN DURING 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D1)					
Was a CUS done on D1?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D1 Thompson & Modified Sarnat scores been completed?	Yes	No
Additional D1 comments/observations		

Additional D1 comments/observations?	Yes	No	Observations: .....
.....			
.....			
.....			
.....			

Study ID:

COOLING PERIOD							DAY 2
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	24				**36**		
	25				**37**		
	26				**38**		
	27				**39**		
	28				**40**		
	29				**41**		
	30				**42**		
	31				**43**		
	32				**44**		
	33				**45**		
	34				**46**		
	35				**47**		

CLINICAL PRESENTATION DURING THE 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG/EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time HR has been moved from Fluid/Feed/TPN/HR block, FiO_2 Oxygen replaced with "Additional oxygen provided (E.g. FiO2 > 21%)"

INVESTIGATIONS DONE AT 36 - 48H								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
Lab: CRP		mg/dL						

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 2ND DAY OF LIFE										
aEEG @ age 48H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Suppressed		

FLUID / FEED / TPNDURING 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

Have the D2 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D2 comments/observations			
Additional D2 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

[1]Unit of measure = mL/kg/day

Study ID:

COOLING PERIOD							DAY 3
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	48				**60**		
	49				**61**		
	50				**62**		
	51				**63**		
	52				**64**		
	53				**65**		
	54				**66**		
	55				**67**		
	56				**68**		
	57				**69**		
	58				**70**		
	59				**71**		

CLINICAL PRESENTATION DURING THE 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg

[2]According to 24H clock time

INVESTIGATIONS DONE AT 72H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 3RD DAY OF LIFE										
aEEG @ age 72H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Severely Suppressed		

FLUID / FEED / TPN DURING 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D3)					
Was a CUS done on D3?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D3 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D3 comments/observations			
Additional D3 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

REWARMING PERIOD							DAY 4
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	72				**84**		
	73				**85**		
	74				**86**		
	75				**87**		
	76				**88**		
	77				**89**		
	78				**90**		
	79				**91**		
	80				**92**		
	81				**93**		
	82				**94**		
	83				**95**		

CLINICAL PRESENTATION DURING THE 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time

INVESTIGATIONS DONE AT 96H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**				

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 4TH DAY OF LIFE										
aEEG after rewarm	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Severely Suppressed		

FLUID / FEED / TPN DURING 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D4)					
Was a CUS done on D4?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D4 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D4 comments/observations			
Additional D4 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

NESHIE STUDY: HOSPITAL COURSE AND DISCHARGE / TRANSFER

DIAGNOSES & SUPPORTIVE MEASURES DURING ENTIRE ADMISSION								
Check relevant block and provide details as requested and as applicable. Refer to Annexure D.e for DEFINITIONS OF TERMS								
Anaemia	Yes	No	**Hypoglycaemia**	Yes	No	**Active bleeding treated with blood product**	Yes	No
Hyponatraemia	Yes	No	**Hypokalaemia**	Yes	No	**Renal failure treated with dialysis**	Yes	No
Pulmonary hypertension	Yes	No	**Pulmonary airleak**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Pneumonia	Yes	No	**Pulmonary haemorrhage**	Yes	No	**Mechanical ventilation**	Yes	No
Major cerebral anomaly	Yes	No	**Subaponeurotic haemorrhage**	Yes	No	**Nasal Cannulae**	Yes	No
Sinus bradycardia < 80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Additional oxygen provided (E.g. FiO_2>21%)**	Yes	No
Necrotising enterocolitis	Yes	No	**Late onset sepsis (> 72h)**	Yes	No	**CPAP**	Yes	No
Meconium aspiration	Yes	No	**Other**	Yes	No	**None of these**	Yes	

[1]Hypotension = MBP persistently <40mmHg

OTHER DIAGNOSES OR DIAGNOSTIC INVESTIGATIONS	Yes	No
Describe if "Other" is "Yes":.....		
.....		
.....		
.....		
.....		

WAS THERE ANY CONDITION OR EVENT LIKELY TO BE DUE TO COOLING TREATMENT OR REWARMING	Yes	No
If yes, detail		
.....		

WAS COOLING STOPPED <72 HOURS	Yes	No
If yes, detail why		
.....		

FLUIDS FOLLOWING COOLING AND REWARMINNG [1]											
Check relevant block and provide details as requested and as applicable											
Fluids: D5	Yes	No	N/A	**Fluids: D6**	Yes	No	N/A	**Fluids: D7**	Yes	No	N/A
D5: Total				**D6: Total**				**D7: Total**			
D5: Enteral				**D6: Enteral**				**D7: Enteral**			

[1]Unit of measure = mL/kg/day

Study ID:

DISCHARGE / TRANSFER / DEATH STATUS							
Discharge home? (If "Yes", indicate Age in days on which this occurred)				Yes	No	N/A	days
Transferred to other hospital? (If "Yes", indicate Age in days on which this occurred)				Yes	No	N/A	days
Name of hospital to which neonate was transferred							
Death? (If "Yes", indicate Age in days on which this occurred)	Yes	No	N/A	days			hours
If "Yes" to "Death", was a PM performed?	Yes	No	N/A	Result:			
.....							

FULL SUCKING / CUP FEEDING ESTABLISHED BY DISCHARGE	Yes	No	N/A
If yes, age established?			days

CLINICAL DETAILS OF BABY AT DISCHARGE OR TRANSFER

MRI INFORMATION

~~Weight at discharge / transfer~~ g ~~Circumference of Head at discharge / transfer~~ cm

Check relevant block and provide details as requested and as applicable. ***MRI files to be stored onto NESHIE External Hard Drive***

~~Have the D5 – D10, Discharge / Transfer Thompson & Modified Sarnat scores been completed?~~							Yes	No
MRI scan performed?	Yes	No	Booked, not performed	N/A	**Date performed**			

MRI Results (if performed) .....

.....

.....

WAS A CUS PERFORMED BETWEEN DAY 5 AND DAY 14, DISCHARGE OR TRANSFER?	Yes	No	N/A
If "Yes", indicate on which day(s) a CUS was performed			
Upload completed CUS report (Annexure D. d) for each applicable day / time period onto REDCap.			

Study ID:

NESHIE STUDY: THOMPSON HIE AND MODIFIED SARNAT SCORES

	0	1	2	3	Pre-Cool < 6h	6h	24h	48h	72h	96h	D5	D6	D7	D8	D9	D10	Discharge OR Transfer
				Date													
Modified Sarnat HIE categories Mod or Sev or Mild or Norm																	
Limb Tone	Norm	HypEr-tonic	HypO-tonic	Flaccid													
LOC	Norm	Hyper-alert or staring	Lethargic or Obtunded	Coma or Stuporose													
Visible Fits	None	Infrequent < 3/day	Frequent > 2/day														
Posture	Norm/ Other	Fisting and / or Cycling	Strong distil flexion	Decerebrate													
Moro	Norm	Partial	Absent														
Grasp	Norm	Poor	Absent														
Suck	Norm	Poor	Absent and/or bites														
Respiration	Norm	Hyper-ventilation	Transient apnoea	Apnoea *requiring* IPPV													
Fontanel	Norm	Full	Tense														
TOTAL																	

Modified Sarnat HIE categories

Category	Moderate Encephalopathy	Severe Encephalopathy
Level of Consciousness	Lethargic	Obtunded/Stuporous/Coma
Spontaneous activity	Decreased spontaneous activity	No spontaneous activity
Muscle tone	Hypotonia	Flaccid (profound hypotonia)
Posture	Distil flexion or extensor posture	Decerebrate
Suck or Moro	Weak suck or partial moro	Absent
Autonomic system *Pupils*	Miosis (fixed pinpoint)	Fixed dilated, slow, absent, unequal or deviated
Heart Rate	Bradycardia (< 100 beats per min)	-
Respiration	Periodic or shallow breathing	Apnoea req ippv

"Moderate" *grade is defined by abnormalities in three or more categories – the majority of the abnormalities under moderate.*
"Severe" *grade is defined by abnormalities in three or more categories – the majority of the abnormalities under severe.*
NOTE: *if equal abnormalities under moderate and severe, the grade is defined by the level of consciousness.*
"Mild" *grade is defined by abnormalities in less than three categories, and/or the isolated presence of one or more of hypertonia, exaggerated reflexes, or hyperalertness.*

"Normal" *is defined by no abnormalities*

FORM CHECKED BY ……

FORM COMPLETED BY ……

DATA CAPTURED BY ……

Amendment v.5 **31 Jan 2020**
(Document v.6)

Study ID:

NESHIE STUDY: MATERNAL & OBSTETRIC DATA

NOTE: IF A TEST / EVALUATION WAS NOT DONE, PLEASE INDICATE THIS USING "ND". IF A TEST / EVALUATION WAS DONE BUT DATA / RESULTS ARE NOT KNOWN, PLEASE INDICATE THIS USING "UK"

MATERNAL & PATERNAL DETAILS
Check relevant block and provide details as requested and as applicable

Maternal Race	Black		Coloured		Indian		White		**Non-South African**	Yes	No
	Other		State "Other":.....								
Paternal Race	Black		Coloured		Indian		White		**Non-South African**	Yes	No
	Other		State "Other":.....								

MATERNAL DEMOGRAPHIC DETAILS
Check relevant block and provide details as requested and as applicable

First language			**Education: Tertiary completed?**		Yes		No		UK	**Highest grade (if no tertiary)**		
Dwelling	House		Flat		Informal settlement		Lodger from rural area outside of metro			Other / UK		
Describe "Dwelling" if "Other / UK"												

MATERNAL PREGNANCY DETAILS
Check relevant block and provide details as requested and as applicable

Confidence in date of last menstrual period	Sure		Unsure		**Did the mother receive antenatal care before deliver (was she "booked")?**	Yes [1]		No				
Date of last menstrual period					**Date of 1st ultrasound scan**							
Gestational age by dates	weeks				**Gestational age at 1st ultrasound scan**	weeks						
Estimated date of delivery by: Dates					**Estimated date of delivery by: 1st ultrasound**							
Was weight measured at 1st booking?	Yes	No	Unknown		***If " Yes"*** **Weight** kg	**Height** cm						
Hospital where birth took place					**Age at time of delivery**	years						
Gravidity		**Parity**		**How many foetuses this pregnancy?**		**No. previous miscarriages (<22w)**		**No. previous stillbirths (≥22w)**				
[1,2 & 3] Please complete Tables 1-3 as appropriate				**No. previous newborns admitted?** [2]		**No. previous newborns with NESHIE / 'Asphyxia'?** [3]						
Postnatal measurements:	Not done		**M/U arm circ** [4] cm		**Weight** kg	**Height** cm						
RVD / HIV status	Positive + ARV		Positive No ARV		Negative		Unknown					
Venereal disease status	Positive Fully treated		Positive Not fully treated		Negative		Unknown					
Venereal disease test(s) used	VDRL		RPR		TPHA							
If HIV positive, last CD4 count	cells/µL		CD4 unknown		**Lowest antenatal Hb**	g/dL						

[1] Please complete Table 1 if antenatal care was provided,

[2] Please complete Table 2 if previous newborns were admitted,

[3] Please complete Table 3 if there had previously been NESHIE / 'Asphyxia' cases

[4] Measurements taken on RIGHT arm

Study ID :

COMPLETE THE INFORMATION IN TABLE 1 ONLY IF: ANTENATAL CARE WAS RECEIVED (MOTHER WAS "BOOKED")

COMPLETE THE INFORMATION IN TABLE 2 ONLY IF: PREVIOUS NEWBORNS HAD BEEN ADMITTED

COMPLETE THE INFORMATION IN TABLE 3 ONLY IF: THERE WERE PREVIOUS NESHIE / 'ASPHYXIA' CASES

Table 1: Antenatal clinic visit booking log

Visit Number:	Date booked for visit	Attended			Visit Number:	Date booked for visit	Attended		
1		Yes	No	Unknown	13		Yes	No	Unknown
2		Yes	No	Unknown	14		Yes	No	Unknown
3		Yes	No	Unknown	15		Yes	No	Unknown
4		Yes	No	Unknown	16		Yes	No	Unknown
5		Yes	No	Unknown	17		Yes	No	Unknown
6		Yes	No	Unknown	18		Yes	No	Unknown
7		Yes	No	Unknown	19		Yes	No	Unknown
8		Yes	No	Unknown	20		Yes	No	Unknown
9		Yes	No	Unknown	21		Yes	No	Unknown
10		Yes	No	Unknown	22		Yes	No	Unknown
11		Yes	No	Unknown	23		Yes	No	Unknown
12		Yes	No	Unknown	24		Yes	No	Unknown

Table 2: Birthweight and reason for previous admissions

Admission number	Birthweight (grams)	Reason for admission	Did the newborn demise?		
1			Yes	No	Unknown
2			Yes	No	Unknown
3			Yes	No	Unknown
4			Yes	No	Unknown
5			Yes	No	Unknown

Table 3: Long-term problems for previous NESHIE / 'Asphyxia' newborns

Admission number	Long-term problem(s)	Did the newborn demise?		
1		Yes	No	Unknown
2		Yes	No	Unknown
3		Yes	No	Unknown
4		Yes	No	Unknown
5		Yes	No	Unknown

Study ID :

MATERNAL MEDICAL CONDITIONS OR TREATMENT PRESENT PRIOR TO PREGNANCY? (NOT PREGNANCY COMPLICATIONS)	Yes	No	UK

Check relevant block and provide details as requested and as applicable below if "Pregnancy complications" is "Yes"

Anaemia	Yes	No	UK	**Cardiac disease**	Yes	No	UK	**Diabetes**	Yes	No	UK
Epilepsy / Seizure	Yes	No	UK	**Hypertension**	Yes	No	UK	**Thyroid disease: Hypothyroidism**	Yes	No	UK
Thyroid disease: Hyperthyroidism	Yes	No	UK	**Other**	Yes	No		**Detail if "Other" is "Yes"**			

.....

