Developing Neuroimaging Methods for Clinical Translation and Better Understanding Neonatal Brain Development

Marianne Aspbury

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

August 2023

Abstract

Understanding and measuring pain and brain development in neonates is essential to be able to provide the best care for this vulnerable population. This is particularly important for premature infants, for whom early life is filled with more painful procedures, and earlier exposure to extrauterine stimuli, which can adversely affect development. Infant pain assessments combine behavioural and physiological measures such as facial expression, crying, and heart rate. However, these metrics are not specific to pain experience, nor sensitive enough to provide reliable outcome measures for clinical trials to validate pain treatments in infants. Neuroimaging techniques provide means to study brain health, development and function. EEG and fMRI measurements of noxious-evoked brain activity could be used to develop more objective and specific pain assessment tools. This thesis focusses on using EEG and MRI to measure infant pain and its relation to overall brain development.

First, I present tests of the validity of an EEG template measure of noxious response in infants recruited at multiple hospital sites. EEG has been used to quantify noxious-evoked activity and study pain interventions in infants, but a standard generalisable approach needs to be established. I tested whether the EEG template discriminates between noxious and non-noxious stimuli, whether the scale of noxious response is equivalent across different hospital sites, and whether noxious response increases with age in premature infants. I found that noxious-evoked responses are significantly greater than non-noxious responses, but that the scale is not equivalent across study sites, and there was no significant age correlation. This suggests that the EEG template can be reliably used as a surrogate measure of pain, with promise for clinical trials. Additionally, data collection site should be accounted for as a confounding factor as needed.

Then, I focus on how MRI can aid our understanding of infant pain and the underlying neurophysiology behind differences in noxious-evoked activity. I present a machine learning model that I developed to predict the magnitude of noxious-evoked responses from resting-state brain activity in infants, using fMRI data. By applying this model to data from the independent Developing Human Connectome Project, I explore how predicted noxious-evoked responses relate to development metrics, including resting-state cortical function and microstructure, as well as prematurity, and assessments of infant cognitive and motor ability at 2-year follow up. I found that prematurity is associated with accelerated development of the nociceptive system, but disrupted neurodevelopment overall.

In summary, this thesis demonstrates the potential for neuroimaging techniques to improve our understanding of infant brain development, and improve clinical assessment and treatment of infant pain.

Acknowledgements

First and foremost, this thesis is dedicated to my Dad, Peter Aspbury. A caring, inquisitive, and intelligent man who was always interested in how my research and my thesis were progressing. He was so eager to read my final thesis, and I promised to print him his own bound copy. Now, I will keep his copy safe.

This thesis would not have been possible without a great deal of people supporting me professionally, personally, and often in both ways. Thank you to my supervisors Rebeccah Slater and Eugene Duff. I am genuinely so grateful for your expert insights and your resilient positivity and support, especially through the unexpected and unprecedented disruption caused by a pandemic! It has been an epic journey working and learning with both of you.

Thank you to my colleagues in the Paediatric Neuroimaging Group, particularly to Luke and Marianne, with whom I've worked most closely. Your knowledge, patience, and kindness has helped me greatly. A special thanks also to Simon, for humouring me with co-working days in the cafés of Oxford, where I made some of my best writing progress.

Thanks, of course, to my partner, Georgie, who has somehow managed to keep her energy and joy by my side, and support me through it all. Thank you to my mum, and to my sisters, who have inspired and supported me every day, you are simply the best people. To the rest of my family, and to my wonderful friends from every city and every corner of my life, you have been amazing and I could not have done it without you.

Not only is it mentioned at points throughout this thesis, but the thesis writing process taught me first-hand; stress is indeed bad for you. I experienced an usual number of adverse events in the final months of writing; I suspect from increased cortisol (although this was not scientifically verified), including a surprise attack of shingles, a first ear infection, and an increased tendency to walk into doors. For this, I am thankful to my mum for identifying the shingles, to the NHS, both for their excellent 111 service and my GP, who quickly prescribed me with remedial medication, and to my partner for keeping me as healthy as possible!

Finally, thanks to everyone who made this research possible: to the participating families, the trained staff involved in recruitment and recording of high quality data, and to the SABS CDT, EPSRC and Reckitt for funding my DPhil.

May the Thesis be over, but the Chapters continue.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACTTION (PPRC) . . Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (Paediatric Pain Research Consortium)

AI . . . . . Artificial Intelligence

AIPE . . . . . Accuracy in Parameter Estimation

AN (l / r) . . . . . (left / right) Auditory Network

atr . . . . . Anterior thalamic radiations

BOLD . . . . . Blood-Oxygen Level-Dependent

BSID (-I, -II, -III) . . . Bayley Scores of Infant Development (version I, II, or III)

CL . . . . . Control Heel Lance

CI . . . . . Confidence Interval

COPE . . . . . Contrast Of Parameter Estimates

COSMIN . . . . . COnsensus-based Standards for the Selection of Health Measurement INstruments

CSF . . . . . Cerebrospinal Fluid

cst . . . . . Corticospinal tracts

CVCR . . . . . Cross-Validated Confound Regression

CV-LOO . . . . . Cross-Validated Leave-One-Out

DAN . . . . . Dorsal Attention Network

dHCP . . . . . Developing Human Connectome Project

DOI . . . . . Digital Object Identifier

DMN . . . . . Default Mode Network

dMRI . . . . . Diffusion Magnetic Resonance Spectroscopy

EEG . . . . . Electroencephalography

ECN . . . . . Executive Control Network

EMA . . . . . European Medicines Agency

EPIPPAIN . . . . . Epidemiology of Procedural Pain in Neonates

ERP . . . . . Event-Related Potential

EUROPPAIN . . . . . EUROpean Pain Audit In Neonates

FA . . . . . Fractional Anisotropy

FC . . . . . Functional Connectivity

FDA . . . . . US Food and Drug Administration

fmi . . . . . Forceps minor

fMRI . . . . . Functional Magnetic Resonance Imaging

FMRIB . . . . . Centre for Functional Magnetic Resonance Imaging of the Brain

fNIRS . . . . . Functional Near Infra-Red Spectroscopy

FSL . . . . . FMRIB Software Library

GA . . . . . Gestational Age

GLM . . . . . General Linear Model

HARKing . . . . . Hypothesising After Results are Known

HL . . . . . Heel Lance

HIE . . . . . Hypoxic-Ischaemic Encephalopathy

HRF . . . . . Haemodynamic Response Function

IASP . . . . . International Association for the Study of Pain

ICA . . . . . Independent Component Analysis

IVH . . . . . Intraventricular Haemorrhage

MAD . . . . . Median Absolute Deviation

MD . . . . . Mean Diffusivity

MDI . . . . . Mental Development Index

M_e . . . . . Effective number of tests

MEG . . . . . Magnetoencephalography

ML . . . . . Machine Learning

MRI . . . . . Magnetic Resonance Imaging

MSE . . . . . Mean Squared Error

NFCS . . . . . Neonatal Facial Coding System

NICU . . . . . Neonatal Intensive Care Unit

NIRS . . . . . Near Infra-Red Spectroscopy

NPS . . . . . Neural Pain Signature

NRMSE . . . . . Normalised Root Mean Squared Error

NSAID . . . . . Non-Steroidal Anti-Inflammatory Drug

OLS . . . . . Ordinary Least Squares

OSF . . . . . Open Science Framework

PC . . . . . Principal Component

PCA . . . . . Principal Component Analysis

PCI-RR . . . . . Peer Community in Registered Reports

PD . . . . . Pharmacodynamics

PIL . . . . . Patient Information Leaflet

PIPP(-R) . . . . . Premature Infant Pain Profile (-Revised)

PK . . . . . Pharmacokinetics

PMA . . . . . Postmenstrual Age

PNA . . . . . Postnatal Age

Poppi . . . . . Procedural Pain in Premature Infants (Clinical Trial)

PROFUMO . . . . . PRObabilistic FUnctional MOdes

RCT . . . . . Randomised Controlled Trial

REC . . . . . Research Ethics Committee

ROI . . . . . Region of Interest

ROP . . . . . Retinopathy Of Prematurity

RMSE . . . . . Root Mean Squared Error

$\boldsymbol{R_{rm}}$ . . . . . Repeated measures correlation

rs-fMRI . . . . . Resting-State Functional Magnetic Resonance Imaging

RSN . . . . . Resting State Network

SMN . . . . . Somatomotor Network

SNR . . . . . Signal to Noise Ratio

SOP . . . . . Standard Operating Procedure

SRT . . . . . Stimulus Response Template

stim-fMRI . . . . . Stimulus-Evoked Functional Magnetic Resonance Imaging

str . . . . . Superior thalamic radiations

SVR . . . . . Support Vector Regression

task-fMRI . . . . . Task-Activation Functional Magnetic Resonance Imaging

TEA . . . . . Term-Equivalent Age

unc . . . . . Uncinate fasciculi

VEP . . . . . Visual-Evoked Potential

VNm . . . . . Medial Visual Network

VNop . . . . . Occipital pole Visual Network

WHO . . . . . World Health Organisation

WM . . . . . White Matter

1 Introduction

1.1 Research Aims

This thesis aims to explore and address open issues in the fields of neonatal and infant pain treatment, health and brain development. I aimed to develop and validate methods using non-invasive neuroimaging techniques to quantify infant noxious-evoked brain activity, as a surrogate measure of pain. I implemented best methodological and research practices in order to do so. I also aimed to use these neuroimaging methods to study how noxious sensitivity relates to indicators of health in infants, such as prematurity and long-term developmental outcomes.

1.2 An Overview of Infant Age Terminology

There are various descriptors of the early life period, which it will be helpful for the reader to understand. The broadest term is that of 'infant' which is typically used to describe the first 0-2 years of life since birth, although the upper age limit is not strict, as reflected in the use of infant by large global projects and organisations concerned with early life development including INTERGROWTH-21st [1], BRIGHT [2], and the World Health Organisation (WHO) [3]. The terms 'neonate', 'newborn' or 'newborn infant' are all used to describe infants within the first 28 days of life since

birth [4, 5]. Infants fall into further categories based on their gestational age (GA), or age at birth, where the gestation is measured from the first day since the biological mother's last menstrual period, which can be assessed by ultrasound or clinical assessment [4, 6, 7]. 'Term' infants are those born from 37 completed weeks to less than 42 completed weeks gestation, whilst 'preterm' infants are those born at less than 37 completed weeks gestation, and 'post-term' infants are born at 42 completed weeks gestation or more [6]. These terms exist to describe the risks associated with post-term and preterm birth. Although risks associated with premature birth increase with decreasing GA, the boundary of prematurity at 37 weeks is relatively arbitrary, since these risks are also increased in infants born $37 - 38$ weeks GA compared to term infants born ≥ 39 weeks [7, 8]. Prematurity is further sub-divided by severity into 'extremely preterm' (< 28 weeks), 'very preterm' ($28 - < 32$ weeks) and 'moderate to late preterm' ($32 - < 37$ weeks) [7].

In this thesis, descriptors of age from birth follow guidelines produced by the Committee on the Fetus and Newborn [9]. As already described, GA is the age counting from the last menstrual period until the day of birth, whilst postnatal age (PNA) counts an infant's days of life since birth (PNA ≥ 0), and postmenstrual age (PMA) describes the total age of an infant since gestation. Before birth, GA and PMA are equivalent, but after birth, GA is fixed by the day of birth, whilst PMA increases with each postnatal day of life since birth (i.e., postmenstrual age equals gestational age plus postnatal age). Finally, for infants born prematurely, there is an additional age descriptor called 'corrected age' which is calculated from the preterm infant's postmenstrual age by subtracting the number of weeks the infant was born before 40 weeks gestation. For example, for an infant with GA $= 30$ weeks, 10 weeks before a 40 week gestation, their corrected age would be 10 weeks less than their PMA. Corrected age is used to appropriately assess developmental outcomes in premature infants compared to their peers in the first few years of life. Additionally, to compare preterm infant development to healthy term-born controls, preterm infants are often studied at term-equivalent age (TEA) which categorises premature infants with PMA in the term range of $37 - 42$ weeks.

1.3 Neonatal Pain

Perhaps surprisingly, given humanity's long personal experience with pain, pain is not a straightforward concept to describe in any population. Pain is particularly difficult to define and determine in a non-verbal population such as neonates, who cannot directly report their pain experience. The appropriate definition and description of pain has been debated for decades. In 1974, a professional organisation dedicated to pain research and management was established to bring together clinicians and multidisciplinary researchers for the first time, named the International Association for the Study of Pain (IASP). The same organisers established a seminal journal dedicated to pain ('PAIN') with its inaugural issue published in 1975 [10]. IASP were the first to publish a standardised taxonomy of pain-related terms and a definition of pain itself, circulated in the PAIN journal in 1979 [11]. It wasn't until 1990 that a group dedicated to improving paediatric pain research and treatment was established by IASP members, named the 'Special Interest Group on Pain in Childhood' (`http://childpain.org/`). Before its inception, IASP had focussed primarily on adult pain (`http://childpain.org/index.php/about-us/history/`). The study and treatment of infant and child pain presents its own challenges compared to the adult population, and necessitates its own careful approach.

There are a number of issues with the historic and current treatment of neonates' and infants' pain; partly a result of poor attitudes and misconceptions, and partly due to inadequate methods of measuring pain and assessing treatment efficacy in this non-verbal population. Studying neonatal pain is complex: behavioural measures are often used but are insufficient, non-specific, not fully validated and not able to determine efficacy in clinical trials for pain management. There is no gold-standard for pain assessment in infants, and more research is required to develop a fully validated and standard assessment tool. There is a low industry drive to develop neonatal pain treatments due to the uncertainty in existing outcome measures, creating a discouraging high-risk high-cost environment for pharmaceutical

companies. The following sections aim to provide a more detailed overview of each of these issues.

1.3.1 Defining Pain and Nociception

The standard definition of pain was recently revised by IASP in 2020 as "An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" [12]. This definition is also accompanied by a series of six short notes which aim to add nuance, clarify possible sources of contention, and highlight the biopsychosocial context of pain in a concise manner. Given the importance of providing understanding and context to the reader, and to fulfil the intention of IASP to produce a combined definition and concise notes to be referenced together, the notes accompanying the definition are presented here:

- Pain is always a personal experience that is influenced to varying degrees by biological, psychological, and social factors.
- Pain and nociception are different phenomena. Pain cannot be inferred solely from activity in sensory neurons.
- Through their life experiences, individuals learn the concept of pain.
- A person's report of an experience as pain should be respected.*
- Although pain usually serves an adaptive role, it may have adverse effects on function and social and psychological well-being.
- Verbal description is only one of several behaviors to express pain; inability to communicate does not negate the possibility that a human or a nonhuman animal experiences pain.

... *The Declaration of Montréal, a document developed during the First International Pain Summit on September 3, 2010, states that "Access to pain management is a fundamental human right." [12].

Of particular interest when considering neonatal and infant pain are the notes separating pain from nociception, highlighting the impact of life experience, and stating that verbal communication is not a necessary aspect of pain experience. The previous IASP definition of pain focused on the presence of "actual or potential tissue damage, or *described* in terms of such damage", which neglected to support

non-verbal humans including infants who cannot describe their experience [11]. The revised definition's removal of the original emphasis on describing pain, and acknowledgement that verbal description is not required to validate pain in the additional notes, is more inclusive of infants [13].

Nociception is described as activity in the nervous system in response to a noxious stimulus [12]. Nociception is separated from pain, since noxious stimuli which stimulate nociceptive pathways do not necessarily lead to report of pain in adults, nor other behavioural or physiological responses which are typically associated with pain report. The converse can also be true, where pain report is not correlated with noxious input, and pain intensity can have emotional and psychological mediators independent of noxious intensity [14]. There are different kinds of pain, including acute and chronic pain, vicarious pain, emotional pain, nociceptive pain, neuropathic pain and nociplastic pain (`https://www.iasp-pain.org/resources/terminology/`). Nociceptive pain must involve nociception; involving activation of nociceptors and nociceptive pathways, but nociception does not necessarily result in pain. Whilst we can be certain of nociception in infants by measuring nociceptive signals in response to noxious stimuli, we can only infer pain in this population. A noxious stimulus is defined by IASP as a stimulus that is damaging to normal tissues or threatens damage to normal tissues (`https://www.iasp-pain.org/resources/terminology/`). Every individual has a pain threshold which represents the minimum noxious stimulus intensity that they perceive as painful. Mild noxious stimuli below an individual's pain threshold can activate nociceptive pathways without resulting in perceived pain [15], which may be indicated by the absence of other physiological and behavioural stress signals in infants.

Where this thesis addresses the topic of neonatal pain, unless stated otherwise, the focus is implicitly on acute nociceptive pain, where pain is assumed to accompany acute tissue-damaging procedures in infants. Acute pain is clinically separate from persistent and chronic pain, which can have different causes, presentations and

consequences [16]. It is more difficult to assess and treat chronic pain in non-verbal infants, as there is limited evidence for diagnosing persistent pain in infants based on current and prior conditions, let alone for providing treatment, whereas most adults can describe their pain state and pain history [16]. Acute pain in infants is more readily assumed from preceding acute noxious stimuli and subsequent correlated acute behavioural and physiological responses. Although progress has been made over the last 40 years of research, there is still an urgent requirement to improve treatment of acute pain in infants. Additionally, there is potential for acute pain studies to provide deeper insight into pain mechanisms and pain-related brain activity in infants to set solid groundwork for future research into the additional complexities of chronic pain in infants. For these reasons, I have focused on acute pain in my thesis research, due to the more immediate potential to address patient need and to provide positive clinical impact now and for the future.

1.3.2 A Brief History of Pain Treatment in Infants

There has been an appalling lack of attention and care to providing appropriate analgesia to infants in recent history. A layperson would consider it obvious that both term and preterm infants feel pain, and yet in the 1980s surgeries were often conducted on preterm infants without adequate analgesia [17–19]. Change began in 1987, when a landmark randomised controlled trial (RCT) by Kanwaljeet Anand and colleagues demonstrated that the common anaesthetic approach in infant surgery known as the "Liverpool Technique", conceived in the 1950s [19, 20], of only administering muscle relaxant to immobilise with a light anaesthetic agent, was wholly inadequate [17]. They found that additionally administering the analgesic fentanyl drastically reduced hormonal stress response in the infants and reduced severe post-operative complications [17]. At a similar time, the story of one infant's post-operative shock and death, following a surgery lasting over 1.5 hours with only induced paralysis without analgesia, became headline news due to the advocacy of the infant's outraged mother, Jill Lawson [21, 22]. Lawson was horrified to discover that this was typical in neonatal and infant surgery, as she recounts in

a letter including distressing details of her child's extensive surgery [21], as were other mothers with equally disturbing accounts of their infants' treatment, or lack thereof [23, 24]. The public were understandably outraged upon becoming aware of the common practice of neonatal surgery performed with muscle relaxants and no analgesia, and deserve recognition for their contribution to updating clinical standards [18]. The combination of Anand's timely research with simultaneous public outcry, resulting from their new-found awareness of the unethical state of neonatal anaesthesia, accelerated the advancement of analgesic use in neonatal surgery [18, 20]. Later the same year, Anand and Hickey published a review of 201 papers presenting comprehensive evidence that newborn infants feel pain, and went one step further than suggesting that analgesia should be provided only to reduce hormonal stress response and post-operative outcomes, but rather that safe analgesia should be provided on humanitarian grounds in its own right [25]. Mere months after the fentanyl trial was published, statements from organisations including the American Academy of Pediatrics [26] and editorials from prominent journals [27–29] were all released in 1987 recommending appropriate use of analgesia. Given that there were strong advocates for the appropriate administration of analgesia to infants to prevent procedural pain even in the 1920s [30], and many neonatologists and nurses were uncomfortable with its absence [24], it is concerning that nearly 70 years passed until including adequate analgesia became a standard in infant surgical care and recommended by relevant authorities.

Even after these formal recommendations, and repeated evidence that analgesia improves infant stress response and health outcomes [17, 25, 31, 32], a number of doctors in the 1980s still protested that infants do not experience pain and that analgesia is not necessary [33]. The practice of male neonatal circumcision is also important, since this surgery is commonly performed on healthy term newborns, in contrast to other infant surgeries performed on vulnerable and premature patients. Throughout the 1990s, analgesia was still not routinely given for neonatal circumcision in the USA, with inconsistencies across regions and between different specialist programs [34–36]. One survey in the USA found that training for up to

40% of clinicians did not include analgesic application for the procedure [35], whilst another survey found that only 24% of physicians performing circumcisions used any analgesia [34]. Half of the surveyed physicians still expressed concern over adverse effects of analgesia, and many held questionable beliefs around the presence and significance of pain in infants [34]. This is despite clear evidence for very simple and safe analgesia for circumcision which significantly improves behavioural [32] and physiological [31] responses in infants, described 20 years prior in 1978 [37]. A similar survey of urologists in the United Kingdom in 1994 did not find such an issue with analgesia and anaesthesia for neonatal circumcision, although the authors suggested that administration of post-operative analgesics for neonates could be improved [38].

Nevertheless, the worldwide rate of change in attitudes and practice in neonatal pain management is remarkable. A survey of board certified neonatal-perinatal physicians in the USA in 1993 found that nearly 100% of physicians believed that premature and term infants perceived pain at all ages (less than 1% believed premature infants <35 weeks did not perceive pain), and most reported the use of anaesthesia and analgesia inter- and post-operatively regardless of infant age, in contrast to the controversy of the late 1980s [39]. These clinicians were aware that their attitudes towards neonatal pain and pain treatment had changed in recent years, citing reasons of increased awareness, scientific literature, and pressure from parents, nurses, and the media [39]. In 1992, a similar comprehensive survey of analgesia and sedation in 87% of the UK's neonatal intensive care units (NICUs) found that pain relief had become common practice, a marked improvement on previous years [40]. The lead author also presents an optimistic reflection on the rapidly updated attitudes and standard-of-care for pain treatment in neonates in the UK in 1998 [41]. Whilst clinical application of analgesia is thankfully now typical in neonatal surgery, its provision remains inconsistent across different settings, as clinicians seek to solve the difficult problem of delivering the safest and most effective anaesthetic regimes for each procedure in the vulnerable infant [19]. Nor is the case of providing appropriate pain relief to infants exclusive to surgical practice;

approaches to pain relief for the many minor routine painful procedures infants receive, such as vaccinations and blood tests, also requires care.