.....

Treatment for pre-existing conditions?	Yes	No	UK	**Detail if "Yes"** .....

.....

.....

COMPLICATIONS DURING PREGNANCY BEFORE ONSET OF LABOUR / PREGNANCY COMPLICATIONS?	Yes	No	UK

Check relevant block and provide details as requested and as applicable below if "Pregnancy complications" is "Yes"

Anaemia (gestational)	Yes	No	UK	**APH / Bleeding**	Yes	No	UK	**Clinical Chorioamnionitis**	Yes	No	UK
Diabetes (gestational)	Yes	No	UK	**Hypertension (gestational)**	Yes	No	UK	**Intra-uterine growth restriction**	Yes	No	UK
Placenta Praevia	Yes	No	UK	**Pre-eclampsia / eclampsia / HELLP**	Yes	No	UK	**PROM > 18h**	Yes	No	UK
Pyrexia	Yes	No	UK	**TB (Active) during pregnancy**	Yes	No	UK	**On treatment if "TB"= "Yes"?**	Yes	No	UK
Other	Yes	No		**Detail if "Other" is "Yes"**							

.....

.....

Infection during pregnancy?	Yes	No	UK	**Selection option only if "Infection" is "Yes"**	Urinary	Vaginal/Vulval	Other

Detail if "Infection" is "Other"

RECREATION / HABITS DURING PREGNANCY?	Yes	No	UK

Check relevant block and provide details as requested and as applicable if "Yes" to any recreational habits

Alcohol	Yes	No	UK	Description:
Cigarettes / Smoker	Yes	No	UK	Description:
Illicit drugs	Yes	No	UK	Description:
Other	Yes	No	UK	Description:

MEDICATIONS (NOT SUPPLEMENTS) USED IN PREGNANCY?	Yes	No	UK

Describe in full if "Yes" .....

.....

Study ID :

CTG DATA BEFORE ONSET OF LABOUR CONTRACTIONS (<24 HOURS BEFORE ONSET OF LABOUR)						
Scan last 30 minutes of CTG before onset of labour if available						
CTG <24 hours before onset of labour	Available		Not done		Done, not available	
Reason for CTG done (main risk factor):						
.....						
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE / DONE <24 HOURS BEFORE ONSET OF LABOUR						
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

ONSET OF LABOUR						
Check relevant block and provide details as requested and as applicable						
Onset of labour: Admission or first noted while admitted	**Date**		**Time**			
Nature of labour	Spontaneous		Induced		Caesarean Section (not in labour)	
Are there other notes on the nature of labour?	Yes	No	Describe if "Yes"			
.....						
Method of induction (Check all that are applicable)	Catheter bulb		PGE2		Misoprostol	
	Oxytocin		AROM		Other	
Describe if "induction" is "Other"						
Fetal movements at onset of labour	Normal		Reduced		Not noted	

FETAL CONDITION AFTER ONSET OF LABOUR CONTRACTIONS						
Scan all available CTG trace after onset of labour						
CTG in labour earlier than the last 2 hours before delivery	Available		Not done		Done, not available	
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE / DONE AFTER ONSET OF LABOUR						
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

MATERNAL CONDITION AFTER ONSET OF LABOUR CONTRACTIONS												
Highest maternal heart rate in last four hours of labour	<100 bpm		101-110 bpm		111-120 bpm		> 120 bpm					
Analgesia in labour (Check all that are applicable)	None		Opiate		Epidural		Other					
Describe if "Analgesia" is "Other"												
Duration of oxytocin augmentation (if used; completed hours)	N/A		hours		**Highest temperature in labour**				°C			
Duration of active phase of first stage			1st time found cervix ≥ **4 cm** to 1st time 10 cm		h	min	1st time found cervix ≥ **5 cm** to 1st time 10 cm		h	min		
Duration of second stage			1st time found cervix 10 cm to birth		min		**Duration of ruptured membranes**		hours			
Liquor	Clear		Mec+		Mec++		Mec+++		**Offensive liquor**	Yes	No	

Amendment v.5 **31 Jan 2020**
(Document v.6)

Study ID :

LAST 2 HOURS OF FHR/CTG IN LABOUR												
Scan all CTG tracing in the last 2 hours before delivery												
CTG from last 2 hours before delivery	Available		Not done		Done, not available							
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE FOR >30 MINS AND DONE IN LAST 2 HOURS OF LABOUR												
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown					

DELIVERY COMPLICATIONS / SENTINEL EVENTS?												
									Yes	No	UK	
Complete block below only if "Delivery Complications" is "Yes"												
Fetomaternal haemorrhage	Yes	No	UK	**Other Ante Partum Haemorrhage**	Yes	No	UK	**Placental abruption**	Yes	No	UK	
Maternal Hypoxia	Yes	No	UK	**Maternal collapse**	Yes	No	UK	**Sudden onset bradycardia**	Yes	No	UK	
Prolapse Cord	Yes	No	UK	**Ruptured Uterus**	Yes	No	UK	**Prolonged 2nd Stage (>2 hours)**	Yes	No	UK	
Shoulder Dystocia	Yes	No	UK	**Difficult breech**	Yes	No	UK	**Other Sentinel Event(s)**	Yes	No	UK	
Detail if "Other Sentinel events" is "YES":												
.....												

UMBILICAL CORD ABNORMALITIES *(Check relevant block)*	None		True knot		Cord around neck twice or more		Single artery	
	Other		Abnormal insertion		Hypercoiling		Unknown	
	Detail "Other" or "Abnormal insertion".....							
	.....							

PLACENTAL ABNORMALITIES *(Check relevant block)*	None		Abnormal		Unknown		
	Detail if "Abnormal"						
	.....						

ROUTE OF DELIVERY *(Check relevant block)*	Pre-labour CS		In-labour CS		Vaginal	
PRESENTATION	Cephalic		Breech			
MODE OF VAGINAL DELIVERY *(Check relevant block)*	Spontaneous		Vacuum		**FAILED VACUUM OR FORCEPS?**	Yes
	Caesarean section		Forceps			No
Reason for vacuum or forceps						
.....						
Reason for caesarean section						
.....						

DIFFICULTIES WITH DELIVERY	Yes	No	UK
If yes, detail			
.....			

Study ID:

Appendix E: A7 CRF

NESHIE STUDY: SCREENING SHEET

NEONATAL SCREENING INFORMATION								
Hospital providing cooling								
DOB				**TOB**				
The neonate was:	Inborn		Outborn		**Sex**	M	F	Intersex

BIRTH SITE INFORMATION		
Province		
Health Sub-district		
Name of Institution		
Is the Province in which the mother gave birth the same as the province in which the mother permanently resides?	Yes	No
If no, in which Province does the mother reside permanently?		

Study ID:

NESHIE STUDY: INCLUSION/EXCLUSION CRITERIA

COMPLETE NESHIE SCREENING FORM if category A, B & C conditions are met.

FOR ENROLMENT INTO THE NESHIE STUDY: All inclusion criteria should be met with no exclusion criteria present

A: (All must be Yes to be cooled and to be included in study)		
≥ 36 weeks gestation*	**Yes**	**No**
≥ 1800g	**Yes**	**No**

* As determined by foetal ultra sound at ≤ 20 weeks; or postnatal Ballard, or Foot length measurements (≥73mm in length)

B: (at least <u>one</u> must be Yes to be cooled and to be included in study)			
BD 1st hr ≥ 16 / pH ≤ 7	**Yes**	**No**	N/A
BD 1st hr ≥ 10 / pH 1st hr ≤ 7.15 and peripartum/ sentinel event	**Yes**	**No**	N/A
5min Apgar < 7 (only if early gas not available)	**Yes**	**No**	
Required resus/assisted ventilation at 10min	**Yes**	**No**	

C1: Which <u>sign</u> of encephalopathy was present before cooling? (At least one must be "Yes" to qualify for enrolment into the study)			**C2: Which other abnormal signs were present before cooling** (If clinical seizures are **NOT** present, at least one of these must be "Yes" to be enrolled in the study)		
Lethargy	**Yes**	**No**	Hypotonia	**Yes**	**No**
Stupor	**Yes**	**No**	Abnormal reflexes	**Yes**	**No**
Coma	**Yes**	**No**	Abnormal suck	**Yes**	**No**
Thompson score ≥ 7	**Yes**	**No**			
Seizures	**Yes**	**No**	*If "**<u>yes</u>**" to **<u>seizures</u>**, describe:*		
.....					

D: Exclusion criteria (Tick all that apply. The presence of any exclusion criteria disqualifies the patient from enrolment.)					
None		aEEG not performed		Neonate not cooled ≤ 6 hours of life	
Moribund and unlikely to benefit from cooling		Asystole		Hypotension or bleeding that is not responding to treatment	
Severe PPHN		Suspected chromosomal abnormality		Congenital infection	
Surgical anomaly		Consent refused		Neonate died prior to obtaining consent	
Consent not obtained for other reason	State:				
Encephalopathy primarily due to non-hypoxic cause	State:				
Congenital abnormality	State:				
Neonate not cooled	State:				
Other	State:				

E: Can it be confirmed that counselling was provided prior to seeking informed consent? (Must be "YES" to be included in the study)	Yes	No
F: Was the patient recruited?	Yes	No
If "no", why not?		

Study ID:

NESHIE STUDY: PRE-COOLING ASSESSMENT AND COOLING METHOD

CLINICAL DETAILS OF BABY AT BIRTH *Check relevant block and provide details as requested and as applicable*									
Method used for GA	Dates		Ballard		Ballard score		**Gestation at birth**		wks
	EUS at ≤20 weeks		Footlength		Length (cm)		**Birth weight**		g
Apgar scores:	1 min		5 min		10 min		**COH**		cm

Adrenaline	Yes	No	**BMV**	Yes	No	**Chest compressions**	Yes	No
Delayed cord clamping / Cord milking	Yes	No	**Intubated**	Yes	No	**Early blood gas results available**[1]	Yes	No

[1] Worst base excess within 60 minutes of birth including cord blood

BLOOD GAS EVALUATION: AT BIRTH *Check relevant block and provide details as requested and as applicable*			**Done and available?**	Yes	No	Not done
Date taken		**Source: Blood gas**	Cord (venous)		Cord (arterial)	
Time taken			Cord (mixed source)		Infant (ABG)	

Lowest blood glucose		mmol/L	**Highest blood glucose**		mmol/L	**pH** (POC)		
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
PCO_2..... (POC)	kPa	mmHg	**PaO_2**.....	kPa	mmHg	**FiO_2**		%
Lab: Hb		g/dL	**Lab: NRBC**		%	**Lab: WBC**		cells/L
Lab: Neutrophils		cells/L	**Lab: Platelets**		cells/L			

Blood culture result for culture taken on this day	Positive	No growth	Not performed
Organism(s) if positive:			

CLINICAL EXAMINATION PRIOR TO COOLING *Check relevant block and provide details as requested and as applicable*								
CPAP	Yes	No	**Nasal Cannulae**	Yes	No	**Additional oxygen provided? (E.g. FiO_2>21%)**	Yes	No
Mechanical ventilation	Yes	No	**Hypotension or Inotropes**[1]	Yes	No	**Active bleeding treated with blood product**	Yes	No
Subaponeurotic haemorrhage	Yes	No	**Sinus bradycardia <80**	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No

[1] Hypotension = MBP persistently < 40mmHg

Study ID:

NEUROLOGICAL ASSESSMENT PRIOR TO COOLING (NOT EARLIER THAN 30 MINS) OR AT ONSET OF COOLING							
Date taken		**Time taken**		**Age at neurological assessment[1]**		hr	min
Thompson HIE score		**Modified Sarnat HIE grade[2]**		**Visible seizures**		Yes	No