1.3.3 Modern Pain Treatment in Infants

A recent large-scale multi-centre study of the Epidemiology of Procedural Pain in Neonates (EPIPPAIN) reported on the occurrence and management of painful procedures in French NICUs [42]. This was the largest prospective observational study of its kind at the time, including 430 neonates across 13 NICUs which served 20% of France's population [42]. The results of this study shocked medical professionals and researchers alike, for highlighting the sheer magnitude and frequency of cumulative painful procedures that term and preterm neonates are exposed to during NICU stays. Neonates experienced a median of 10 painful procedures per day in hospital, and a median of 75 total painful procedures for each neonate over their hospital stay (up to the first 14 days were included in the study). The highest number of painful procedures recorded for one neonate was 51 painful procedures in one day, and 362 total painful procedures during their hospital stay. The EPIPPAIN study also demonstrated the diversity of pain management strategies, and the inadequacy of their implementation, with 49.1% of painful procedures in neonates performed without any form of procedure-specific or non-specific concurrent analgesia [42]. Additionally, Cruz et al. systematically reviewed all prospective, retrospective and cross-sectional studies of procedural pain in neonates admitted to intensive care units published prior to 2015 [43]. These studies covered 13 different countries across Asia, Africa, Europe, North and South America, and a total of 3156 infants. They collectively found a mean of $7.5 - 17.3$ painful procedures per neonate per day in hospital, and a range of $15 - 100\%$ of procedures performed without analgesia, depending on geographic location. A follow-up to EPIPPAIN, EPIPPAIN-2, was conducted 5 years after the former, in 16 French NICUs in 2011 [44, 45]. They found an improvement in analgesia provision with 75.2% of heel lances performed with any form of analgesia [45], compared to 62.2% in the earlier study [42], and 76% of venipunctures performed with pre-procedural analgesia [44], compared to 71.9%

in the original study [42]. It should also be noted that all of these studies consider the supply of sucrose and other sweet-solution as a form of non-pharmacological analgesia, and the Cruz review recommends increasing its use, despite conflicting reports over whether sucrose is helpful or harmful [46–49].

A further Europe-wide extension to these epidemiological projects was completed in 2015, enrolling 6680 neonates across 243 NICUs in 18 European countries under the EUROPAIN (EUROpean Pain Audit In Neonates) study into practices of sedation and analgesia [50]. EUROPAIN found widespread provision of sedation or analgesia across sites, although pain assessment was only made in 42% of cases, and there were large unjustified variations in practices between NICUs and countries regarding frequency (0-100%) and type of sedation or analgesia given to infants receiving different levels of supportive ventilation [50]. The percentage of painful procedures including pain treatment compared to the number of painful procedures being performed in neonates revealed by these epidemiological studies, cumulates to an incredible number of untreated painful procedures in this vulnerable population. Each of these studies revealed different standard-of-care between NICUs e.g. different uses of swaddling, sucrose, and local anaesthetic, depending on procedure. There is still widespread uncertainty and inconsistency in pain treatment strategies in the NICU. Attitudes are shifting towards non-pharmacological interventions such as gentle touch and skin-to-skin contact (also known as 'kangaroo care') [51], which are not thought to pose any potential risks, unlike pharmacological interventions, whilst demonstrating efficacy for reducing pain and distress [51, 52]. Sucrose is often reported as non-pharmacological in the literature, but some authors believe this is a misclassification. Instead, proposing that sucrose be reported as a pharmacological intervention, as some research suggests sucrose has a similar mechanism of action to opioids [53], with possible adverse effects [48, 53, 54], and requiring appropriate dosing management [53, 54]. In recent years, physicians have also been reducing the number of painful and distressing procedures infants' receive [51]. This is a good strategy to reduce excessive pain exposure and its negative consequences. When considering the ethical and humanitarian concerns, as well as evidence for adverse

effects of pain exposure on long-term health and wellbeing [e.g., 36, 55, 56], neonatal pain management needs improvement.

1.3.4 Clinical Pain Assessment Tools

The infant experience is particularly difficult to quantify due to their limited tools for self expression. Foremost, they cannot articulate their experiences, unlike verbal populations of older children and adults. Limitations of self-report are particularly important with regard to the subjective and variable context of pain; different people can be subject to the same noxious stimulus and have vastly different perceptions of pain. Clinical practice for neonates, and much of current neonatal research, relies on "pain scores" which aim to quantify an infant's pain experience using a combination of behavioural measures (e.g. facial grimacing), physiological measures (e.g. heart rate and oxygen saturation) and contextual factors (e.g. age, sleep-wake state). These benefit from straightforward implementation, important for clinical practice, and common scales such as the Premature Infant Pain Profile - Revised edition (PIPP-R) [57] and Neonatal Facial Coding System (NFCS) [58, 59] are considered to be fairly reliable [60, 61]. Over 40 pain assessment tools have been developed [62] since the first in the 1980s [58, 59], and yet none are of these are an accepted "gold-standard" in infant pain measurement, and most have been developed with only acute and/or procedural pain in mind [62]. Furthermore, whilst certain measures have been shown to correlate significantly with noxious events (e.g. blood tests), behavioural and physiological signs underpinning these assessments may sometimes show inconsistent or even opposite effects [46, 49, 62–65]. Infants' behavioural expressions are not specific to pain over distress [64, 66, 67] and their ability to communicate pain also depends on contextual factors including age, energy to cry, ill-health, and ongoing or prior pain [62]. These factors can also affect physiological indicators of pain, for example, inflammation related to infection can suppress behavioural and physiological indicators of pain compared to noxious-evoked brain activity [68].

Clinical measurement tools must be practical and provide accurate results in order to benefit patient care. The evaluation of these tools must also be robust and relevant to determine whether the measurement truly reflects the intended construct. A Delphi study of over 40 international experts in epidemiology, statistics, psychology, and clinical medicine was conducted to clarify and standardise the relevant measurement properties for evaluating health measurement instruments [69]. The study produced the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN). The COSMIN taxonomy presents three key domains relevant for evaluating health measurement tools (a) reliability, (b) validity; (c) responsiveness. Reliability and validity can be broken down into key measurement properties: reliability includes test-retest, inter-rater and intra-rater reliability, measurement error, and internal consistency; validity includes content validity, face validity, construct validity, and criterion validity. The main definitions are outlined in Table 1.1. Interpretability, the degree to which score values or changes in value can be assigned a clinical or commonly understood meaning, whilst not strictly a COSMIN measurement property, is also considered an essential quality for health measurement tools [69].

Systematic reviews have demonstrated the poor quality of common infant pain assessment scales, with over half of scales lacking validity and reliability [71], along with inadequate demonstration of appropriate responsiveness of tools to pain reduction [72], calling their utility into question. The former review assessed the quality of 57 infant pain assessment scales [71]. They found that over half lacked construct validity, internal consistency and inter-rater reliability, and of the remaining 28 scales which were adequately designed with these qualities, only a quarter (7) had low risk of bias [71]. The authors determined that the highest-quality scales with low risk of bias for acute pain assessment in preterm and term infants were COMFORT, NFCS, N-PASS and PIPP. These are all commonly used in clinical and research settings, although less favourable scales identified by the review are also in common practice. According to this review, not all scales suggested for use by the American Academy of Paediatrics have appropriately demonstrated validity and reliability in

Measurement Term	Definition
Reliability Domain	The extent to which scores for patients whose state has not changed are the same for repeated measurement under several conditions, e.g. over time (test-retest), by different people for the same occasion (inter-rater) or repeated examination by the same person (intra-rater).
Reliability	Proportion of measurement variance due to true differences among patients.
Internal Consistency	Degree of interrelatedness among measurement items, i.e. the extent to which each item of a tool measures the same construct.
Measurement Error	Difference in measurement which is not due to true changes in the construct being measured, but due to systematic and random error.
Validity Domain	Degree to which the measurement tool actually measures the intended construct(s).
Content Validity	Degree to which the content of the measurement tool adequately reflects the construct to be measured; whether the measure is relevant and comprehensive. Within this 'face validity' is the subjective assessment of whether the tool appears to reflect the construct to be measured.
Construct Validity (There is no gold-standard)	Degree to which scores of the measurement tool concur with a-priori hypotheses such as expected differences between groups, with different stimuli, or relationships to other variables, for the intended construct. i.e. how well measurements agree with evidence-based expectations.
Criterion Validity (There is a gold-standard)	Degree of agreement with assessment using a pre-existing gold-standard.
Responsiveness Domain	Responsiveness is treated as its own domain rather than a subcategory of validity, since it relates to multiple measurements rather than individual measurements. The only property is responsiveness, defined below.
Responsiveness	The ability of the measurement tool to detect changes in the intended construct over time. Responsiveness is only a reflection of longitudinal changes in the construct, not differences between different patients at the same point in time. I.e., when the same patients are tested at two different timepoints, if there has been a change in the construct (whether due to intervention or otherwise) the measurement tool should detect this.

Table 1.1: Key measurement properties for health tool evaluation according to COnsensus-based Standards for the Selection of Health Measurement INstruments (COSMIN) taxonomy [69, 70].

the populations in which they are used, and some were based on very limited samples or flawed assumptions [71]. In the latter complementary systematic review, Meesters et al. found that only 10 of 43 published pain assessment scales for neonates have had direct studies of responsiveness to intervention, and none of these responsiveness studies were of good quality [72]. Standards for measurement quality assessment were not clear until the publication of COSMIN guidelines in 2010 [69], which were used to assess quality of the scales in both of these reviews [71–74], therefore inconsistencies in validation protocols and metrics considered by original studies prior to this publication by COSMIN are somewhat understandable. Nevertheless, pre-existing tools need to be held to the same standards as modern developments, and these reviews help to highlight the inadequacies of many accepted tools, promote selection of the highest-quality and most relevant assessment tools available, and drive further necessary research. A greater degree of attention needs to be given to appropriately validating pain assessment scales in their target populations.

A number of research groups have been working to develop cortical activity measures as potentially more objective and consistent measurements of pain experience, since pain perception is processed in the brain. Neonatal neuroimaging tools will be discussed later in Section 1.4, and their application to the field of infant pain and pain measurement will be discussed in Section 1.4.4.

1.3.5 Impact of Pain on Long-Term Development

Pain and procedural medical history have a significant influence on infant development, particularly in premature infants who are subject to many more procedures than healthy term-born infants due to the extra and more intensive care that preterms require in their first weeks of life, e.g. through eye examinations and blood tests. As revealed earlier, untreated procedural pain can cause adverse short-term stress responses to the point of fatality in the extreme [17], but there can also be serious long-term consequences. Repetitive pain in the early life of premature infants has been linked to both hyper- and hypo-sensitivity to pain in later life with altered

sensory processing [56, 75–78], and persistent abnormal brain development not only at infancy, but also in school age children and adolescents [56, 79–81].

In the late 1990s, Anna Taddio and colleagues demonstrated that circumcision in newborn males resulted in hyperalgesic responses to vaccination at 4-6 months since birth, compared to uncircumcised males [36]. They also demonstrated that providing topical local anaesthesia for the circumcision procedure reduced the extent of circumcised infants' hyperalgesia to vaccination, compared to those who did not receive anaesthesia for circumcision (as was standard at the time). This landmark study demonstrated both that early life exposure to pain can increase sensitivity to subsequent painful procedures, even many months later, and also that providing pain relief for these procedures can reduce the long-term negative outcomes. Whilst the circumcised males who did receive a topical local anaesthetic still exhibited hyperalgesia compared to uncircumcised males, it is possible that the post-operative pain was not adequately managed in the circumcision group, since this was not administered through the study protocol. Therefore, comprehensive pain relief for early-life painful procedures could potentially reduce hyperalgesic side-effects further towards baseline. However, the most straightforward method to reduce long-term adverse outcomes associated with pain is to reduce early life pain exposure where possible.

Many studies which demonstrate adverse outcome associated with pain exposure have been conducted since. One study by Grunau et al. firmly demonstrated that the number of painful procedures a preterm infant experiences profoundly influences their neurodevelopment at 8 and 18 month follow-up [55]. The mean number of skin-breaking procedures, as a simple marker of cumulative pain exposure, for the cohort of 137 preterm infants born at less than 32 weeks GA, was 121 procedures. Grunau and colleagues found that the number of skin-breaking procedures was strongly correlated with impaired cognitive and motor development at 8 and 18 months of age, after controlling for early illness severity, days on morphine and on dexamethasone [55]. Interestingly, lower GA was not associated with lower

cognitive and motor outcomes when controlling for the same factors, even though lower GA is strongly correlated with increased skin-breaking procedures. This demonstrates the importance of reducing painful experiences in preterm infants, and that understanding and managing pain must be a key goal for improving neonatal care.

1.3.6 Clinical Unmet Need

There has long been great disparity between pain relief provided to infants compared to adults, including post-operatively and for acutely painful procedures, and this remains the case. The provision of pain relief to infants is not standardised, medication may be prescribed but never administered, and there is a poor understanding of when and how much pain relief to administer due to the challenges in assessing pain state in infants. This not only limits assessment of pain to provide analgesia to infants in practice, but impedes any demonstration of analgesic efficacy in clinical trials. These problems mean that licensing analgesics for use in infants and neonates is extremely difficult, and most drugs are therefore used off-label in these groups. The anxious clinician's concern about providing off-label medication to infants is not without cause, since infants can respond differently to adults when given the same active ingredient, appropriate doses will vary significantly within this rapidly growing population, and pharmacokinetic (PK) and pharmacodynamic (PD) studies, which adult prescriptions rely on to determine safe and effective dose regimes, are all but non-existent in infants [82, 83]. A lack of adequate literature and potential adverse side-effects associated with neonatal drug use lead to unclear choices for appropriate analgesic provision [84, 85].

Most drugs are used off-label in neonates due to limited clinical trials into the efficacy, safety, and appropriate dose regimes. To improve this, the European Medicines Agency (EMA) [86] (currently under revision [87]) and US Food and Drug Administration (FDA) [88, 89] both produced legislation to encourage paediatric drug investigations, however this has not seen follow-through of increased licensing as hoped [90, 91]. The EMA are currently updating their guidelines, which may have a

positive impact [87], and both the EMA and FDA have hosted workshops to increase cross-disciplinary collaboration, to understand the challenges in regulatory approval, and work to resolve these issues [92, 93]. A recent multicentre study of on-label and off-label neonatal drug use in The Netherlands found large variation between NICUs [94]. The authors used a broad definition of on-label drug prescription, where medicines were considered on-label if the active ingredient had an indication (i.e., use-case) described for neonates, including if there is a non-specific indication for children in general, according to the Dutch Medicines Evaluation Board [94]. This is over-optimistic, since an indication for neonates, or all children, may be described despite a medication not being licensed in neonates under all circumstances, the license depends on delivery method such as intravenous or oral dosage, and a particular dose strength would not be appropriate for both preterm and term neonates of different PMA and weight. For example, they included all use of paracetamol as on-label, which is not licensed in (a) infants under 2 months (b) under 3 months (c) weighing less than 10 kg or (d) preterm neonates, depending on delivery method (`https://bnfc.nice.org.uk/drugs/paracetamol/#unlicensed-use`). With these considerations, they still found that at least 31% of all nervous system drugs were prescribed off-label for the neonatal age. Another study of the use of off-label and unlicensed medicines in NICUs used a more specific classification of off-label and unlicensed drug use in neonates, by classification according to information provided by the FDA [95]. This study found a more realistic proportion of 49% of medications used off-label and 25% unlicensed in neonates, that almost all neonates (96%) were exposed to off-label drugs and two-thirds were exposed to unlicensed medications. A fifth of these neonates received nervous system drugs of which 48% were used off-label and 50% unlicensed. Fentanyl, an opioid analgesic given intravenously, was the most prescribed off-label drug. Whilst this research was conducted in Brazil, use of off-label and unlicensed medications in neonates is a global issue, with similarly high frequencies reported internationally [96–102]. Recommendations and analgesic drug labelling also varies internationally, in the US only oral paracetamol has approved use in neonates, with the next approved infant drug being ibuprofen

in infants aged 6 months or more [103]. Whereas in Germany, morphine is approved in neonates, paracetamol is approved in term neonates, and ibuprofen is approved from 3 months [103]. Lack of validated objective measures of pain relief create a poor understanding of the balance of short- and long-term benefits and adverse effects of these drugs in infants, and contribute to discrepancies in international drug guidelines.

The EUROPAIN study of sedation and analgesic practices in NICUs in 18 European countries found that 6 countries had national guidelines for neonatal sedation and analgesia, and 75% of NICUs in the study had local protocols [50]. The widespread variation in sedation and analgesia provision across the EUROPAIN countries would be reduced if there were informed international guidelines for the use of sedation and analgesia in the NICU [50]. To produce adequate international guidelines, we need more trials with robust methods to investigate the therapeutic and adverse effects of existing and novel medications in neonates for pain-relief, as well as the efficacy of safe non-pharmacological interventions. The benefits of these evidence-based guidelines, and validated safe and effective pain-relief in infants, would not just serve a European population, but all populated continents. Pre-existing treatments for acute infant pain including morphine and sucrose cannot be recommended outright, as long-term adverse effects have been observed with both including poorer motor development [47, 55, 104], as well as growing evidence that morphine [105, 106] and sucrose [48, 49, 54, 107] are not effective analgesics for acute procedural pain in infants. Two of these sucrose studies do not themselves dispute the use of sucrose as an analgesic, since reduced behavioural expression of pain was evident in sucrose treatment arms [48, 107], however they show that sucrose did not reduce hyperalgesia - which is associated with increased pain experience [56, 108] - during subsequent procedures [48], nor reduce physiological signs of pain and distress as the authors expected, including heart rate, oxygen saturation and energy expenditure [107]. These results suggest that whilst sucrose exhibits a visibly calming effect on infants, it lacks therapeutic benefit on pain experience and may not necessarily qualify as an adequate analgesic.

Even when drugs are recognised as having a therapeutic benefit, such as opioids for ventilated preterm infants, delivering an optimised efficacious and safe dose remains challenging. Neonates have greater sensitivity to opioids and lower drug clearance, and PK/PD profiles depend on the rapidly changing infant weight and developmental stage [82, 109–112]. Research suggests misunderstanding of differing drug sensitivity and clearance in neonates may lead to unnecessarily high opioid dosage above the therapeutic limit [110]. The EMA published guidelines on the role of PK in developing paediatric medicinal products in 2006 [113], which are due an update [114]. Appropriate trial designs would account for developmental differences by stratifying neonates by factors including GA, PNA, birth weight, and accounting for prematurity and co-morbidity, but recruitment in the neonatal population is already difficult, and ethical considerations mean reducing risk through minimising sample numbers [82, 115]. Optimal dosing studies are required for each drug indication, for example, optimal doses for ventilation will be different for acute painful procedures and post-surgery. The paucity of neonatal PK/PD studies is partly due to ethical challenges; PK studies typically require regular blood sampling which conveys pain and distress, but you cannot increase harm to infants through increased pain exposure without providing any potential direct benefit to the infants in the study. This issue is avoided in older participants who can provide consent for additional procedures. The challenge of profiling developmental PD is also aggravated by the lack of validated and objective measures of pain in infants which can be used to quantify pain relief.

1.3.7 Clinical Trials

Conducting clinical trials for pain relief in neonates is complex [116]. A consensus meeting of the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION; `www.acttion.org`) Pediatric Pain Research Consortium (PPRC) recently convened to synthesise the evidence around acute pain models and appropriate clinical trial design attributes for neonatal, infant, and older paediatric populations [116]. This consortium included participants

from universities, government agencies, and industries with relevant expertise in treating acute pain or conducting clinical trials for acute pain in paediatric populations. Acute pain models for drug treatment trials in adults are not directly applicable to paediatric cohorts. Recommended adult acute pain models include conditions such as bunions and dental care [117], which are not relevant to paediatric groups. Recommended neonatal and infant acute pain models represent the most common procedures; heel lances and circumcision [116]. These are both frequent and typical in standard clinical practice, although the latter only applies to male infants and its prevalence varies internationally and culturally. Therefore, heel lances, for routine blood tests in the clinic, are the preferred acute pain model, and non sex-specific general surgical procedures such as hernia repair can also be considered [116]. The ACTTION PPRC recognised challenges in neonatal and infant clinical trials for pain relief, including the aforementioned lack of a gold-standard for infant pain assessment. Infant pain scales are not sensitive enough for clinical trials, owing to the subjective measurement of pain and dependence on infant maturation [66, 116]. Pain assessment tools and outcome measures must consider cognitive maturation and behaviour [66, 116]. Subjectivity and inter-individual variability in noxious sensitivity increase the uncertainty and error in pain assessment, presenting a challenge for using pain assessment scales as outcome measures in clinical trials [118–120]. Multimodal assessment of infant pain encompassing cortical, behavioural, and physiological measures could provide improved outcome measures for pain in future trials [121]. Methods to adjust for inter-individual variability in noxious sensitivity could also improve the sensitivity of pain scores as outcome measures in pain treatment interventions [122]. A review by Olsson et al. reported that clinical trials don't always use a contextually valid pain scale either (16% of cases) [123], e.g. they use a scale validated only for acute pain for the assessment of prolonged pain, or use scales which haven't been validated in the target population [123]. The reviewers also note an unmet need in terms of infant post-operative and prolonged pain relief trials, which are greatly under-represented, as the large majority of clinical trials concerned acute pain procedures which could be considered excessively repetitive [123]. They

suggest that it would be more worthwhile to direct some of these resources towards the issue of developing care for prolonged and chronic pain in infants, an area of great clinical relevance [123]. Although this is a far more challenging task, and acute pain management remains unresolved.

Another related issue raised by the ACTTION consortium is that deciding the minimal clinically important decrease in pain according to any infant pain assessment method remains unclear [116, 124, 125]. In adult studies, using verbal pain rating scales, minimal clinically significant reductions in pain vary depending on the initial pain severity [126], and developing appropriate thresholds for clinical significance in adult patients is facilitated by the patients' ability to directly report whether they consider the pain reduction clinically meaningful [127]. Similar testing to identify minimally clinically meaningful thresholds for pain reduction are not possible in infants, and our best approximations include general guidelines from research in older populations [128], or the subjective perceptions of parents [125] and clinicians [120, 125, 129]. Pilots or prior research are also often used to determine meaningful difference estimates for pain assessment metrics [130, 131], but some incorrectly cite prior research which does not justify their chosen threshold [132, 133], and there is no good way of determining whether reduction thresholds are optimistically narrow or pessimistically wide. These uncertainties could lead to overly stringent requirements for approving a pain relieving intervention if pain reduction boundaries are set too high, or lead to over-medication if the boundaries are set too low.