[1] Same as Thomson score age, [2] worst grade prior to cooling

CFM DETAILS						
Model used:					**No. of channels used**	
How were measurements taken?			Needle		Skin patch	
Indicate the CFM /aEEG monitoring status			Continuous		Intermittent	
Initiation of aEEG	Prior to cooling		At onset of cooling		Other	
Describe if aEEG initiation is "Other":						

BASELINE aEGG ASSESSMENT							Time of assessment			:	
Is the baseline aEEG assessment different to the aEEG evaluation at 6H							Yes		No		
Electrical seizures present?							Yes		No		
Assessment status				Normal			Abnormal		Not done		
Voltage:				Normal			Moderately abnormal		Suppressed		
Pattern:	Normal		FT		BS		CLV		DNV	CNV	

aEGG ASSESSMENT: 6 HOURS OF LIFE* **Complete this table only if 6H aEEG is different from baseline assessment*							Time of assessment			:	
Electrical seizures present?							Yes		No		
Assessment status				Normal			Abnormal		Not done		
Voltage:				Normal			Moderately abnormal		Suppressed		
Pattern:	Normal		FT		BS		CLV		DNV	CNV	

Study ID:

BASELINE LABORATORY INVESTIGATION: ON ADMISSION OR AS CLOSE TO COOLING AS POSSIBLE *COMPLETE ONLY IF DIFFERENT FROM BLOOD GAS AT BIRTH Check relevant block and provide details as requested and as applicable.*								
Date taken		**Source: Blood gas**	Cord (venous)		Cord (arterial)			
Time taken			Cord (mixed source)		Infant (ABG)			
Lowest blood glucose		mmol/L	**Highest blood glucose**		mmol/L	**pH** (POC)		
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
PCO_2..... (POC)	kPa	mmHg	**PaO_2**.....	kPa	mmHg	**FiO_2**		%
Lab: Hb		g/dL	**Lab: NRBC**		%	**Lab: WBC**		cells/L
Lab: Neutrophils		cells/L	**Lab: Platelets**		cells/L			
Blood culture result for culture taken on this day						Positive	No growth	Not performed
Organism(s) if positive:								

COOLING INDUCTION, METHOD AND SETTING											
Date *admitted* to cooling centre			**Time *admitted* to cooling centre**			hr	min				
Date cooling *commenced*			**Time cooling *commenced***			hr	min				
Cooling method used:	Automated whole body cooling		State make:								
	Servo-controlled gel bag method		Coolcap cooling		Manual Techotherm		Other				
Describe if "Other"											
Temperature before cooling started		°C	**Target core temperature**		°C	**Intubation/IPPV available in cooling unit?**		Yes	No		
Target core Temp =attained: Date			**Target core Temp attained: Time**			**Invasive BP monitoring available in cooling unit?**		Yes	No		
What is the average number of babies per nurse?						**Inotrope infusions available in cooling unit?**		Yes	No		
What level of care is the baby nursed in?		NICU	High Care	**If "High Care" selected, was the patient moved to NICU at any point in time?**			N/A	Yes	No		

Study ID:

NESHIE STUDY: SAMPLE COLLECTION SHEET

Fill in details or check the applicable answer

PLACENTA SENT FOR HISTOLOGY/PATHOLOGY	Yes	Date sent:	No
If yes, brief result:			
If no, why not:			

Placenta: Pathomicrobiome sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Placenta: Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

Cord blood (Venous): DNA & RNA sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Cord blood (Venous): Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

Blood spot: <60 min sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
[1] Blood spot: 1-6 hour sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Blood spot: 48-72 sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[1] Only applicable if blood spots cannot be obtained within the first hour of neonatal life

Neonatal urine: Metabolomics sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[2] Peripheral blood: Neonatal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Peripheral blood: Maternal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					
Peripheral blood: Paternal sample collected?	Yes	Date Collected:	Time Collected:	No	N/A
If no, explain why not:					

[2] Only applicable if neonatal venous cord blood cannot be obtained within the first hour of neonatal life

Study ID:

NESHIE STUDY: DAILY GENERAL MONITORING DATA *after cooling commencement*

COOLING PERIOD							DAY 1
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	0				**12**		
	1				**13**		
	2				**14**		
	3				**15**		
	4				**16**		
	5				**17**		
	6				**18**		
	7				**19**		
	8				**20**		
	9				**21**		
	10				**22**		
	11				**23**		

CLINICAL PRESENTATION DURING THE 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock

INVESTIGATIONS DONE AT 24H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
Lab: Sodium (Na)		mmol/L	**Lab: Potassium (K)**		mmol/L	**Lab: Calcium (Ca)**		mmol/L
Lab: Magnesium (Mg)		mmol/L	**Lab: Urea**		mmol/L	**Lab: Creatinine**		mmol/L

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 1ST 24 HOURS OF LIFE										
aEEG @ age 24H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Suppressed		

FLUID / FEED / TPN DURING 1ST 24 HOURS OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D1)					
Was a CUS done on D1?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D1 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D1 comments/observations			
Additional D1 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

COOLING PERIOD							DAY 2
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	24				**36**		
	25				**37**		
	26				**38**		
	27				**39**		
	28				**40**		
	29				**41**		
	30				**42**		
	31				**43**		
	32				**44**		
	33				**45**		
	34				**46**		
	35				**47**		

CLINICAL PRESENTATION DURING THE 2ND DAY OF LIFE									
Check relevant block and provide details as requested and as applicable									
Clinical or aEEG/EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**		Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**		Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**		Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**			Yes	No	Not performed	
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**						

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time HR has been moved from Fluid/Feed/TPN/HR block, FiO_2 Oxygen replaced with "Additional oxygen provided (E.g. FiO2 > 21%)"

INVESTIGATIONS DONE AT 36 - 48H								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L
Lab: CRP		mg/dL						

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 2ND DAY OF LIFE										
aEEG @ age 48H	**Pattern:**	FT		BS		CLV		DNV		CNV	
	Voltage:		Normal			Moderately abnormal			Suppressed		

FLUID / FEED / TPNDURING 2ND DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Have the D2 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D2 comments/observations			
Additional D2 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

COOLING PERIOD							DAY 3
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	48				**60**		
	49				**61**		
	50				**62**		
	51				**63**		
	52				**64**		
	53				**65**		
	54				**66**		
	55				**67**		
	56				**68**		
	57				**69**		
	58				**70**		
	59				**71**		

CLINICAL PRESENTATION DURING THE 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time

INVESTIGATIONS DONE AT 72H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**	**POC: pH**	**POC: Hb**		g/dL
POC: Sodium (Na)		mmol/L	**POC: Potassium (K)**		mmol/L	**POC: Bicarb**		mmol/L
POC: Lactate		mmol/L	**POC: Base Excess**		mmol/L	**POC: iCa**		mmol/L

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 3RD DAY OF LIFE										
aEEG @ age 72H	**Pattern:**	FT		BS		CLV		DNV		CNV
	Voltage:		Normal			Moderately abnormal			Severely Suppressed	

FLUID / FEED / TPN DURING 3RD DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D3)					
Was a CUS done on D3?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D3 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D3 comments/observations			
Additional D3 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

REWARMING PERIOD							DAY 4
NOTE: Please include a value for every hour of cooling. If a recording was not made for any particular hour, please indicate this.							
Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)	Time	Hours from: Start cool	Target Temp (°C)	Core Temp (°C)
	72				**84**		
	73				**85**		
	74				**86**		
	75				**87**		
	76				**88**		
	77				**89**		
	78				**90**		
	79				**91**		
	80				**92**		
	81				**93**		
	82				**94**		
	83				**95**		

CLINICAL PRESENTATION DURING THE 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Clinical or aEEG /EEG Seizures	Yes	No	**CPAP**	Yes	No	**Nasal Cannulae**	Yes	No
Additional oxygen provided (E.g. FiO_2>21%)	Yes	No	**Mechanical ventilation**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Sinus bradycardia <80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Bleeding/SAH**	Yes	No
Min HR **BPM**	**Min HR Time taken[2]**		**Sepsis (Proven on culture)**	Yes	No	Not performed		
Max HR **BPM**	**Max HR Time taken[2]**		**Organism(s) if positive:**					

[1]Hypotension = MBP persistently <40mmHg [2]According to 24H clock time

INVESTIGATIONS DONE AT 96H (± 6H) OF LIFE								
Lowest blood glucose		mmol/L	**POC: PCO_2**.....	kPa	mmHg	**POC: PaO_2**.....	kPa	mmHg
Highest blood glucose		mmol/L	**POC: FiO_2**	**%**				

Study ID:

SEDATION / ANTICONVULSANTS GIVEN DURING 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Sedation / Anticonvulsants	None		Morphine		Phenobarb		Midazolam	
	Lignocaine		Phenytoin		Levetiracetam		Other	
Describe in full if "Other" is indicated								

CFM FINDINGS DURING 4TH DAY OF LIFE										
aEEG after rewarm	**Pattern:**	FT		BS		CLV		DNV		CNV
	Voltage:		Normal			Moderately abnormal		Severely Suppressed		

FLUID / FEED / TPN DURING 4TH DAY OF LIFE								
Check relevant block and provide details as requested and as applicable								
Intravenous fluids (excl TPN)	Yes	No	**Enteral feeds**	Yes	No	**TPN**	Yes	No
Intravenous fluids (excl TPN)[1]			**Enteral feeds[1]**			**TPN[1]**		
Any additional notes?	Yes	No	Describe if "additional notes" is "Yes"					
.....								
.....								

[1]Unit of measure = mL/kg/day

Cranial Ultrasound (D4)					
Was a CUS done on D4?	Yes	No	**If "Yes", Age taken**	Hrs	**Observations:** .....
.....					
.....					
Complete CUS report (Annexure D. d) and upload onto REDCap					

Have the D4 Thompson & Modified Sarnat scores been completed?	Yes	No

Additional D4 comments/observations			
Additional D4 comments/observations?	Yes	No	**Observations:** .....
.....			
.....			
.....			
.....			

Study ID:

NESHIE STUDY: HOSPITAL COURSE AND DISCHARGE / TRANSFER

DIAGNOSES & SUPPORTIVE MEASURES DURING ENTIRE ADMISSION								
Check relevant block and provide details as requested and as applicable. Refer to Annexure D.e for DEFINITIONS OF TERMS								
Anaemia	Yes	No	**Hypoglycaemia**	Yes	No	**Active bleeding treated with blood product**	Yes	No
Hyponatraemia	Yes	No	**Hypokalaemia**	Yes	No	**Renal failure treated with dialysis**	Yes	No
Pulmonary hypertension	Yes	No	**Pulmonary airleak**	Yes	No	**Hypotension or Inotropes[1]**	Yes	No
Pneumonia	Yes	No	**Pulmonary haemorrhage**	Yes	No	**Mechanical ventilation**	Yes	No
Major cerebral anomaly	Yes	No	**Subaponeurotic haemorrhage**	Yes	No	**Nasal Cannulae**	Yes	No
Sinus bradycardia < 80	Yes	No	**Arrhythmia (other than SB <80)**	Yes	No	**Additional oxygen provided (E.g. FiO_2>21%)**	Yes	No
Necrotising enterocolitis	Yes	No	**Late onset sepsis (> 72h)**	Yes	No	**CPAP**	Yes	No
Meconium aspiration	Yes	No	**Other**	Yes	No	**None of these**	Yes	

[1]Hypotension = MBP persistently <40mmHg

OTHER DIAGNOSES OR DIAGNOSTIC INVESTIGATIONS	Yes	No
Describe if "Other" is "Yes":.....		
.....		
.....		
.....		
.....		

WAS THERE ANY CONDITION OR EVENT LIKELY TO BE DUE TO COOLING TREATMENT OR REWARMING	Yes	No
If yes, detail		
.....		

WAS COOLING STOPPED <72 HOURS	Yes	No
If yes, detail why		
.....		

FLUIDS FOLLOWING COOLING AND REWARMINNG [1]											
Check relevant block and provide details as requested and as applicable											
Fluids: D5	Yes	No	N/A	**Fluids: D6**	Yes	No	N/A	**Fluids: D7**	Yes	No	N/A
D5: Total				**D6: Total**				**D7: Total**			
D5: Enteral				**D6: Enteral**				**D7: Enteral**			

[1]Unit of measure = mL/kg/day

Study ID:

DISCHARGE / TRANSFER / DEATH STATUS							
Discharge home? (If "Yes", indicate Age in days on which this occurred)				Yes	No	N/A	days
Transferred to other hospital? (If "Yes", indicate Age in days on which this occurred)				Yes	No	N/A	days
Name of hospital to which neonate was transferred							
Death? (If "Yes", indicate Age in days on which this occurred)	Yes	No	N/A	days			hours
If "Yes" to "Death", was a PM performed?	Yes	No	N/A	Result:			
.....							