Tensions are high around neonatal and infant trials, due to concerns over ethics, including the use of placebo-controlled trials [134], even where standard clinical care is included in all trial arms, due to misunderstandings around the meaning and implementation of placebo-controls as necessarily with-holding care [135, 136]. However, outrage at the existence of no-treatment arms in RCTs and infant pain research is justified [e.g. 137][1], since all neonates should receive the best standard of

[1]The authors state that "To avoid confounding by other potential pain-relieving mechanisms, non-nutritive sucking and touching, or swaddling of participants was not permitted during the study." [137], thereby denying the included infants standard-of-care to reduce pain and distress

pain relieving care at time-of-study. Bellieni and Johnston reviewed the prevalence of unnecessary infant pain in clinical trials published between $2013-2015$ [134], but grouped together trials with no-treatment arms which restricted interventions [e.g. 137] with trials using a placebo/control arm which included routine care [e.g. 138]. The authors admonished all of these randomised placebo-controlled trials equally, but only the former is unequivocally unethical. Placebo arms are ethically sound in clinical trials for analgesia provided that the placebo arm still includes standard-of-care or proven analgesic treatments for the procedure under study. The comparative treatment arm should include the treatment under study in addition to any treatments and standard-of-care present in the placebo arm. The placebo should serve to ensure blinding to the treatment and eliminate intervention effects unrelated to the active treatment in the clinical trial population, e.g. by providing an oral placebo *vs* an oral drug on top of routine care, not to restrict pain relief provision. ACTTION also recommend randomised placebo-controlled trials in infants with the use of rescue analgesia, given that excess analgesia, such as overexposure to opioids, has also been associated with adverse outcomes [116]. Therefore, minimising harmful overexposure through rescue doses balances risk, rather than being unethical in withholding analgesia [116]. Additional approaches which attempt to address ethical considerations are also possible, such as a flexible randomisation design where infants can be assigned to a group based on parental preference, or randomised if there is no preference, or designs which assign more infants to the treatment arm on the assumption that it will be beneficial [116].

Where neonatal clinical trials for analgesics are attempted, they do not always have the desired outcomes. One recent clinical trial assessing the safety and efficacy of morphine for Procedural Pain in Premature Infants (Poppi) was stopped early by the data monitoring committee due to increased respiratory difficulties in morphine-treated infants compared to controls [106]. This represents one case where the balance of risk associated with increased hypoxic events, which can be life-threatening and lead to numerous impairments [139], outweighed the benefits of morphine for analgesia [106]. When there is so little evidence for the right treatment approaches,

and drugs are routinely used off-label, it is equally important to ascertain when and what not to prescribe for neonates to prevent avoidable harm, as well as to validate the safe and effective medications. The balance between risk and benefit for drug use is context-dependent, and each potential use-case needs to be rigorously, and safely, tested to determine the best treatment action for infants [140]. However, from the point-of-view of pharmaceutical companies, it is not worth investing money into expensive trials for neonatal analgesia when they are likely to fail to prove clinical efficacy and meet regulatory requirements in order to receive return on their investment. There are a growing number of calls to action from academics and clinicians to improve the landscape of evidence-based and licensed treatment in the vulnerable infant [85, 91, 93, 103, 116, 140]. Future progress depends on the collaboration of regulators, funders, researchers, clinicians and pharmaceutical and drug-discovery companies, to improve the evidence-base for safe and effective neonatal medicines [85, 93, 103].

1.3.8 Summary

Although a complex task, improving infant pain assessment tools both for clinical practice and for analgesic clinical trials is imperative for providing adequate pain relief. As well as humane considerations to do no harm applying to reducing acute procedural pain, the adverse long-term effects of pain experience further cement that we need to improve the treatment of pain in infants. This issue is particularly important for preterm infants, who experience numerous painful procedures during NICU stays, in order to improve their developmental outcomes.

1.4 Neonatal Brain Imaging Techniques

In this section I will describe the major non-invasive neuroimaging techniques used for studies of human neonatal and infant brain development, and which are relevant to the research methods in this thesis. The history of *in vivo* non-invasive neonatal neuroimaging dates to the early 1900s with the application of electroencephalography (EEG) [141, 142]. However, complementary neuroimaging techniques have a very

short history: magnetoencephalography (MEG) dates to the 1960s [143], the first newborn functional near infrared spectroscopy (fNIRS) studies were in the 1980s [144, 145], and *in vivo* magnetic resonance imaging (MRI) began in the 1980s [146]. EEG is the only neuroimaging technique commonly used for routine clinical assessment at the cot-side in NICUs, owing to its portability, ease-of-use, and known clinical markers of health. In this section I will describe how pioneering EEG studies provided early insight into developmental stages of infant brain function. I will then explore recent advances in neonatal neuroimaging, which have mainly related to developments in neonatal MRI within the past two decades. Finally, I will discuss the challenges associated with neonatal neuroimaging, and how neuroimaging techniques can be used to study neonatal pain and long-term health.

1.4.1 Electroencephalography

EEG is a long-standing technique for measuring brain activity by measuring the electrical fields produced by neurons firing using a set of electrodes positioned on the head. Neonatal EEG has a long history, with recordings dating back to the 1930s [142], and even foetal EEG results were published shortly after in the early 1940s [147]. Since then, researchers have conducted numerous studies to characterise normal and pathological EEG activity in term and preterm neonates of different ages. Pioneering work in this field was largely produced by French neurophysiologists in the 1950s [148–150], therefore certain internationally recognised technical terms describing EEG patterns remain in their original French [149]. In the late 1990s, a world-class team of French and English-speaking physicians sought to harmonise the use of French and English terminology in neonatal EEG research, drawing from literature of both languages to publish the first glossary of neonatal EEG [148], which has since been updated in 2010 to reflect advances from analogue to digitised EEG recordings [149].

Interpreting neonatal EEG requires consideration of GA, PMA and arousal state, amongst other clinical details [149]. Characteristics of normal EEG are rapidly evolving from preterm to post-term; the EEG trace of a healthy infant at

30 weeks observed in a 32 week infant would indicate dysfunction and dysmaturity. In extremely preterm neonates, background EEG is characterised by discontinuous traces (*tracé discontinu)* comprising bursts of activity separated by long intervals of inactivity, 'delta' waves which describe high-amplitude low-frequency activity, and sharp 'theta' waves of variable amplitude, with no discernible difference in EEG activity across arousal states. With increasing PMA in the preterm, background EEG becomes a mix of discontinuous and continuous traces, with periods of continuity lengthening, and number of theta waves increasing, with increasing age. By 30 weeks, arousal or sleep-state is identifiable from the EEG traces, differentiated by the degree of trace continuity and artefact. At the same time, the slow delta waves become progressively superimposed with low amplitude high frequencies, known together as 'delta brushes' [150, 151]. Delta brushes are a hallmark of prematurity which disappear from healthy EEG by 38-40 weeks PMA [152]. By term age, and in the term newborn, EEG should comprise predominantly continuous activity across all arousal states [151, 152]. Given the complex evolution of EEG activity with age, it is imperative to have accurate infant demographic and clinical data and to consider these factors when interpreting infant EEG data.

By neonatal neuroimaging standards, EEG has a long history, and its features are well-understood and well-characterised across infancy with respect to both healthy and pathological brain activity [151]. EEG is routinely used for cot-side monitoring in neonatal intensive care and, owing to its portability and flexible set-up, it is used for continuous measurement without disrupting normal infant care, such as for pre- and post-surgical monitoring. Recent and future advances in EEG surround increasing automation, whether using artificial intelligence (AI) for accurate artefact detection, automated identification of pathologies and infants requiring care, or for short- or long-term outcome prediction models. One such implementation of recent advancements in AI algorithms in infant EEG research has been the development of accurate prediction models of infant brain-age from EEG data [153]. However, significant advances in neonatal brain imaging techniques in the past 10-20 years

have largely concerned the application of MRI in neonatal populations, which are discussed in further detail below.

1.4.2 Recent Advances in Neonatal Brain Imaging

Modern advances in non-invasive neuroimaging techniques have revolutionised our understanding of the human brain, including expanding our knowledge of infant brain structure and function, and early brain development. In combination with results from adult and animal studies, this technology has also allowed us to probe questions of how infants may experience the world, which are incredibly difficult to answer without the capability of direct report. In particular, imaging techniques offer value towards understanding pain experience.

1.4.2.1 Functional MRI

One key neuroimaging technique for examining the functional development of the brain is functional MRI (fMRI), which has been vital to uncovering the early organisation of functional networks of activity in the preterm and newborn. fMRI measures blood-oxygen level-dependent (BOLD) signals related to fluctuations in brain activity [154]. This signal arises since an increase in brain activity increases metabolic demand in the region of activity, which is followed by an increase in blood flow to deliver oxygenated haemoglobin to the active region [154]. These fluctuations can be measured at rest, as well as in response to tasks and stimulation. fMRI can provide insight into resting-state brain function as well as evidence for spatially-localised functionally-specific brain regions which are active in response to specific stimuli, through non-invasive high-resolution images of deep brain structures in the order of millimetres [154, 155].

Resting-state fMRI

The advent of resting-state fMRI (rs-fMRI) has been fundamental to uncovering the functional organisation and connectivity of the infant brain, and studying normal and abnormal early brain development. Understandably, in infants and young children,

the use of task-activation fMRI to probe specific task-related activity, is difficult in a population who cannot follow instructions, whereas rs-fMRI can be acquired in any state of rest, typically during sleep in infants. rs-fMRI can be used to examine correlated activity known as functional connectivity (FC) between pairs of brain regions, as well as measure spatial patterns of temporally-correlated brain activity known as resting-state networks (RSNs). Changes in FC and the shape and strength of RSNs in infants can be mapped with age and developmental stage [155–158]. Whilst the first rs-fMRI of FC in adults was published nearly 30 years ago [159], the first study in neonates was only published in the last 15 years [160]. The number of rs-fMRI studies in neonates has grown significantly since. A cross-sectional study comparing FC in neonates to that of 1- and 2-year old infants was published one year after the first neonatal fMRI study [161], and several key studies into the longitudinal development of functional networks in premature infants were published soon after [156, 162]. Although several studies have investigated the relationship between features of structural and microstructural MRI in neonates to their longer-term neurodevelopmental outcomes, such as cognitive and motor function at school age [163, 164], there are very few studies comparing functional development using rs-fMRI data in neonates to later neurodevelopmental outcomes [165, 166]. Despite these advances, the field of infant rs-fMRI research remains small, with a PubMed search of key terms identifying up to 216 infant rs-fMRI studies published since 2007[1], compared to almost 9,000 adult rs-fMRI studies in the same time-frame[2]. A similar search by Zhang et al. in 2019 concluded there were fewer than 100 individual infant rs-fMRI studies published within the previous two years compared to over 2000 total rs-fMRI studies in the same period [167].

A recent study by Smyser et al., in 2010, characterised the longitudinal development of FC and RSN localisation in preterm infants born at $23 - 34$ weeks,

[1]PubMed search on 01/08/2022: '(infant) AND (resting-state functional MRI)) NOT (review[Publication Type])', using filters to include only human studies of newborns and infants 0-23 months. Search results were not manually checked and the true number of original infant rs-fMRI studies in this search may be fewer due to imprecision of the search methods [167].