FULL SUCKING / CUP FEEDING ESTABLISHED BY DISCHARGE	Yes	No	N/A
If yes, age established?	days		

CLINICAL DETAILS OF BABY AT DISCHARGE OR TRANSFER *Check relevant block and provide details as requested and as applicable*			
Weight at discharge / transfer	g	**Circumference of Head at discharge / transfer**	cm
Have the D5 – D10, Discharge / Transfer Thompson & Modified Sarnat scores been completed?	Yes	No	

MRI INFORMATION						
Check relevant block and provide details as requested and as applicable. ***MRI files to be stored onto NESHIE External Hard Drive***						
MRI scan performed?	Yes	No	Booked, not performed	N/A	**Date performed**	
MRI Results (if performed)						
.....						
.....						

WAS A CUS PERFORMED BETWEEN DAY 5 AND DAY 14, DISCHARGE OR TRANSFER?	Yes	No	N/A
If "Yes", indicate on which day(s) a CUS was performed			
Upload completed CUS report (Annexure D. d) for each applicable day / time period onto REDCap.			

Study ID:

NESHIE STUDY: THOMPSON HIE AND MODIFIED SARNAT SCORES

	0	1	2	3	Pre-Cool < 6h	6h	24h	48h	72h	96h	D5	D6	D7	D8	D9	D10	Discharge OR Transfer
	Date																
Modified Sarnat HIE categories Mod or Sev or Mild or Norm																	
Limb Tone	Norm	HypEr-tonic	HypO-tonic	Flaccid													
LOC	Norm	Hyper-alert or staring	Lethargic or Obtunded	Coma or Stuporose													
Visible Fits	None	Infrequent < 3/day	Frequent > 2/day														
Posture	Norm/ Other	Fisting and / or Cycling	Strong distil flexion	Decerebrate													
Moro	Norm	Partial	Absent														
Grasp	Norm	Poor	Absent														
Suck	Norm	Poor	Absent and/or bites														
Respiration	Norm	Hyper-ventilation	Transient apnoea	Apnoea *requiring* IPPV													
Fontanel	Norm	Full	Tense														
TOTAL																	

Modified Sarnat HIE categories

Category	Moderate Encephalopathy	Severe Encephalopathy
Level of Consciousness	Lethargic	Obtunded/Stuporous/Coma
Spontaneous activity	Decreased spontaneous activity	No spontaneous activity
Muscle tone	Hypotonia	Flaccid (profound hypotonia)
Posture	Distil flexion or extensor posture	Decerebrate
Suck or Moro	Weak suck or partial moro	Absent
Autonomic system *Pupils*	Miosis (fixed pinpoint)	Fixed dilated, slow, absent, unequal or deviated
Heart Rate	Bradycardia (< 100 beats per min)	-
Respiration	Periodic or shallow breathing	Apnoea req ippv

*"**Moderate**" grade is defined by abnormalities in three or more categories – the majority of the abnormalities under moderate.*
*"**Severe**" grade is defined by abnormalities in three or more categories – the majority of the abnormalities under severe.*
***NOTE:** if equal abnormalities under moderate and severe, the grade is defined by the level of consciousness.*
*"**Mild**" grade is defined by abnormalities in less than three categories, and/or the isolated presence of one or more of hypertonia, exaggerated reflexes, or hyperalertness.*

*"**Normal**" is defined by no abnormalities*

FORM CHECKED BY ……

FORM COMPLETED BY ……

DATA CAPTURED BY ……

Study ID:

NESHIE STUDY: MATERNAL & OBSTETRIC DATA

NOTE: IF A TEST / EVALUATION WAS NOT DONE, PLEASE INDICATE THIS USING "ND". IF A TEST / EVALUATION WAS DONE BUT DATA / RESULTS ARE NOT KNOWN, PLEASE INDICATE THIS USING "UK"

MATERNAL & PATERNAL DETAILS
Check relevant block and provide details as requested and as applicable

Maternal Race	Black		Coloured		Indian		White		**Non-South African**	Yes	No
	Other		State "Other":.....								
Paternal Race	Black		Coloured		Indian		White		**Non-South African**	Yes	No
	Other		State "Other":.....								

MATERNAL DEMOGRAPHIC DETAILS
Check relevant block and provide details as requested and as applicable

First language				**Education: Tertiary completed?**		Yes		No		UK	**Highest grade (if no tertiary)**	
Dwelling	House		Flat		Informal settlement		Lodger from rural area outside of metro				Other / UK	
Describe "Dwelling" if "Other / UK"												

MATERNAL PREGNANCY DETAILS
Check relevant block and provide details as requested and as applicable

Confidence in date of last menstrual period	Sure		Unsure		**Did the mother receive antenatal care before deliver (was she "booked")?**	Yes [1]		No			
Date of last menstrual period					**Date of 1st ultrasound scan**						
Gestational age by dates	weeks				**Gestational age at 1st ultrasound scan**	weeks					
Estimated date of delivery by: Dates					**Estimated date of delivery by: 1st ultrasound**						
Was weight measured at 1st booking?	Yes	No	Unknown		***If " Yes"*** **Weight** kg	**Height** cm					

Hospital where birth took place						**Age at time of delivery**				years	
Gravidity		**Parity**		**How many foetuses this pregnancy?**		**No. previous miscarriages (<22w)**		**No. previous stillbirths (≥22w)**			
[1,2 & 3] Please complete Tables 1-3 as appropriate				**No. previous newborns admitted?** [2]		**No. previous newborns with NESHIE / 'Asphyxia'?** [3]					
Postnatal measurements:	Not done			**M/U arm circ** [4] cm		**Weight** kg		**Height** cm			

RVD / HIV status	Positive + ARV		Positive No ARV		Negative		Unknown				
Venereal disease status	Positive Fully treated		Positive Not fully treated		Negative		Unknown				
Venereal disease test(s) used	VDRL		RPR		TPHA						
If HIV positive, last CD4 count	cells/µL		CD4 unknown		**Lowest antenatal Hb**		g/dL				

[1] Please complete Table 1 if antenatal care was provided,

[2] Please complete Table 2 if previous newborns were admitted,

[3] Please complete Table 3 if there had previously been NESHIE / 'Asphyxia' cases

[4] Measurements taken on RIGHT arm

Study ID:

COMPLETE THE INFORMATION IN TABLE 1 ONLY IF: ANTENATAL CARE WAS RECEIVED (MOTHER WAS "BOOKED")

COMPLETE THE INFORMATION IN TABLE 2 ONLY IF: PREVIOUS NEWBORNS HAD BEEN ADMITTED

COMPLETE THE INFORMATION IN TABLE 3 ONLY IF: THERE WERE PREVIOUS NESHIE / 'ASPHYXIA' CASES

Table 1: Antenatal clinic visit booking log

Visit Number:	Date booked for visit	Attended			Visit Number:	Date booked for visit	Attended		
1		Yes	No	Unknown	13		Yes	No	Unknown
2		Yes	No	Unknown	14		Yes	No	Unknown
3		Yes	No	Unknown	15		Yes	No	Unknown
4		Yes	No	Unknown	16		Yes	No	Unknown
5		Yes	No	Unknown	17		Yes	No	Unknown
6		Yes	No	Unknown	18		Yes	No	Unknown
7		Yes	No	Unknown	19		Yes	No	Unknown
8		Yes	No	Unknown	20		Yes	No	Unknown
9		Yes	No	Unknown	21		Yes	No	Unknown
10		Yes	No	Unknown	22		Yes	No	Unknown
11		Yes	No	Unknown	23		Yes	No	Unknown
12		Yes	No	Unknown	24		Yes	No	Unknown

Table 2: Birthweight and reason for previous admissions

Admission number	Birthweight (grams)	Reason for admission	Did the newborn demise?		
1			Yes	No	Unknown
2			Yes	No	Unknown
3			Yes	No	Unknown
4			Yes	No	Unknown
5			Yes	No	Unknown

Table 3: Long-term problems for previous NESHIE / 'Asphyxia' newborns

Admission number	Long-term problem(s)	Did the newborn demise?		
1		Yes	No	Unknown
2		Yes	No	Unknown
3		Yes	No	Unknown
4		Yes	No	Unknown
5		Yes	No	Unknown

Study ID:

MATERNAL MEDICAL CONDITIONS OR TREATMENT PRESENT PRIOR TO PREGNANCY? (NOT PREGNANCY COMPLICATIONS)											
									Yes	No	UK
Check relevant block and provide details as requested and as applicable below if "Pregnancy complications" is "Yes"											
Anaemia	Yes	No	UK	**Cardiac disease**	Yes	No	UK	**Diabetes**	Yes	No	UK
Epilepsy / Seizure	Yes	No	UK	**Hypertension**	Yes	No	UK	**Thyroid disease: Hypothyroidism**	Yes	No	UK
Thyroid disease: Hyperthyroidism	Yes	No	UK	**Other**	Yes	No		**Detail if "Other" is "Yes"**			
.....											
.....											
Treatment for pre-existing conditions?	Yes	No	UK	**Detail if "Yes"**							
.....											
.....											

COMPLICATIONS DURING PREGNANCY BEFORE ONSET OF LABOUR / PREGNANCY COMPLICATIONS?											
									Yes	No	UK
Check relevant block and provide details as requested and as applicable below if "Pregnancy complications" is "Yes"											
Anaemia (gestational)	Yes	No	UK	**APH / Bleeding**	Yes	No	UK	**Clinical Chorioamnionitis**	Yes	No	UK
Diabetes (gestational)	Yes	No	UK	**Hypertension (gestational)**	Yes	No	UK	**Intra-uterine growth restriction**	Yes	No	UK
Placenta Praevia	Yes	No	UK	**Pre-eclampsia / eclampsia / HELLP**	Yes	No	UK	**PROM > 18h**	Yes	No	UK
Pyrexia	Yes	No	UK	**TB (Active) during pregnancy**	Yes	No	UK	**On treatment if "TB"= "Yes"?**	Yes	No	UK
Other	Yes	No		**Detail if "Other" is "Yes"**							
.....											
.....											
Infection during pregnancy?	Yes	No	UK	**Selection option only if "Infection" is "Yes"**	Urinary		Vaginal/Vulval		Other		
Detail if "Infection" is "Other"											

RECREATION / HABITS DURING PREGNANCY?						
					Yes	No	UK
Check relevant block and provide details as requested and as applicable if "Yes" to any recreational habits						
Alcohol	Yes	No	UK	Description:		
Cigarettes / Smoker	Yes	No	UK	Description:		
Illicit drugs	Yes	No	UK	Description:		
Other	Yes	No	UK	Description:		

MEDICATIONS (NOT SUPPLEMENTS) USED IN PREGNANCY?			
	Yes	No	UK
Describe in full if "Yes"			
.....			

Study ID:

CTG DATA BEFORE ONSET OF LABOUR CONTRACTIONS (<24 HOURS BEFORE ONSET OF LABOUR)						
Scan last 30 minutes of CTG before onset of labour if available						
CTG <24 hours before onset of labour	Available		Not done		Done, not available	
Reason for CTG done (main risk factor):						
.....						
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE / DONE <24 HOURS BEFORE ONSET OF LABOUR						
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

ONSET OF LABOUR						
Check relevant block and provide details as requested and as applicable						
Onset of labour: Admission or first noted while admitted	**Date**		**Time**			
Nature of labour	Spontaneous		Induced		Caesarean Section (not in labour)	
Are there other notes on the nature of labour?	Yes	No	Describe if "Yes"			
.....						
Method of induction (Check all that are applicable)	Catheter bulb		PGE2		Misoprostol	
	Oxytocin		AROM		Other	
Describe if "induction" is "Other"						
Fetal movements at onset of labour	Normal		Reduced		Not noted	

FETAL CONDITION AFTER ONSET OF LABOUR CONTRACTIONS						
Scan all available CTG trace after onset of labour						
CTG in labour earlier than the last 2 hours before delivery	Available		Not done		Done, not available	
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE / DONE AFTER ONSET OF LABOUR						
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

MATERNAL CONDITION AFTER ONSET OF LABOUR CONTRACTIONS												
Highest maternal heart rate in last four hours of labour	<100 bpm		101-110 bpm		111-120 bpm		> 120 bpm					
Analgesia in labour (Check all that are applicable)	None		Opiate		Epidural		Other					
Describe if "Analgesia" is "Other"												
Duration of oxytocin augmentation (if used; completed hours)	N/A		hours		**Highest temperature in labour**				°C			
Duration of active phase of first stage			1st time found cervix ≥ **4 cm** to 1st time 10 cm		h	min	1st time found cervix ≥ **5 cm** to 1st time 10 cm		h	min		
Duration of second stage			1st time found cervix 10 cm to birth		min		**Duration of ruptured membranes**		hours			
Liquor	Clear		Mec+		Mec++		Mec+++		**Offensive liquor**	Yes	No	

Study ID:

LAST 2 HOURS OF FHR/CTG IN LABOUR							
Scan all CTG tracing in the last 2 hours before delivery							
CTG from last 2 hours before delivery	Available		Not done		Done, not available		
COMPLETE LINE BELOW ONLY IF NO CTG AVAILABLE FOR >30 MINS AND DONE IN LAST 2 HOURS OF LABOUR							
If "no CTG", was FHR normal/reactive or reassuring?	Yes	No	Unknown	**Decelerations noted?**	Yes	No	Unknown

DELIVERY COMPLICATIONS / SENTINEL EVENTS?											
									Yes	No	UK
Complete block below only if "Delivery Complications" is "Yes"											
Fetomaternal haemorrhage	Yes	No	UK	**Other Ante Partum Haemorrhage**	Yes	No	UK	**Placental abruption**	Yes	No	UK
Maternal Hypoxia	Yes	No	UK	**Maternal collapse**	Yes	No	UK	**Sudden onset bradycardia**	Yes	No	UK
Prolapse Cord	Yes	No	UK	**Ruptured Uterus**	Yes	No	UK	**Prolonged 2nd Stage (>2 hours)**	Yes	No	UK
Shoulder Dystocia	Yes	No	UK	**Difficult breech**	Yes	No	UK	**Other Sentinel Event(s)**	Yes	No	UK
Detail if "Other Sentinel events" is "YES":											

UMBILICAL CORD ABNORMALITIES *(Check relevant block)*	None		True knot		Cord around neck twice or more		Single artery	
	Other		Abnormal insertion		Hypercoiling		Unknown	
	Detail "Other" or "Abnormal insertion".....							

PLACENTAL ABNORMALITIES *(Check relevant block)*	None		Abnormal		Unknown	
	Detail if "Abnormal"					

ROUTE OF DELIVERY *(Check relevant block)*	Pre-labour CS		In-labour CS		Vaginal	
PRESENTATION	Cephalic		Breech			
MODE OF VAGINAL DELIVERY *(Check relevant block)*	Spontaneous		Vacuum		**FAILED VACUUM OR FORCEPS?**	Yes
	Caesarean section		Forceps			No
Reason for vacuum or forceps						
Reason for caesarean section						

DIFFICULTIES WITH DELIVERY	Yes	No	UK
If yes, detail			

Appendix F: Supplementary Data

Table 23: The A3 database design according to the data dictionary: Screening sheet and inclusion/ exclusion criteria form

Variable / Field Name	Form Name	Section Header	Field Type	Field Label	Choices, Calculations, OR Slider Labels
neonate_study_id	neonate_screen_inclusion_and_exclusion_criteria		text	Study ID (Neonate)	
neonate_crf_v_no	neonate_screen_inclusion_and_exclusion_criteria		dropdown	What is the version number of the Neonatal Screening, and Inclusion/Exclusion form?	3, A3 | 4, A4 | 5, A5 | 6, A6 | 7, A7 | 8, A8 | 9, A9 | 10, A10 | 11, A11 | 12, A12 | 13, A13 | 14, A14 | 15, A15 | 16, A16 | 17, A17 | 18, A18 |19, A19 | 20, A20
cooling_hospital	neonate_screen_inclusion_and_exclusion_criteria	Neonatal Screening Sheet	dropdown	Name of hospital providing cooling	1, Groote Schuur | 2, Mowbray Maternity | 3, Tygerberg | 4, Kalafong | 5, Steve Biko | 6, Charlotte Maxeke | 7, Baragwanath | 8, Edendale | 9, Somerset | 0, Other
dob_tob_neonate	neonate_screen_inclusion_and_exclusion_criteria		text	Date and time of birth	
in_outborn_status	neonate_screen_inclusion_and_exclusion_criteria		radio	Was the neonate inborn or outborn?	1, Inborn | 0, Outborn
birth_site_province	neonate_screen_inclusion_and_exclusion_criteria		dropdown	Birth site: Province	0, Northern Cape | 1, Western Cape | 2, Eastern Cape | 3, Mpumalanga | 4, KwaZulu-Natal | 5, Limpopo | 6, Gauteng | 7, North West | 8, Free State
birth_site_sub_district	neonate_screen_inclusion_and_exclusion_criteria		text	Birth Site: Health Sub-district	
birth_site_institution	neonate_screen_inclusion_and_exclusion_criteria		dropdown	Birth site: Name of Institution	1, Groote Schuur | 2, Mowbray Maternity | 3, Tygerberg | 4, Kalafong | 5, Steve Biko | 6, Charlotte Maxeke | 7, Baragwanath | 8, Edendale | 9, Somerset | 0, Other
birth_site_inst_other	neonate_screen_inclusion_and_exclusion_criteria		text	Name of Institution	
mat_home_province_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	Is the province in which the mother gave birth the same as the province in which the mother permanent resides?	
mat_home_province	neonate_screen_inclusion_and_exclusion_criteria		dropdown	Select province in which mother permanently resides?	1, Northern Cape | 2, Western Cape | 3, Eastern Cape | 4, Mpumalanga | 5, KwaZulu-Natal | 6, Limpopo | 7, Gauteng | 8, North West | 9, Free State | 0, Outside of RSA
gest_age_yesno	neonate_screen_inclusion_and_exclusion_criteria	Inclusion Criteria A	yesno	Gestational age 36 weeks or more?	
incl_weight_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	Neonate weight at birth 1800g or more	
ctg_stop_time	neonate_screen_inclusion_and_exclusion_criteria		yesno	Able to start cooling before age 6 hours	

Table 23: The A3 database design according to the data dictionary: Screening sheet and inclusion/ exclusion criteria form

bd16_ph7_yesno	neonate_screen_inclusion_and_exclusion_criteria	Inclusion Criteria B	yesno	BD 1st hour 16 or more/ pH 7 or less	
bd16_ph7_pp_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	BD 1st hour 10 or more / pH 7 or less AND peripartum / sentinel event	
apgar5min_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	Apgar score < 7 at 5 min	
resus_supp_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	Required resus / respiratory support (assisted ventilation) at 10 mins	
incl_c_enceph_signs	neonate_screen_inclusion_and_exclusion_criteria	Inclusion Criteria C: Signs of encephalopathy and other abnormal signs present before cooling	radio	Which of the following signs were observed before cooling?	1, Lethargy | 2, Stupor | 3, Coma | 0, None of these
thompson_yn	neonate_screen_inclusion_and_exclusion_criteria		yesno	Was Thompson score of at least 7 observed?	
incl_c_seiz_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	Were clinical seizures present before cooling?	
clin_seiz_descrip	neonate_screen_inclusion_and_exclusion_criteria		text	Description of Clinical Seizures	
other_abnorm_signs_yn	neonate_screen_inclusion_and_exclusion_criteria		yesno	Were any of the following abnormal signs present before cooling: Hypotonia, Abnormal reflexes, or Abnormal suck?	
other_abnorm_signs	neonate_screen_inclusion_and_exclusion_criteria		checkbox	Which of the following signs were observed before cooling?	0, Hypotonia | 1, Abnormal reflexes | 2, Abnormal suck
aeeg_abnorm_yesno	neonate_screen_inclusion_and_exclusion_criteria	Inclusion Criteria C: amplitude-integrated EEG (aEEG) appearance before or at the onset of cooling	yesno	Was an abnormal aEEG background voltage/pattern observed before the onset or at the time of cooling?	
aeeg_background_descrip	neonate_screen_inclusion_and_exclusion_criteria		radio	aEEG background description	0, Discontinuous normal voltage | 1, Burst suppression | 2, Low Voltage | 3, Flat or Iso-Electic
aeeg_seiz_yesno	neonate_screen_inclusion_and_exclusion_criteria		yesno	Were aEEG seizures observed before the onset of cooling?	
excl_crit_present_yn	neonate_screen_inclusion_and_exclusion_criteria	Exclusion criteria	yesno	Were exclusion criteria present	
excl_crit_descrip	neonate_screen_inclusion_and_exclusion_criteria		checkbox	Select all exclusion criteria present	0, Refused Consent | 1, Severe PPHN | 2, Severe Hypotension | 3, Bleeding that is not responding to treatment | 4, Congenital abnormality | 5, Known chromosomal abnormality | 6, Moribund and unlikely to benefit from cooling | 7, Asystole despite 10 min of resuscitation | 8, Congenital infection | 9, Surgical anomaly | 10, Encephalopathy primarily due to non-hypoxic causes

Table 23: The A3 database design according to the data dictionary: Screening sheet exclusion criteria form

					| 11, Infant died prior to obtaining informed consent | 12, Discontinuation of cooling before 72 hours of treatment (not solely for prognostic reasons) | 13, Other
excl_crit_other	neonate_screen_inclusion_and_exclusion_criteria		text	Please describe "Other" exclusion criteria if applicable	
counselled_yn	neonate_screen_inclusion_and_exclusion_criteria		yesno	Can it be confirmed that counselling was provided prior to seeking informed consent?	
recruit_failure	neonate_screen_inclusion_and_exclusion_criteria	Recruitment status	radio	Was the patient recruited?	0, No
recruited_yes	neonate_screen_inclusion_and_exclusion_criteria		radio	Was the patient recruited?	1, Yes

Table 24: The A3 database design according to the data dictionary: Neonate clinical data

Variable / Field Name	Form Name	Section Header	Field Type	Field Label	Choices, Calculations, OR Slider Labels
gest_age_value	neonatal_history_and_delivery_sheet	Clinical Details of Baby at Birth	text	What was the gestational age at birth?	
method_for_gest_age	neonatal_history_and_delivery_sheet		checkbox	Which method(s) was used to determine the gestational age?	0, Ballard | 1, Footlength | 2, EUS = Sonar at 20 weeks or less | 3, Dates
ballard_score_neonate	neonatal_history_and_delivery_sheet		text	Ballard Score?	
footlength_measure	neonatal_history_and_delivery_sheet		text	Footlength measurement?	
sex_neonate	neonatal_history_and_delivery_sheet		radio	Sex (Neonate)?	0, Male | 1, Female
birth_weight_neonate	neonatal_history_and_delivery_sheet		text	Birth weight (Neonate)	
coh_neonate	neonatal_history_and_delivery_sheet		text	COH (circumference of head)	
apgar_1min	neonatal_history_and_delivery_sheet		text	Apgar score at 1 min	
apgar_5min	neonatal_history_and_delivery_sheet		text	Apgar score at 5 min	
apgar_10min	neonatal_history_and_delivery_sheet		text	Apgar score at 10 min	
bmv_yesno	neonatal_history_and_delivery_sheet		yesno	Bag-mask ventilation (BMV) used?	
nnate_chest_compress_yn	neonatal_history_and_delivery_sheet		yesno	Chest compressions performed?	
nnate_adrenaline_yesno	neonatal_history_and_delivery_sheet		yesno	Adrenaline administered?	
nnate_intub_yesno	neonatal_history_and_delivery_sheet		yesno	Intubation?	
delay_clamp_milking_yn	neonatal_history_and_delivery_sheet		yesno	Delayed cord clamping/cord milking present?	
clamp_milking_diff	neonatal_history_and_delivery_sheet		radio	Indicate whichever option was used	0, Delayed cord clamping | 1, Cord milking | 2, Both delayed cord clamping and cord milking
early_bld_gas_done_yn	neonatal_history_and_delivery_sheet		radio	Are early blood gas results (worst base excess within 60 minutes of birth including cord blood) available?	1, Yes | 0, No | 2, Unrecordable
early_bld_gas_sample_type	neonatal_history_and_delivery_sheet		radio	Source of the early blood gas results?	0, Cord (venous) | 1, Cord (arterial) | 2, Infant
bloodgas_ph_result	neonatal_history_and_delivery_sheet		text	Early blood gas result: pH	
bloodgas_b_excess_result	neonatal_history_and_delivery_sheet		text	Early blood gas result: Base Excess	
bloodgas_lactate_result	neonatal_history_and_delivery_sheet		text	Early blood gas result: Lactate	
cpap_o2_yesno	neonatal_precooling_assessment_and_cooling_method	Clinical Examination Prior to Cooling	yesno	CPAP?	