[2]PubMed search on 01/08/2022: '(resting-state functional MRI) NOT (review[Publication Type])', using filters to include only human studies of adults of 19+ years.

scanned at multiple timepoints throughout prematurity and at term-equivalent age (from $26-42$ weeks PMA), with comparison to term-born controls [156]. Smyser et al. also described 11 major RSNs identified at each timepoint and mapped the maturation of these RSNs through changes in their strength and shape with age. They identified each of these RSNs in preterm infants at all scans, including the earliest at 26 weeks, demonstrating that even extremely premature infants have key functional networks. However, there were characteristic developmental changes in these functional networks with age. RSNs were initially more diffuse in nature in premature neonates, becoming more localised with increasing PMA. In general, long-range interhemispheric FC increased with PMA, however, differences in development of long-range FC between hemispheres depended on the network of interest. For example, increases in interhemispheric connectivity with increasing PMA were evident in the sensorimotor cortex, whilst long-range FC in the temporal cortex remained low throughout all ages studied, and anterior cingulate cortices had stable significant interhemispheric connectivity across all ages studied including even the most preterm infants. This demonstrated clear heterogeneity in the maturation of different functional networks and emergence of bilateral connectivity. They also found evidence that prematurity impacted development of functional networks by comparing preterms at TEA with term-born controls. Term-born controls had further localised RSNs with increased long-range interhemispheric FC to homotopic regions compared to their preterm counterparts. Doria et al. detailed a complementary study of preterm and term-born infants ($29-43$ weeks PMA) a few months later, and was the first to illustrate the emergence of facsimiles of all major adult RSNs by term age [162]. Doria et al. identified a more complex default mode network (DMN) at term age, involving cortices comprising the adult DMN, where previous studies only reported fragmented DMN components similar to those found by Doria et al. in the preterm infants studied before term [162]. Doria et al. reproduced findings of the Smyser study concerning the maturation of FC in the early infant brain, finding increases in interhemispheric connectivity of sensorimotor networks, amongst others, with increasing PMA [156, 162]. Together, these studies have shown that

infant development of functional networks does not progress equally, with different trajectories of spatial organisation, connectivity, and strength for each network; nevertheless, all major adult RSNs, including facsimiles of higher-order default mode and executive control networks, are present in healthy term and preterm infants at TEA [155, 156, 160, 162]. However, unlike Smyser et al., Doria did not find differences in FC between preterms at term and term-born controls. The studies differed in their sedation practices, and both studies included only $10-12$ term-born controls, so they may have been underpowered to test for differences between groups. Doria et al. also suggested that, although both studies aimed to recruit healthy preterms without brain injury or abnormality, the discrepancy may be explained by undetected clinical differences in the preterms, which may only become apparent with long-term follow-up. The tendency for research to restrictively recruit only the healthiest preterm infants does not provide a truly representative cross-section of the premature infant population [168], and until very recently, there had been insufficient studies to provide any further clarification.

Novel large-scale Open Science projects such as the developing Human Connectome Project (dHCP; `www.developingconnectome.org`), have enabled more detailed and powerful studies of early infant brain development by pooling multi-centre resources to gather high-quality neuroimaging data of an unprecedented number of diverse infants, and releasing this data for wider academic research. Eyre et al. harnessed this dHCP dataset release to conduct the first large-scale study of RSNs and FC in over 300 infants studied at term, including 65 preterm infants, in 2021 [158]. Their study sought to address the outlined discrepancies in prior research, such as the nature of the developmental trajectory and emergence of an adult-like topology of higher-order networks including the DMN in infants, and to determine precisely how preterm birth affects RSN development and FC by term-equivalent age, compared to term-born controls. With this large dataset, they were able to precisely map the typical development of FC and organisation of RSNs throughout the perinatal period. They found agreement with prior studies, showing that FC increases in premature infants with increasing PMA, and finding

heterogeneous development of functional networks, whereby primary networks, such as visual and sensorimotor networks, mature earlier than higher-order networks such as the DMN. Meanwhile, despite the clear maturation of preterm infants' functional networks as they age, they found that preterm-birth is associated with significant disrupted development of functional networks, with FC at TEA largely reduced compared to the standard of term-born controls. This disruption was dependent on degree of prematurity, as lower GA was associated with greater decreases in FC at TEA compared to term-borns. They also concluded that whilst they observed DMN-like network fragments, which showed the greatest increases in within-network connectivity throughout the term period, these RSNs did not include all cortices implicated in the adult DMN, and therefore these results support a longer developmental trajectory of adult-like higher-order networks, which other studies suggest completes over the first year or two of life [157, 169, 170].

Stimulus-evoked fMRI

Whilst much neonatal fMRI research has utilised resting-state data, stimulus-evoked fMRI (stim-fMRI) studies, the counterpart to task-fMRI in adults, have also generated significant advances in neonatal neuroimaging in the past decade. Initial stim-fMRI studies successfully characterised BOLD responses to sensorimotor stimulation, by using programmable passive motor and tactile stimuli of the hand [171], and robotic devices to measure both passive and active motor function of the wrist [172, 173]. These investigations found localised somatosensory cortex and primary motor cortex activation contralateral to the stimulated areas in preterm and term infants [171, 173], matching patterns of activity found in adults [174]. These results demonstrated that the sensorimotor cortex is already considerably organised in newborns, and even in early prematurity, including the functionally lateralised response pattern observed in adults [175]. Longitudinal analyses in preterm infants from preterm to term-equivalent PMA found that the BOLD responses progressed with increasing age towards more complex activation patterns, including bilateral responses [171, 173], increased localisation of functional activity, and integration of

further sensorimotor regions including the secondary somatosensory cortex and basal ganglia [173]. These patterns may demonstrate maturation in these infants towards more adult-like sensorimotor function, as they are similar to the complex responses to active motor stimulation observed in adults [174]. These results have been recently reproduced by Scheef et al. using novel neonatal-specific MRI coil technology, which improves signal-to-noise in infant data by fitting closely to their markedly smaller heads compared to standard MRI coils designed to suit adults [175]. Scheef et al. also conducted a 2-year follow up to demonstrate that these early functionally-lateralised sensorimotor responses to passive motor stimulation were present in preterm infants without evidence of abnormal motor function at 2-years [175]. Similarly, sensory fMRI studies of noxious [176–178], gentle tactile [176, 179], visual [180], auditory music [181] and language [181, 182] stimulation, the majority published in the past 10 years, have all demonstrated that the infants' functional networks activated by these sensory stimuli are similar to that of adults. Marked differences to that of the fully-developed adult brain are also noted, including that regional specificity of visual responses to faces over other object scenes, which are present in adult functional activity, were not evident in infants aged $4-6$ months [180], and how certain patterns of activity relating to the complex psychological constructs around pain experience in adults, such as expectation and contextualisation, are not observed in infants [177, 178]. Determining the timescale over which these mature adult-like features of stimulus-related activity emerge, and identifying the key developmental stages for maturation of functional network organisation, is an unanswered field of research, requiring detailed longitudinal studies from infancy, through childhood, and beyond.

1.4.2.2 Diffusion MRI

Before the recent advancement of non-invasive brain imaging techniques, understanding of structural development of the human brain was based on *post-mortem* studies. This is a limited and sub-optimal approach, since the sample available for neonatal *post-mortem* investigations are more likely to represent abnormal development,

and it is not possible to study functional correlations of brain structure this way. However, these histological examinations have provided a solid groundwork and complementary information to interpret non-invasive MRI data to study white matter (WM) development [164, 183]. Diffusion MRI (dMRI) scans can be acquired to examine tissue microstructure and anatomical connectivity of the brain non-invasively [184].

The principle of dMRI is to use magnetic field gradient sequences to measure the diffusion of water in brain tissue, which is encoded in the MRI signal through spatial and temporal variation in the magnetic field [185, 186]. Two common quantitative measures of diffusion which can be extracted from dMRI are fractional anisotropy (FA) and mean diffusivity (MD), which measure the directionality of the diffusion and the average amount of diffusion over all directions, respectively. High FA represents greater anisotropy, or more directional diffusion, whilst low FA represents more isotropic diffusion. MD increases with greater mean diffusion over all directions. The diffusion of water is influenced by tissue structure, whereby diffusion is greater along the direction of tissue fibres (axial diffusivity) and impeded perpendicular to the direction of tissue fibres (radial diffusivity). Images acquired from dMRI can therefore be used to measure correlates of brain microstructure, primarily spatial anisotropy of diffusion and rates of diffusivity in the brain, which relate to tissue organisation (WM bundles) and tissue myelination. dMRI investigations are typically focused on WM, where axons are concentrated, but there are also advanced dMRI techniques which allow investigation of grey matter [184, 187]. The anatomical connectivity of WM fibres can be mapped using *tractography* algorithms, which reveal the architecture of WM tracts connecting different regions in the brain. Tractography is the only way to investigate structural connectivity of the brain *in vivo* [184], and allows investigation of changes in structural connectivity and relation to FC.

Using dMRI, all major adult WM bundles can be identified in the infant brain, although they are less myelinated at this stage [183, 188, 189]. dMRI studies have

detailed the architecture of the premature brain, demonstrating that at 30 weeks PMA, structural connections are modularly organised much like in adulthood [183, 190–192]. From prematurity to term age, neonatal structural networks become more densely connected and clustered [183, 193, 194], with more efficient connections [194]. There are further significant increases in global efficiency and mature organisation of the structural connectome over the first two years of life [183, 195]. Increases in global and local efficiency during early infancy are linked to WM maturation, which provides a more compactly organised network and efficient information-transfer [190, 194]. On a macrostructural level, brain volume also increases rapidly to reach 90% of the adult size by 2-years of age [196]. Ouyang et al. provide a detailed review of distinct foetal and infant brain maturation processes from conception to 2 years post-birth, and several months beyond [197]. Much like infant functional development, infant structural connectivity develops heterogeneously across brain regions [191, 194]. Regions with greater rates of developmental changes in structural efficiency during the neonatal period are consistent with the literature surrounding functional development. For example, Zhao et al. found significant increases in structural efficiency in the left and right precentral and postcentral gyrus from late preterm to term-age [194], regions which are essential for primary sensorimotor functions, which, as discussed earlier, fMRI has identified as an early maturing functional network in the preterm infant [156, 162]. A recent review of foetal and infant brain development explains that structural architecture involved in primary functions are more mature in the neonatal brain, and this architecture evolves to support high-order functional networks with increasing PNA [191].

Structural connectivity studies have also revealed that short-range connectivity increases faster postnatally than long-range connectivity in preterm neonates [194], which is similarly reflected in functional connectivity literature [191]. However, studies of structural development following preterm birth, whether longitudinal or cross-sectional comparisons to term-born neonates, do not necessarily describe the typical in-utero structural development of term neonates. Evidence of marked differences in functional development in preterm compared to term infants should

provide a clear warning against any assumption that preterm development mirrors in-utero development, and signal that the extrauterine environment can dramatically impact trajectories of infant brain development. Although the global organisation of structural networks does not appear to be disrupted by prematurity, except in cases of brain injury and lesions, prematurity has been shown to affect the quality or strength of structural connections [191]. For example, Ball et al. demonstrated that prematurity is related to altered strength of short- and long-range connectivity and reduced network capacity [192]. Recent advances in foetal neuroimaging techniques and improved resolution could provide key insights into the structural development of the term neonate compared to age-matched preterm infants.