Table 24: The A3 database design according to the data dictionary: Neonate clinical data

nasal_cann_yn	neonatal_precooling_assessment_and_cooling_method		yesno	Nasal Cannulae?	
extra_o2_yn	neonatal_precooling_assessment_and_cooling_method		yesno	Was additional Oxygen provided?	
mech_ventilation_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Mechanical ventilation?	
hypotens_inotropes_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Hypotension or Inotropes?	
active_bleed_w_trtmnt_yn	neonatal_precooling_assessment_and_cooling_method		yesno	Active bleeding treated with blood products?	
subaponeurotic_haem_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Subaponeurotic haemorrhage?	
sinus_bradycard_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Sinus bradycardia < 80?	
arrhythmia_other_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Other Arrhythmia?	
time_at_neuro_assess	neonatal_precooling_assessment_and_cooling_method	Neurological Assessment Not Earlier Than 30 mins Prior to Cooling	text	Date and time at which neurological assessment was performed?	
age_at_neuro_assess	neonatal_precooling_assessment_and_cooling_method		calc	Age at neurological assessment (same as Thompson score age)	datediff([dob_tob_neonate],[time_at_neuro_assess], "h", "dmy")
visible_seizures_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Visible seizures?	
thompsonscore_baseline	neonatal_precooling_assessment_and_cooling_method		text	Thompson HIE score?	
mod_sarnat_grade_baseline	neonatal_precooling_assessment_and_cooling_method		radio	Modified Sarnat HIE Grade (worst grade prior to cooling)?	0, 1 | 1, 2 | 2, 3
cfm_age_baseline	neonatal_precooling_assessment_and_cooling_method	CFM findings (Worst CFM within first 6 hours of life)	text	Age of the neonate when performing the CFM?	
cfm_voltage_descrip	neonatal_precooling_assessment_and_cooling_method		dropdown	CFM assessment: Voltage?	0, Normal | 1, Moderately abnormal | 2, Suppressed
cfm_pattern_descrip	neonatal_precooling_assessment_and_cooling_method		dropdown	CFM assessment: Pattern?	0, FT | 1, BS | 2, CLV | 3, DNV | 4, CNV
electrical_seizures_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Electrical seizures observed?	
aeeg_model	neonatal_precooling_assessment_and_cooling_method		text	Model used for CFM findings?	
aeeg_measure_type	neonatal_precooling_assessment_and_cooling_method		checkbox	Select which method(s) were used for CFM	0, Needle | 1, Skin patch
aeeg_channel_no	neonatal_precooling_assessment_and_cooling_method		dropdown	No. of channels used for CFM	1, 1 | 2, 2 | 3, 3 | 4, 4 | 5, 5 | 6, 6
date_time_base_labs	neonatal_precooling_assessment_and_cooling_method	Baseline Laboratory Investigation (on	text	Date and Time baseline lab tests were taken?	

Table 24: The A3 database design according to the data dictionary: Neonate clinical data

		admission or as close to cooling as possible; whichever earliest)			
glucose_base_low	neonatal_precooling_assessment_and_cooling_method		text	Lowest blood glucose	
glucose_base_high	neonatal_precooling_assessment_and_cooling_method		text	Highest blood glucose	
poc_sodium_base	neonatal_precooling_assessment_and_cooling_method		text	Point of Care: Baseline Sodium (Na)	
poc_potass_base	neonatal_precooling_assessment_and_cooling_method		text	Point of Care: Baseline Potassium (K)	
ph_base	neonatal_precooling_assessment_and_cooling_method		text	Baseline pH	
pco2_baseline	neonatal_precooling_assessment_and_cooling_method		text	Baseline PCO2	
pco2_measure_method	neonatal_precooling_assessment_and_cooling_method		radio	Baseline PCO2 unit of measure	0, kPa | 1, mmHg
fio2_baseline_value	neonatal_precooling_assessment_and_cooling_method		text	Baseline FiO2 value at time of the blood gas measurement?	
pao2_baseline_value	neonatal_precooling_assessment_and_cooling_method		text	Baseline PaO2 value at time of the blood gas measurement?	
pao2_measure_method	neonatal_precooling_assessment_and_cooling_method		radio	PaO2 unit of measure	0, kPa | 1, mmHg
bicarb_base	neonatal_precooling_assessment_and_cooling_method		text	Baseline Bicarb	
lactate_base	neonatal_precooling_assessment_and_cooling_method		text	Baseline Lactate	
base_excess_base	neonatal_precooling_assessment_and_cooling_method		text	Baseline Base Excess	
poc_icalc_base	neonatal_precooling_assessment_and_cooling_method		text	Point of Care: Ionize Calcium (iCa)	
hb_lab_base	neonatal_precooling_assessment_and_cooling_method		text	Lab results: Hb	
nrbc_lab_base	neonatal_precooling_assessment_and_cooling_method		text	Lab results: Nucleated Red Blood Cells (NRBC)	
tot_wbc_base	neonatal_precooling_assessment_and_cooling_method		text	Lab results: Total White Blood Cells (WBC)	
neutrophil_lab_base	neonatal_precooling_assessment_and_cooling_method		text	Lab results: Neutrophils (Neut)	
platelets_lab_base	neonatal_precooling_assessment_and_cooling_method		text	Lab results: Platelets	
blood_culture_outcome	neonatal_precooling_assessment_and_cooling_method		radio	Baseline blood culture results?	1, Positive | 0, No growth | 2, Not performed
blood_culture_organism	neonatal_precooling_assessment_and_cooling_method		text	For what organism(s) were positive results found?	
date_cooling_admit	neonatal_precooling_assessment_and_cooling_method	Cooling Induction, Method & Setting	text	Date and Time patient ADMITTED to the cooling centre	

Table 24: The A3 database design according to the data dictionary: Neonate clinical data

age_cooling_admit	neonatal_precooling_assessment_and_cooling_method		calc	At what age was the patient ADMITTED to the cooling centre?	datediff([dob_tob_neonate],[date_cooling_admit], "h", "dmy")
date_cooling_commence	neonatal_precooling_assessment_and_cooling_method		text	Date and Time cooling COMMENCED?	
age_cooling_commence	neonatal_precooling_assessment_and_cooling_method		calc	At what age was cooling COMMENCED on the patient?	datediff([dob_tob_neonate],[date_cooling_commence], "h", "dmy")
temp_before_cooling	neonatal_precooling_assessment_and_cooling_method		text	Temperature BEFORE cooling started?	
cooling_method_used	neonatal_precooling_assessment_and_cooling_method		radio	Please select the cooling method used	0, Automated whole body cooling | 1, Servo-controlled gel bag method | 2, Coolcap method | 3, Techotherm (Manual) method | 4, Other
auto_cooling_model	neonatal_precooling_assessment_and_cooling_method		text	Automated whole body cooling: Make and model?	
cool_meth_other_descrip	neonatal_precooling_assessment_and_cooling_method		text	Describe "Other" cooling method used	
date_time_target_temp	neonatal_precooling_assessment_and_cooling_method		text	Date and time target core temp of 33.5 degrees celsius was attained (34.5 degrees for Coolcap)?	
age_at_target_temp	neonatal_precooling_assessment_and_cooling_method		calc	Age when target core temp of 33.5 degrees celsius was attained (34.5 degrees for Coolcap)?	datediff([dob_tob_neonate],[date_time_target_temp], "h", "dmy")
intubation_possible	neonatal_precooling_assessment_and_cooling_method		yesno	Can centre performing cooling offer intubation/IPPV if needed?	
invasive_bp_monitor_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Can centre performing cooling offer invasive blood pressure monitoring?	
inotropes_avail_yesno	neonatal_precooling_assessment_and_cooling_method		yesno	Are inotrope infusions available at the cooling centre?	
care_level	neonatal_precooling_assessment_and_cooling_method		radio	What level of care is the baby nursed in?	0, NICU | 1, High Care
high_care_nicu_move	neonatal_precooling_assessment_and_cooling_method		yesno	If care was initiated in High Care, was the patient moved to NICU at any point in time during care?	
nurse_baby_ratio_avr	neonatal_precooling_assessment_and_cooling_method		text	What is the average number of babies per nurse?	

Table 25: Branching logic in A3 database: Screening and inclusion/ exclusion criteria and Precooling data

Variable / Field Name	Branching Logic
birth_site_sub_district	[birth_site_province] = '0' or [birth_site_province] = '1' or [birth_site_province] = '2' or [birth_site_province] = '3' or [birth_site_province] = '4' or [birth_site_province] = '5' or [birth_site_province] = '6' or [birth_site_province] = '7' or [birth_site_province] = '8'
birth_site_institution	[birth_site_sub_district] <> ""
birth_site_inst_other	[birth_site_institution] = '0'
mat_home_province_yesno	([birth_site_institution] = '0' and [birth_site_inst_other] <> "") or ([birth_site_institution] = '1' or [birth_site_institution] = '2' or [birth_site_institution] = '3' or [birth_site_institution] = '4' or [birth_site_institution] = '5' or [birth_site_institution] = '6' or [birth_site_institution] = '7' or [birth_site_institution] = '8' or [birth_site_institution] = '9')
mat_home_province	[mat_home_province_yesno] = '0'
gest_age_yesno	[mat_home_province_yesno] = '1' or [mat_home_province] <> ""
incl_weight_yesno	[gest_age_yesno] = '1'
ctg_stop_time	[incl_weight_yesno] = '1'
bd16_ph7_yesno	[gest_age_yesno] = '1' and [incl_weight_yesno] = '1' and [cool_within_6hrs_yesno] = '1'
bd16_ph7_pp_yesno	[bd16_ph7_yesno] = '1' or [bd16_ph7_yesno] = '0'
apgar5min_yesno	[bd16_ph7_pp_yesno] = '1' or [bd16_ph7_pp_yesno] = '0'
resus_supp_yesno	[apgar5min_yesno] = '1' or [apgar5min_yesno] = '0'
incl_c_enceph_signs	([bd16_ph7_yesno] = '1' or [bd16_ph7_pp_yesno] = '1' or [apgar5min_yesno] = '1' or [resus_supp_yesno] = '1')
thompson_yn	[incl_c_enceph_signs] <> ""
incl_c_seiz_yesno	[thompson_yn] <> ""
clin_seiz_descrip	[incl_c_seiz_yesno] = '1'
other_abnorm_signs_yn	(([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '1')) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '0')) or (([incl_c_enceph_signs] = '0') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '0')) or (([incl_c_enceph_signs] = '0') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '1')) or (([incl_c_enceph_signs] = '0') and ([incl_c_seiz_yesno] = '0') and ([thompson_yn] = '1')) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([incl_c_seiz_yesno] = '0') and ([thompson_yn] = '1')) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([incl_c_seiz_yesno] = '0') and ([thompson_yn] = '0'))
other_abnorm_signs	[other_abnorm_signs_yn] = '1'

Table 25: Branching logic in A3 database: Screening and inclusion/ exclusion criteria and Precooling data

aeeg_abnorm_yesno	(([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '1') and ((([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')) or ([other_abnorm_signs_yn] = '0'))) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '0') and ((([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')) or ([other_abnorm_signs_yn] = '0'))) or ((([incl_c_enceph_signs] = '0') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '0') and ((([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')) or ([other_abnorm_signs_yn] = '0'))) or ((([incl_c_enceph_signs] = '0') and ([clin_seiz_descrip] <> "") and ([thompson_yn] = '1') and ((([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')) or ([other_abnorm_signs_yn] = '0'))) or ((([incl_c_enceph_signs] = '0') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1'))) and ([thompson_yn] = '1')) or ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1'))) and ([thompson_yn] = '1')) or ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1'))) and ([thompson_yn] = '0'))
aeeg_background_descrip	[aeeg_abnorm_yesno] = '1'
aeeg_seiz_yesno	([aeeg_background_descrip] = '0') or ([aeeg_background_descrip] = '1') or ([aeeg_background_descrip] = '2') or ([aeeg_background_descrip] = '3') or ([aeeg_abnorm_yesno] = '0')
excl_crit_present_yn	([gest_age_yesno] = '1' and [incl_weight_yesno] = '1' and [cool_within_6hrs_yesno] = '1') and ([bd16_ph7_yesno] = '1' or [bd16_ph7_pp_yesno] = '1' or [apgar5min_yesno] = '1' or [resus_supp_yesno] = '1') and ([aeeg_background_descrip] = '0' or [aeeg_background_descrip] = '1' or [aeeg_background_descrip] = '2' or [aeeg_background_descrip] = '3' or [aeeg_seiz_yesno] = '1')
excl_crit_descrip	[excl_crit_present_yn] = '1'
excl_crit_other	[excl_crit_descrip(13)] = '1'
counselled_yn	([excl_crit_present_yn] = '0')
recruit_failure	([gest_age_yesno] = '0' or [incl_weight_yesno] = '0' or [cool_within_6hrs_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd16_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([incl_c_enceph_signs] = '0' and [thompson_yn] = '0' and [incl_c_seiz_yesno] = '0') or ([incl_c_seiz_yesno] = '0' and [other_abnorm_signs_yn] = '0') or ([aeeg_abnorm_yesno] = '0' and [aeeg_seiz_yesno] = '0') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(0)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(1)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(2)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(3)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(4)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(5)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(6)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(7)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(8)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(9)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(10)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(11)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(12)] = '1') or ([excl_crit_present_yn] = '1' and [excl_crit_descrip(13)] = '1' and [excl_crit_other] <> "") or ([counselled_yn] = '0')
recruited_yes	([gest_age_yesno] = '1' and [incl_weight_yesno] = '1' and [cool_within_6hrs_yesno] = '1') and ([bd16_ph7_yesno] = '1' or [bd16_ph7_pp_yesno] = '1' or [apgar5min_yesno] = '1' or [resus_supp_yesno] = '1') and (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3' or [thompson_yn] = '1' or [incl_c_seiz_yesno] = '1') or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3' or [thompson_yn] = '1') and ([incl_c_seiz_yesno] = '0' and ([other_abnorm_signs(0)] = '1' or [other_abnorm_signs(1)] = '1' or [other_abnorm_signs(2)] = '1')))) and ([aeeg_abnorm_yesno] = '1' or [aeeg_seiz_yesno] = '1') and ([excl_crit_present_yn] = '0') and ([counselled_yn] = '1')

Table 26: Example of the advancement of branching logic from the A3 database to the A4/A5 database

Variable name	Branching logic
recruited_yes **(A3 database)**	([gest_age_yesno] = '1' and [incl_weight_yesno] = '1' and [cool_within_6hrs_yesno] = '1') and ([bd16_ph7_yesno] = '1' or [bd16_ph7_pp_yesno] = '1' or [apgar5min_yesno] = '1' or [resus_supp_yesno] = '1') and (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3' or [thompson_yn] = '1' or [incl_c_seiz_yesno] = '1') or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3' or [thompson_yn] = '1') and ([incl_c_seiz_yesno] = '0' and ([other_abnorm_signs(0)] = '1' or [other_abnorm_signs(1)] = '1' or [other_abnorm_signs(2)] = '1')))) and ([aeeg_abnorm_yesno] = '1' or [aeeg_seiz_yesno] = '1') and ([excl_crit_present_yn] = '0') and ([counselled_yn] = '1')
recruited_yes **(A4/A5 database)**	((([gest_age_yesno] = '1' and [incl_weight_yesno] = '1') and (([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1')) and ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '1') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '0') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or (([incl_c_enceph_signs] = '0') and ([thompson_yn] = '0') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or (([incl_c_enceph_signs] = '0') and ([thompson_yn] = '1') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or ((([incl_c_enceph_signs] = '0') and ([thompson_yn] = '1') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs_yn] = '1') and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1'))))) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '1') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs_yn] = '1') and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1'))))) or (([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '0') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs_yn] = '1') and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))))) and ([excl_crit(0)] = '1') and ([counselled_yn] = '1')) or ((([gest_age_yesno] = '1' and [incl_weight_yesno] = '1') and (([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '1') or ([bd16_ph7_yesno] = '1' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '1' and [apgar5min_yesno] = '0' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '0' and [apgar5min_yesno] = '1' and [resus_supp_yesno] = '0') or ([bd16_ph7_yesno] = '0' and [bd10_ph7_pp_yesno] = '0' and

Table 26: Example of the advancement of branching logic from the A3 database to the A4/A5 database

	[apgar5min_yesno] = '0' and [resus_supp_yesno] = '1')) and ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '1') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '0') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or ((([incl_c_enceph_signs] = '0') and ([thompson_yn] = '0') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or ((([incl_c_enceph_signs] = '0') and ([thompson_yn] = '1') and ([clin_seiz_descrip] <> "") and (([other_abnorm_signs_yn] = '0') or (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))) or ((([incl_c_enceph_signs] = '0') and ([thompson_yn] = '1') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs_yn] = '1') and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1'))))) or ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '1') and ([incl_c_seiz_yesno] = '0' and ((([other_abnorm_signs_yn] = '1') and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))))) or ((([incl_c_enceph_signs] = '1' or [incl_c_enceph_signs] = '2' or [incl_c_enceph_signs] = '3') and ([thompson_yn] = '0') and ([incl_c_seiz_yesno] = '0' and (([other_abnorm_signs_yn] = '1') and (([other_abnorm_signs(0)] = '1') or ([other_abnorm_signs(1)] = '1') or ([other_abnorm_signs(2)] = '1')))))) and ([excl_crit_a5(0)] = '1') and ([counselled_yn_a5] = '1')))

Table 27: The A7 database design according to the data dictionary: Screening sheet and inclusion/ exclusion criteria form

Variable / Field Name	Form Name	Section Header	Field Type	Field Label	Choices, Calculations, OR Slider Labels
please_complete_1	neonatal_screening_sheet		descriptive	<font color="navy"> <h3> Amendment 7 Documents are now in use! <h3 style="text-align:center;"> Form must be completed in full in order to proceed to Neonatal Inclusion/Exclusion Form. </h3>	
cooling_hospital	neonatal_screening_sheet	<font size=3> Neonatal Screening Information	dropdown	Name of hospital providing cooling	1, Groote Schuur | 2, Mowbray Maternity | 3, Tygerberg | 4, Kalafong | 5, Steve Biko | 6, Charlotte Maxeke | 7, Baragwanath | 8, Edendale | 9, Somerset | 0, Other
birth_site_province	neonatal_screening_sheet	<font size=3> Birth Site Information	dropdown	Province	0, Northern Cape | 1, Western Cape | 2, Eastern Cape | 3, Mpumalanga | 4, KwaZulu-Natal | 5, Limpopo | 6, Gauteng | 7, North West | 8, Free State
cov_mat_test_stat	neonatal_screening_sheet	<font size=3> COVID-19 Status: Mother	dropdown	Mother: Tested for COVID-19	1, Yes | 0, No | 99, Unknown | 2, N/A - A3/A5 data
cov_mat_test_date	neonatal_screening_sheet		text	Date tested: Mother	
cov_mat_samp_type	neonatal_screening_sheet		dropdown	Mother: Type of sample collected	1, Swab: Nasopharangeal | 2, Swab: Buccal | 3, Blood | 99, Unknown
cov_mat_test_meth	neonatal_screening_sheet		dropdown	Mother: Testing method for sample	1, PCR | 2, Antibody | 99, Unknown
cov_mat_inf_stat	neonatal_screening_sheet		dropdown	Mother: Positive for COVID-19	1, Yes | 0, No | 99, Unknown
cov_mat_sev_stat	neonatal_screening_sheet		dropdown	Mother: Infection severity	1, Asymptomatic | 2, Mild | 3, Moderate | 4, Severe | 99, Unknown
cov_mat_other_details	neonatal_screening_sheet		notes	Indicate any other details RE Maternal COVID status	
cov_neo_test_stat	neonatal_screening_sheet	<font size=3> COVID-19 Status: Neonate	dropdown	Neonate: Tested for COVID-19	1, Yes | 0, No | 99, Unknown | 2, N/A - A3/A5 data
cov_neo_test_date	neonatal_screening_sheet		text	Date tested: Neonate	

Table 27: The A7 database design according to the data dictionary: Screening sheet and inclusion/ exclusion criteria form

cov_neo_samp_type	neonatal_screening_sheet		dropdown	Neonate: Type of sample collected	1, Swab: Nasopharangeal | 2, Swab: Buccal | 3, Blood | 99, Unknown
cov_neo_test_meth	neonatal_screening_sheet		dropdown	Neonate: Testing method for sample	1, PCR | 2, Antibody | 99, Unknown
cov_neo_inf_stat	neonatal_screening_sheet		dropdown	Neonate: Positive for COVID-19	1, Yes | 0, No | 99, Unknown
cov_neo_sev_stat	neonatal_screening_sheet		dropdown	Neonate: Infection severity	1, Asymptomatic | 2, Mild | 3, Moderate | 4, Severe | 99, Unknown
cov_neo_other_details	neonatal_screening_sheet		notes	Indicate any other details RE Neonatal COVID status	
please_proceed_cov1	neonatal_screening_sheet		descriptive	<font color="black"> Please proceed.	
complete_only_abc_met	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color= "navy"> <h3> Amendment 7 Documents are now in use! <h3 style="text-align:center;"> Complete only if category A, B and C conditions are met according to source.</h3>	
gest_age_yesno	neonatal_inclusion_and_exclusion_criteria	<font size=3> Inclusion Criteria A	yesno	Gestational age 36 weeks or more?	
screen_cond_not_met_1	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color="black"> Screening conditions not met. Remove data.	
bd16_ph7_yesno	neonatal_inclusion_and_exclusion_criteria	<font size=3> Inclusion Criteria B	yesno	BD 1st hour 16 or more / pH 7 or less	
screen_cond_not_met_2	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color="black"> Screening conditions not met. Remove data.	

Table 27: The A7 database design according to the data dictionary: Screening sheet and inclusion/ exclusion criteria form

incl_c_enceph_signs	neonatal_inclusion_and_exclusion_criteria	<font size=3> Inclusion Criteria C1: Sign(s) of Encephalopathy Present Before Cooling	radio	Which of the following signs were observed before cooling?	1, Lethargy | 2, Stupor | 3, Coma | 0, None of these
screen_cond_not_met_3	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color="black"> Screening conditions not met. Remove data.	
other_abnorm_signs_yn	neonatal_inclusion_and_exclusion_criteria	<font size=3> Inclusion Criteria C2: Other Abnormal Sign(s) Present Before Cooling	yesno	Were any of the following abnormal signs present before cooling: Hypotonia, Abnormal reflexes, or Abnormal suck?	
screen_cond_not_met_4	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color="black"> Screening conditions not met. Remove data.	
excl_crit	neonatal_inclusion_and_exclusion_criteria		checkbox	Select all exclusion criteria present	0, None | 1, Neonate not cooled within 6 hours of life | 2, Moribund and unlikely to benefit from cooling | 3, Asystole | 4, Hypotension or bleeding that is not responding to treatment | 5, Severe PPHN | 6, Suspected chromosomal abnormality | 7, Congenital infection | 8, Surgical anomaly | 9, Consent refused | 10, Neonate died prior to obtaining informed consent | 11, Consent not obtained for other reason | 12, Encephalopathy primarily due to non-hypoxic cause | 13, Congenital abnormality | 14, Neonate not cooled | 15, Other
counselled_yn	neonatal_inclusion_and_exclusion_criteria	<font size=3> Criteria E: Patient Counselling	yesno	Can it be confirmed that counselling was provided prior to seeking informed consent?	

Table 27: The A7 database design according to the data dictionary: Screening sheet and inclusion/ exclusion criteria form

recruit_failure	neonatal_inclusion_and_exclusion_criteria	<font size=3> Criteria F: Recruitment Status	yesno	Was the patient recruited?	
please_upload_2	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color="black"> Please upload the Screening Sheet and Inclusion & Exclusion Form to LogicalDOC.	
please_upload_2_b	neonatal_inclusion_and_exclusion_criteria		descriptive	<font color="black"> Please upload the Screening Sheet and Inclusion & Exclusion Form to LogicalDOC.	

Table 28: The A7 database design according to the data dictionary: Neonate clinical data

Variable / Field Name	Form Name	Section Header	Field Type	Field Label	Choices, Calculations, OR Slider Labels
no_data_upload_1	neonatal_precooling_assessment_and_cooling_method		descriptive	Patient not recruited. No data to upload.	
picf_no_cond_html_pre	neonatal_precooling_assessment_and_cooling_method		descriptive	<font color="black"> Informed Consent conditions not met or completed incorrectly! Correct before proceeding.	
no_recontact_3	neonatal_precooling_assessment_and_cooling_method		descriptive	<font color="black"> This form will not function unless permission for recontact was provided.	
method_for_gest_age	neonatal_precooling_assessment_and_cooling_method	<font size=3> Pre-Cooling Assessment and Cooling Method	checkbox	Method used for GA	3, Dates | 2, EUS < 20 weeks | 0, Ballard | 1, Footlength
blood_gas_eval_birth	neonatal_precooling_assessment_and_cooling_method	<font size=3>Blood Gas Evaluation: At Birth	radio	Blood Gas Evaluation: At birth - Done and available?	1, Yes | 0, No | 2, Not done
date_base_labs	neonatal_precooling_assessment_and_cooling_method		text	Date and Time taken	
early_bld_gas_sample_type	neonatal_precooling_assessment_and_cooling_method		radio	Source: Blood gas	0, Cord (venous) | 1, Cord (arterial) | 2, Cord (mixed source) | 3, Infant (ABG)
no_data_upload_1	neonatal_precooling_assessment_and_cooling_method		descriptive	Patient not recruited. No data to upload.	
hb_lab_base	neonatal_precooling_assessment_and_cooling_method	<font size=3> Laboratory Investigations: At Birth	text	Lab: Hb	
platelets_lab_base	neonatal_precooling_assessment_and_cooling_method		text	Lab: Platelets	
blood_culture_outcome	neonatal_precooling_assessment_and_cooling_method	<font size=3> Blood Culture Investigations: At Birth	radio	Blood culture result for culture taken on this day	1, Positive | 0, No growth | 2, Not performed
blood_culture_organism	neonatal_precooling_assessment_and_cooling_method		text	Organism(s) if positive:	