Insights from Foetal dMRI

Foetal MRI studies, especially involving large datasets, are rare, since the technical challenges are even greater than ex-utero infant MRI scanning. These challenges include the constant motion of the foetus which causes image artefact, and effects of maternal tissue which need to be accounted for in image processing. A recent study by Sian Wilson et al. used state-of-the-art foetal MRI techniques to map WM development in-utero from 22-37 weeks gestation [198]. The foetal MRI data analysed by the authors in this study was acquired under the dHCP project, and comprises the largest and highest-quality open foetal MRI dataset to date, with over 100 healthy subjects with high-quality images. An impressive $\tilde{7}5\%$ of all 151 scanned foetuses were able to be included after uncorrectable motion artefact rejection. By comparison, an earlier high-quality foetal dMRI study by Zanin et al. in 2011 was only able to include 17 (28%) of 61 healthy foetuses' dMRI scans after rejection based on excessive motion corruption [199]. This novel high-resolution dHCP dataset allowed the authors to examine normal WM tract development during the second and third trimesters with an unprecedented degree of sensitivity and specificity. They focused on the in-utero developmental trajectories of five WM tracts which have been implicated in neurodevelopmental disorders under abnormal development. Interestingly, non-linear maturational changes in FA and MD were observed in most

tracts, whilst on a whole-brain level FA increased linearly with GA. These non-linear trajectories were a novel finding, since most earlier studies only captured development during the later stages of the third trimester, and observing trends only from 28- or 30-weeks onwards would appear to show a linear relationship as these ages coincide with the turning point in the non-linear trajectories. Although, a limited number of earlier studies did also suggest a non-linear relationship between MD and GA during foetal development [199, 200]. The direction of changes in MD and FA within each tract were inverted, which is expected since the same underlying organisational changes are reflected in these metrics by changing in opposite directions. These dMRI data cannot themselves explain the neurobiological basis for the non-linear trends in FA and MD with age, nor the heterogeneity observed between different tracts. However, when considered in relation to histological findings, likely causes can be postulated. Initial early increases in MD could be explained by the early stages of tissue organisation, which initially increases overall diffusion since diffusion is less hindered by coherent fibres compared to an entangled tissue state [198, 199]. However, at a critical point around 30-weeks, further tissue organisation, pre-myelination processes, and increased production of extracellular glial matter serves to decrease overall diffusivity by reducing diffusion in the extracellular matrix and directing diffusion predominantly along fibre bundles in a more controlled manner [199, 201], which is observed as a 'linear' decrease of MD with GA from 30-weeks [198, 199]. As with MD, the turning point in the non-linear relationship of FA with GA appears to be around 30 weeks in Wilson et al.'s data, which corresponds to the age at which several developmental processes which could affect FA transition, such as the generation of immature oligodendrocytes and their organisation around axons to form a premyelin sheath [201]. Linear increases in FA at later GA are likely due to increases in axonal growth and insulation by developing premyelin sheaths, improving the directional diffusion along the fibre tracts rather than escaping through the walls of more porous unsheathed axons [199, 201]. Detailed in-utero studies of healthy WM development are imperative for providing a template of normal trajectories with which abnormal WM development can be identified, whether in-utero or in

ex-utero premature infants. This could be used to identify infants with or at risk of neurocognitive impairments at an earlier stage, and direct care to support these infants. The dHCP foetal MRI data has only recently been published, and it remains a relatively untapped source of potential knowledge to provide further insights into early human brain development.

1.4.3 Common Challenges in Neonatal Brain Imaging

There are a number of challenges to conducting neonatal neuroimaging research, particularly MRI studies. Many of these challenges result in small sample sizes in neonatal neuroimaging research. Reasons for this include difficulties recruiting for MRI research due to (i) the importability of MRI scanners and need to transport the infant off the hospital wards or co-ordinate with patients to visit from home, (ii) parental concerns around safety and (iii) limitations on infant eligibility for studies, such as strict age limits due to the rapidly developing brain or stringent health requirements. MRI scanner time is also expensive, and requires co-ordination of researchers, medical professionals to monitor infant vital signs, MRI radiographers and technicians to conduct the scans. Scan times for infants are often longer than adults to ensure infant comfort and allow for waking and feeding breaks. Specialist equipment and acquisition methods are also required to keep the infant safe and comfortable throughout, such as providing adequate ear protection and ensuring scan noise and length is kept to a minimum [202]. Collectively, this resource-intensive set-up makes recruiting, co-ordinating, and funding larger sample sizes very difficult. Although small sample sizes in infant neuroimaging research are commonly criticised [203], these studies are worthwhile. Since these studies are resource-intensive and costly, using small samples for exploratory research, generating hypotheses and piloting high-risk designs, is a valuable use of time and money. This approach directs resources wisely by developing well-designed large-scale studies built on prior knowledge [203], and is important for protecting infant safety if using new sequences and imaging protocols. Large infant cohorts are not practical to recruit for every research question, but a combination of small-scale studies and larger

multi-centre studies can be used to generate and investigate specific hypotheses. In order to utilise resources effectively, researchers need to embrace open research and data sharing practices. By sharing infant neuroimaging datasets we prevent waste of available resource and information, and enable the acceleration of research development and our understanding. Neonatal MRI data is a valuable resource which should be used to its full extent. Initiatives like the dHCP which provide unprecedented large open infant neuroimaging datasets help to combat these issues and advance neonatal research.

Infant MRI scans are typically performed while participants are in a natural sleep slate [203, 204]. This beneficially reduces motion and indicates the infant is comfortable, unless they awake. If infants awake and become unsettled, time is required to try and re-settle the infant to continue scanning. The feed and swaddle technique is a safe, low-cost method of reducing infant motion comfortably, and promoting sleep during the scan, although this method becomes increasingly ineffective in older infants [202, 204]. Whilst infant sedation is possible for clinical MRI it is questionable for research purposes, as infant safety is of the utmost importance. Therefore, acquiring high-quality neuroimaging data becomes difficult beyond the first postnatal week as neonates become increasingly active and images become further corrupted with motion artefact. The fact that neonates are typically asleep may affect the nature of fMRI activity recorded, but this remains the safest and highest-quality method to acquire usable fMRI data, since fMRI is easily corrupted by motion artefact [203]. Acquiring infant MRI data in naturally-asleep infants is still demonstrably useful, with many research papers dedicated to identifying patterns of typical and atypical development, and demonstrating predictability of long-term outcomes from MRI data acquired in naturally-asleep neonates. Some sequences are particularly sensitive to motion, so that even with all reasonable motion-reduction measures in place, motion artefact is too great which leads to very high rejection rates and makes analysis impractical [e.g., for Arterial Spin Labelling, 205].

There are advanced neuroimaging methods which would be very informative in neonatal research, but challenges pose a great barrier to implementing and adopting these methods. For example, simultaneous EEG-fMRI combines the benefits of EEG and fMRI to provide brain activity data with both high-quality spatial and temporal resolution. However, it took nine years since its inception before an infant EEG-fMRI study was conducted [206], and it remains the only published infant EEG-fMRI study 5 years later. This is doubtless due to the complexity of collecting MRI data alone in infants, without adding to the challenge by combining MRI with an additional neuroimaging technique. Another area of interest is to perform awake stimulus-fMRI in infants, e.g., to investigate visual and auditory responses in the awake infant. This also proves prohibitively difficult, since the infant is unlikely to be content with the noisy MRI environment to (i) complete the scanning session restfully awake, and (ii) remain still enough whilst awake to acquire high-quality MRI data. Stimulus-fMRI is performed in sleeping infants since they are more amenable to scanning. If neuroimaging acquisition and pre-processing methods improve enough to allow scanning protocols which provide high-quality task-activation data in the awake and settled infant, this would be really interesting for advancing our knowledge of functional brain development and comparing to task-activation during sleep.

Other future areas of interest relate to hardware improvements. One potential area is acquiring high-field images using the recently available 7 T MRI, compared to current standards of $1.5 - 3$ T MRI. The Wellcome Centre for Integrative Neuroimaging at the John Radcliffe Hospital (Oxford) site installed a 7 T MRI scanner in 2011 (`https://www.win.ox.ac.uk/about/facilities/7-tesla-human-mri-scanner`). Higher resolution images provided by 7 T scans would enable us to extract fine structural detail of the infant brain, but first, safe scanning protocols for neonates and infants need to be developed. Another significant improvement is to be found by using neonatal and infant-specific head coils [207]. Neonatal head coils are far smaller than the standard MRI coils which are designed for adult anatomy. These specialised coils improve signal:noise (SNR) significantly by reducing distance

of the coil from the infant's head. The signal attenuates quickly with increasing distance of the coil from the head, so the coil size is one of the main factors influencing SNR [207]. Using a smaller infant-specific coil sounds a simple improvement, but the barrier to uptake includes the great expense of MRI equipment, as well as time and resource required to learn how to use a new coil, to refine protocols in order to acquire optimal images with the new set-up, and to receive a worthwhile return on the investment.

1.4.4 Neonatal Brain Imaging to Measure Pain

Observing brain activity and features of brain maturation, rather than observing outwardly visible facial features, could be an objective method to better understand neurobiology underlying infant pain sensitivity, and develop methods to tailor pain treatment to the individual to reduce adverse effects associated with pain exposure and inappropriate analgesia. Cortical measurement of noxious-evoked activity as a pain assessment tool has high face validity (Table 1.1). Since pain is mediated by the brain, logic follows that measurement of noxious-specific brain activity can reflect pain experience. The first measurements of cortical activity in response to acute noxious stimuli in neonates used fNIRS, measuring cerebral haemodynamics, to identify noxious-evoked haemodynamic responses in the somatosensory cortex [208, 209]. However, fNIRS has lower temporal resolution than EEG and lower spatial resolution than fMRI, therefore these latter two techniques have the potential to provide deeper insights in neonatal pain research. As with other streams of neonatal neuroimaging research, EEG is popular in the pain field, due its portability, low-cost and ease-of-use in clinical settings, where EEG is already used routinely. Although fMRI is a more restrictive technique which is not feasible for routine clinical use, by furthering general understanding of the infant pain experience through fMRI research, it may be possible to develop detailed pain models which can ultimately lead to better clinical outcomes. For example, by using fMRI models in clinical trials for pain treatment, or by identifying pain-intensity encoded brain signals which can be measured with more portable techniques, or investigating factors which influence

pain sensitivity or pain management outcomes that can be translated into better clinical management.

The use of neuroimaging techniques to study infant pain and pain treatment is becoming increasing prevalent, including to study the impact of pain, analgesia and anaesthetics on neurodevelopmental outcomes in later life [210]. Multimodal MRI techniques have also been recognised as pain-specific and anaesthetic-agent specific by paediatric professionals [210].