Table 28: The A7 database design according to the data dictionary: Neonate clinical data

cpap_o2_yesno	neonatal_precooling_assessment_and_cooling_method	<font size=3>Clinical Examination Prior to Cooling	radio	CPAP	1, Yes | 0, No
aeeg_done_time	neonatal_precooling_assessment_and_cooling_method	<font size=3>CFM Details	checkbox	Was an aEEG evaluation performed? If yes, please indicate the time points.	0, Not done |1, Pre-Cooling/ Cooling |2, 6 Hours |3, 24 Hours |4, 48 Hours |5, 72 Hours | 6, After rewarm
date_at_neuro_assess	neonatal_precooling_assessment_and_cooling_method	<font size=3>Neurological Assessment Prior to Cooling (Not Earlier than 30 mins) or at Onset of Cooling	text	Date and Time taken	
precool_aeeg_time	neonatal_precooling_assessment_and_cooling_method	<font size=3>Pre-cooling or Cooling aEEG Assessment	text	Time of Assessment	
aeeg_time_6h	neonatal_precooling_assessment_and_cooling_method	<font size=3> aEEG Assessment: 6 Hours of Life	text	Time of Assessment	
base_lab_inves_close_cool	neonatal_precooling_assessment_and_cooling_method	<font size=3>Baseline Laboratory Investigation: On Admission or as Close to Cooling as Possible	radio	Were these evaluations done and are the results different from the blood gas at birth?	1, Done, Results different | 0, Done, Results not different |2, Not Done
hb_lab_base_cool	neonatal_precooling_assessment_and_cooling_method	<font size=3> Laboratory Investigations: On Admission or as Close to Cooling as Possible	text	Lab: Hb	
blood_culture_out_cool	neonatal_precooling_assessment_and_cooling_method	<font size=3> Blood Culture Investigations: On Admission or as Close to Cooling as Possible	radio	Blood culture result for culture taken on this day	1, Positive | 0, No growth | 2, Not performed
date_cooling_admit	neonatal_precooling_assessment_and_cooling_method	<font size=3>Cooling Induction, Method & Setting	text	Date and Time admitted to cooling centre	
age_cooling_admit	neonatal_precooling_assessment_and_cooling_method		calc	At what age was the patient ADMITTED to the cooling centre?	datediff([dob_tob_neonate],[date_cooling_admit], "h", "dmy", true)
date_cooling_commence	neonatal_precooling_assessment_and_cooling_method		text	Date and Time cooling commenced	

Table 28: The A7 database design according to the data dictionary: Neonate clinical data

age_cooling_commence	neonatal_precooling_assessment_and_cooling_method		calc	At what age did cooling COMMENCE?	datediff([dob_tob_neonate],[date_cooling_commence], "h", "dmy", true)
cooling_method_used	neonatal_precooling_assessment_and_cooling_method		dropdown	Cooling method used	0, Automated whole body cooling | 1, Servo-controlled gel bag method | 2, Coolcap method | 3, Manual Techotherm | 4, Other
age_at_target_temp	neonatal_precooling_assessment_and_cooling_method		calc	Age when target core temp was attained?	datediff([dob_tob_neonate],[date_target_temp], "h", "dmy", true)

Table 29: An overview of the values and proportions of error codes across Instruments

	A3						A4						A5						A7					
	'0' (%)	'1' (%)	'555' (%)	'777' (%)	'888' (%)	'999 (%)'	'0' (%)	'1' (%)	'555' (%)	'777' (%)	'888' (%)	'999' (%)	'0' (%)	'1' (%)	'555' (%)	'777' (%)	'888' (%)	'999' (%)	'0' (%)	'1' (%)	'555' (%)	'777' (%)	'888' (%)	'999' (%)
	Site 1																							
Screen	3 (2.8)	84 (77.8)	0	21 (19.4)	0	0	8 (9.5)	58 (69)	1 (1.2)	45 (53.6)	0	0	2 (5.6)	27 (75)	0	7 (19.4)	0	0	3 (2.3)	106 (80.3)	1 (0.8)	22 (16.7)	0	0
Inclusion / Exclusion	2 (1.1)	112 (62.2)	5 (2.8)	61 (33.9)	0	0	4 (2.7)	97 (66)	1 (0.7)	45 (30.6)	0	0	0 (1.3)	41 (68.3)	0	19 (31.7)	0	0	3 (1.4)	143 (65)	2 (0.9)	72 (32.7)	0	0
Precool	22 (2.2)	536 (53.2)	51 (5.1)	376 (37.3)	5 (0.5)	18 (1.8)	26 (3.3)	526 (67)	48 (6)	136 (17.3)	2 (0.3)	46 (5.9)	2 (0.6)	229 (68.2)	14 (4.2)	42 (12.5)	2 (0.6)	47 (14)	12 (1)	964 (78.2)	32 (2.6)	102 (8.3)	15 (1.2)	107 (8.7)
	Site 2																							
Screen	1 (0.9)	73 (67.6)	8 (7.4)	26 (24.1)	0	0	10 (8.3)	84 (70)	0	65 (54.2)	0	0	5 (6.9)	50 (69.4)	0	17 (23.6)	0	0	1 (1.2)	63 (75)	1 (1.2)	19 (22.6)	0	0
Inclusion /Exclusion	0	117 (61.9)	9 (4.8)	63 (33.3)	0	0	0	145 (69)	0	65 (31)	0	0	1 (0.8)	84 (66.7)	1 (0.8)	40 (31.7)	0	0	0	100 (68)	0	47 (32)	0	0
Precool	12 (1.2)	587 (58.2)	43 (4.3)	359 (35.6)	6 (0.6)	1 (0.1)	45 (4.0)	648 (57.9)	54 (4.8)	166 (14.8)	16 (1.4)	191 (17.1)	9 (1.3)	239 (35.6)	31 (4.6)	88 (13.1)	2 (0.3)	50 (7.4)	9 (1.1)	604 (77)	22 (2.8)	77 (9.8)	1 (0.1)	71 (9.1)
	Site 3																							
Screen	0	79 (82.3)	0	17 (17.7)	0	0	0	70 (83.3)	0	44 (52.4)	0	0	3 (4.2)	56 (77.8)	0	13 (18.1)	0	0	0	39 (81.3)	0	9 (18.8)	0	0
Inclusion /Exclusion	0	112 (66.7)	5 (3)	51 (30.4)	0	0	3 (2.0)	95 (64.6)	5 (3.4)	44 (29.9)	0	0	0	85 (67.5)	0	41 (32.5)	0	0	0	57 (67.9)	0	27 (32.1)	0	0
Precool	11	556	11	295	19	0	27	521	25	127	36	48	9	434	17	77	11	124	0	345	2	49	26	26

Table 29: An overview of the values and proportions of error codes across Instruments

	(1.2)	*(62.1)*	*(1.2)*	*(32.9)*	*(2.1)*		*(3.4)*	*(66.5)*	*(3.2)*	*(16.2)*	*(4.6)*	*(6.1)*	*(1.3)*	*(64.6)*	*(2.5)*	*(11.5)*	*(1.6)*	*(18.5)*		*(77)*	*(0.4)*	*(10.9)*	*(5.8)*	*(5.8)*
	'0'	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*	*'0'*	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*	*'0'*	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*	*'0'*	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*
	Site 4																							
Screen	*0*	*51 (70.8)*	*4 (5.6)*	*17 (23.6)*	*0*	*0*	-	-	-	-	-	-	*4 (4.8)*	*63 (75)*	*1 (1.2)*	*16 (19.0)*	*0*	*0*	*1 (0.9)*	*87 (80.6)*	*1 (0.9)*	*19 (17.6)*	*0*	*0*
Inclusion /Exclusion	*1 (0.8)*	*67 (53.2)*	*18 (14.3)*	*40 (31.7)*	*0*	*0*	-	-	-	-	-	-	*4 (2.7)*	*92 (62.6)*	*6 (4.1)*	*45 (30.6)*	*0*	*0*	*0*	*125 (66.1)*	*4 (2.1)*	*60 (31.7)*	*0*	*0*
Precool	*29 (4.3)*	*305 (45.4)*	*54 (8)*	*248 (36.9)*	*34 (5.1)*	*2 (0.3)*	-	-	-	-	-	-	*7 (0.9)*	*430 (54.8)*	*72 (9.2)*	*88 (11.2)*	*44 (5.6)*	*143 (18.2)*	*21 (2.1)*	*461 (45.7)*	*105 (10.4)*	*62 (6.2)*	*176 (17.5)*	*182 (18.1)*
	Site 5																							
Screen	*0*	*19 (79.2)*	*0*	*5 (20.8)*	*0*	*0*	*0*	*30 (83.3)*	*0*	*19 (16.7)*	*0*	*0*	*4 (11.1)*	*24 (66.7)*	*0*	*8 (22.2)*	*0*	*0*	*11 (10.2)*	*77 (71.3)*	*0*	*20 (18.5)*	*0*	*0*
Inclusion /Exclusion	*0*	*28 (66.7)*	*0*	*14 (33.3)*	*0*	*0*	*0*	*43 (68.3)*	*1 (1.6)*	*19 (30.2)*	*0*	*0*	*0*	*40 (63.5)*	*2 (3.2)*	*21 (33.3)*	*0*	*0*	*2 (1.1)*	*122 (64.6)*	*4 (2.1)*	*61 (32.3)*	*0*	*0*
Precool	*33 (14.7)*	*115 (51.3)*	*6 (2.7)*	*70 (31.3)*	*0*	*0*	*1 (0.3)*	*262 (78)*	*16 (4.8)*	*29 (8.6)*	*8 (2.4)*	*14 (4.2)*	*19 (15.1)*	*82 (65.1)*	*1 (0.8)*	*41 (32.5)*	*0*	*0*	*24 (2.4)*	*698 (69.2)*	*25 (2.5)*	*55 (5.5)*	*35 (3.5)*	*152 (15.1)*
	Site 6																							
Screen	*1 (4.2)*	*16 (66.7)*	*2 (8.3)*	*5 (20.8)*	*0*	*0*	*0*	*21 (87.5)*	*0*	*3 (12.5)*	*0*	*0*	*3 (4.2)*	*54 (75)*	*0*	*15 (20.8)*	*0*	*0*	-	-	-	-	-	-
Inclusion /Exclusion	*0*	*27 (64.3)*	*2 (4.8)*	*13 (31)*	*0*	*0*	*0*	*29 (69)*	*0*	*13 (31)*	*0*	*0*	*2 (1.6)*	*79 (62.7)*	*3 (2.4)*	*41 (32.5)*	*1 (0.8)*	*0*	-	-	-	-	-	-

Table 29: An overview of the values and proportions of error codes across Instruments

Precool	*6 (2.7)*	*117 (52.2)*	*12 (5.4)*	*75 (33.5)*	*14 (6.3)*	*0*	*7 (3.1)*	*172 (76.8)*	*10 (4.5)*	*24 (10.7)*	*10 (4.5)*	*0*	*25 (3.7)*	*453 (67.6)*	*11 (1.6)*	*53 (7.9)*	*86 (12.8)*	*42 (6.3)*	-	-	-	-	-	-
	'0'	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*	*'0'*	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*	*'0'*	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*	*'0'*	*'1'*	*'555'*	*'777'*	*'888'*	*'999'*
	Site 7																							
Screen	*0*	*72 (85.7)*	*0*	*12 (14.3)*	*0*	*0*	*1 (1.7)*	*47 (78.3)*	*0*	*12 (20)*	*0*	*0*	*0*	*54 (75)*	*0*	*18 (25)*	*0*	*0*	*8 (9.5)*	*57 (67.9)*	*0*	*19 (22.6)*	*0*	*0*
Inclusion /Exclusion	*3 (2.0)*	*100 (68.0)*	*1 (0.7)*	*43 (29.3)*	*0*	*0*	*2 (1.9)*	*70 (66.7)*	*3 (2.9)*	*30 (28.6)*	*0*	*0*	*3 (2.4)*	*87 (69.0)*	*0*	*36 (28.6)*	*0*	*0*	*6 (4.1)*	*92 (62.6)*	*2 (1.4)*	*47 (32)*	*0*	*0*
Precool	*25 (3.2)*	*579 (73.9)*	*14 (1.8)*	*136 (17.3)*	*3 (0.4)*	*0*	*11 (2.0)*	*423 (75.5)*	*39 (7.0)*	*36 (6.4)*	*27 (4.8)*	*0*	*26 (3.9)*	*520 (77.4)*	*39 (5.8)*	*27 (4.0)*	*4 (0.6)*	*20 (3.0)*	*25 (3.2)*	*629 (80.2)*	*20 (2.6)*	*72 (9.2)*	*8 (1)*	*28 (3.6)*