1.4.4.1 EEG Studies of Noxious-Evoked Brain Activity

The first EEG study to measure noxious-evoked brain activity in infants was published in 2010 [211]. This study identified a noxious-specific event-related potential (ERP) that discriminates noxious from non-noxious tactile stimulation. EEG studies have determined that extremely preterm infants display delta brushes, the non-specific neuronal bursts, to noxious and tactile stimuli, which gradually transition to modality-specific responses with increasing age [212, 213]. Early investigations reported that noxious-specific responses start to dominate in the late preterm, at approximately 34 weeks [212, 213], at the same time as non-noxious tactile-specific responses emerge [213]. These studies identified minimal to non-existent noxious-specific brain activity observed in infants at less than 32 weeks [212, 213], and at ages younger than 32 weeks there was substantial overlap between non-noxious tactile and noxious-evoked cortical responses [213]. Interpreting this finding is difficult, perhaps this implies that non-noxious touch and noxious stimuli are both unpleasant sensory experiences in extreme prematurity, but early skin contact and positive affective touch has been linked to physiological stability and pain reduction [214–216], whilst pain exposure has been linked to negative outcomes [56, 77, 79, 214]. Noxious-evoked responses increase in magnitude with increasing PMA [217], which matches infants' increasing behavioural discrimination of noxious and non-noxious stimuli with increasing age [213]. Research suggests that sole behavioural discrimination of non-noxious tactile and acute noxious stimuli is apparent from approximately 34 weeks, at the same time as modality-specific

cortical activity appears to emerge [213]. Recent research by van der Vaart et al. demonstrated that noxious-evoked brain activity undergoes an additional transitory stage between $30-33$ weeks PMA, with a morphologically distinct EEG pattern including a negative peak $400-500$ ms after noxious stimulation [218]. This is different to the previously reported modality non-specific delta brushes in infants <34 weeks [212], and the characteristic template of noxious-evoked activity for infants >34 weeks involving a positive peak $\sim 400-700$ ms after noxious stimulation [211, 219]. Importantly, noxious-evoked brain activity was discriminable from non-noxious activity even in these $30-33$ week-old infants, and multi-modal models including EEG activity, autonomic and behavioural responses were able to discriminate noxious from non-noxious stimuli with nearly 80% accuracy in infants as young as 28 weeks PMA [218]. This is promising for the implementation of cortical activity as a surrogate measure of pain, complementing behavioural and physiological signals, to optimise the delivery of pain treatment and validate analgesia for the most vulnerable premature infants.

The noxious-evoked EEG response from 34 weeks PMA was characterised in detail by Hartley et al. in 2017 [219]. The authors identified and validated a template of nociceptive brain activity from infant noxious-evoked EEG activity, that can be used to quantify infant noxious-evoked responses to acute stimuli [219]. The detailed publication included a study which demonstrated the template measure's ability to discriminate between noxious and non-noxious activity in both term and preterm infants ($34-43$ weeks PMA), correlation of the template measure with behavioural indicators of pain, as well as the template measure's responsiveness to pain relieving interventions [219]. The authors shared this template freely as a supplementary file with the original publication, and their methods, and other similar EEG ERP approaches, have been implemented in a variety of settings to study infant noxious-evoked activity. This approach has been used to investigate a range of acute tissue damaging clinical procedures including heel lancing [211, 216, 219, 220], vaccination [122, 221], and cannulation [222], as well as mild non-tissue damaging experimental noxious stimuli [215, 219, 223]. It has also been

used to evaluate the efficacy of numerous pharmacological and non-pharmacological analgesic interventions, including morphine [106], paracetamol [68], topical local anaesthetics [219], sucrose [49, 220], kangaroo care [216, 224], breastfeeding [220], and gentle touch [122, 215]. In addition to age [212, 217] noxious-evoked EEG response magnitudes are stimulus intensity encoded [225], and are affected by inflammation [68], mode of delivery [223], sex [226], stress [227] and premature birth [228]. EEG is a highly practical technique for infant pain assessment which can be implemented cot-side, is familiar to clinicians, and provides high temporal resolution.

1.4.4.2 fMRI Studies of Noxious-Evoked Brain Activity

fMRI provides complementary insight into cortical processing of pain in infants compared to EEG studies. fMRI measures haemodynamic responses rather than electrophysiological activity, provides whole-brain coverage with access to deeper sub-cortical brain regions and greater spatial resolution than EEG, although with lower temporal resolution. The spatial resolution afforded by fMRI allows us to compare cortical activity between adult and infant populations which offer insights into potential similarities and differences between adult and infant pain experience. This method could also be used in research settings to probe the neurophysiological basis for inter-individual variability in pain sensitivity, and lead to greater understanding of factors underlying the experience of infant pain and how to treat infant pain.

The first infant noxious-evoked fMRI study was published by Williams et al. in 2015 [176]. This established an experimental paradigm for studying infant noxious-evoked activity in MRI scanners using experimental noxious stimuli [176], since MRI scanners are not a safe nor practical environment to conduct clinical procedures. This was an exploratory study demonstrating the feasibility of collecting infant noxious-evoked stim-fMRI data. Their results confirmed prior EEG findings [225], that infants display intensity-encoded nociceptive responses [176]. They also demonstrated that noxious stimulation evoked activity in several brain regions including primary and

secondary somatosensory cortices, and the pattern of noxious-evoked activity differed from that evoked by non-noxious tactile stimulation [176].

Soon after this proof-of-concept study, Goksan et al. published a landmark fMRI paper revealing the spatial overlap between noxious-evoked activity in adults and infants [177]. This study established high similarity between regions of noxious-evoked brain activity between 10 adults and 10 term neonates. Neonates exhibited noxious-evoked activity in 18 of 20 adult-activated regions, implying that nociceptive processing in term neonates is well-developed, and infant pain perception is likely similar to adult experience. The regions of noxious-evoked activity in term neonates included both primary and secondary sensory networks, as well as networks associated with emotional aversion, including the anterior cingulate cortex [177]. This study also illustrated that infants have heightened nociceptive sensitivity compared to adults, as similar functional activation intensities were observed between neonates and adults with a quarter of the stimulus force applied to infants [177]. The two regions which were activated in adults but not infants were the amygdala and orbitofrontal cortex, both of which are associated with emotional processing. The orbitofrontal cortex is associated with anticipation and reward, whilst the amygdala is involved in fear and anxiety. Their absence suggests that pain processing in term neonates is not yet intertwined with these negative emotional states and they are unable to anticipate the stimulus in the experimental context. Instead, these results suggest relative immaturity of strong emotional aversion to pain, which likely develops as infants age and learn the ability to contextualise and fear pain. As noted in the updated IASP definition of pain, pain experience is impacted by life experience [12]. All neonates were scanned within the first postnatal week, where ex-utero life experience is still very limited, and represent a healthy term cohort with minimal pain exposure [177]. The timescale with which nociceptive processing matures to involve higher-order emotional networks, and the effect of cumulative pain exposure on this development, has not been studied. However, this result is not definitive, and could just be due to insufficient statistical power to detect a true effect in the amygdala and orbitofrontal cortex, or the noxious stimulus could simply be too mild to elicit a pain response

which would pass the pain threshold to involve the relevant negative emotional states. These experimental stimuli are only considered ethically acceptable due to their mild nature, the fact that they do not elicit behavioural or physiological distress signs in infants [225], unlike acutely painful clinical procedures, and EEG measures of noxious-evoked activity are significantly weaker in response to these experimental stimuli compared to clinical skin-breaking procedures [219, 225]. Although, even if these regions are implicated in more intensely painful events in infants, these results suggest that emotional processing of pain may become more prominent with ageing, since these regions were still activated in adult responses to mild nociceptive stimuli [177]. We lack evidence to test whether the amygdala and orbitofrontal cortex would activate in response to more intense nociceptive stimuli on par with clinical skin-breaking procedures in infants. Such a study is not currently possible with fMRI due to ethical considerations, and inability to collect MRI data during painful clinical procedures.

Additional evidence for differences in emotional experiences of nociception in adults and infants was provided by Duff et al. in 2020 [178]. This study included 10 adults and 37 term neonates scanned within 11 postnatal days. The authors found that an adult nociception-specific intensity-encoded neurologic pain signature [229] was expressed in term neonates in response to experimental noxious stimuli, but there was no expression of an adult pain signature which represents stimulus-intensity independent emotional encoding of pain perception [14, 178]. This reinforces the interpretation that neonates possess core nociceptive pain processing, which matches adults, but with an absence of adult experiences of expectation, fear, and contextualisation of pain in the neonatal noxious response [178], as suggested by the original study by Goksan et al. [177].

The number of fMRI studies investigating neonatal nociception remains very low, with only 5 published to date [176–178, 230, 231], and an additional two studies presenting optimal methods for neonatal noxious-evoked stim-fMRI research [232, 233]. The number of infants with noxious-evoked stim-fMRI data included in each

of these five studies ranged from just $1-37$ (median=13). Despite the value of high spatial resolution fMRI data, compared to EEG or fNIRS, recruitment and research growth in this area is slow, due to difficulties with data acquisition and its high cost (general challenges are discussed in Section 1.4.3). Recruitment is particularly difficult in the experimental nociception paradigm, as parents are understandably reticent to consent to studies which require their infant to receive experimental noxious stimuli, even with evidence that these stimuli activate nociceptive networks but are too mild to cause emotional or behavioural distress associated with pain experience [176, 177, 225]. In the past few years, methodological developments have enabled higher-quality infant nociception fMRI experiments. Goksan et al. determined optimal echo times for noxious stimulus fMRI acquisition [232] and Baxter et al. developed an optimised pipeline for processing and analysing neonatal noxious stimulus fMRI data [233]. This optimised pipeline provides robust methodology for researchers to study neonatal noxious-evoked activity, with greater accuracy, confidence and efficiency of analysis. These recent developments will hopefully encourage more widespread fMRI studies of infant nociception by demonstrating how to achieve high quality data and reducing barriers to entry.

1.4.4.3 Future Directions and Potential for Neuroimaging Studies of Infant Pain

EEG measures of cortical activity have been used in clinical trials to quantify pain response for the purpose of testing analgesic efficacy [106], and are being used for ongoing clinical trials testing pain-relieving interventions [131, 224]. However, this is still a relatively untapped methodology, and more clinical trials for analgesics could be conducted which incorporate noxious-evoked activity as an outcome measure. Furthermore, each of these ongoing trials focus on acute procedural pain, but there is also the potential to utilise cortical activity measurement to assess prolonged pain, such as post-operative pain. Group measures of pain response ignore individual infant variability in pain sensitivity [234]. EEG could be used to assess infants' baseline nociceptive sensitivity [122], in order to optimise clinical trial designs and personalise appropriate provision of analgesia in clinical practice for painful

procedures. Additionally, machine learning (ML) techniques could be used to improve the assessment of individual infant pain sensitivity, to predict optimal analgesic provision, and tailor pain management strategies [235]. Part of this thesis focuses on developing an ML model to predict infant nociceptive sensitivity from resting-state brain activity (Chapter 3), suggesting that we could infer an individual infant's analgesic requirements from non-specific measures of their development and contextual factors [231]. Whilst such measures of resting-state brain activity that inform these prediction models could not be routinely measured in clinical practice, knowledge ascertained from these models could be translated to improving clinical pain assessment and analgesic provision.

Using or testing EEG as a pain-assessment tool also has disappointingly limited uptake in general research. A shallow search of the literature shows that infant pain research incorporating EEG measures are globally-limited and restricted to a few research groups. The majority of this research originates from European and North American labs, including our own research group in Oxford, UK [219], and our ongoing collaborations with researchers in Belgium and in France, along with independent research in London, UK [212, 227], Switzerland [223], Canada [220, 224], and the USA [236], with one exception from Israel [237]. There is much work to be done in broadening the application of this technique across the continents, particularly when cot-side EEG has been implemented in NICUs on an international scale. In Chapter 2 I investigate the validity of using EEG activity for infant pain assessment.

A noxious-specific EEG template for premature infants aged 30-33 weeks has also only recently been published [218]. This novel tool opens possibilities for more detailed studies into factors affecting pain sensitivity and strategies for pain management in these vulnerable young infants. The majority of neonatal pain research has focused on individual metrics and pain scales to quantify pain response, but research has demonstrated that multimodal measures combining cortical activity, behaviour, and physiology are better able to discriminate noxious

response [121]. Future research should validate and develop these multimodal measurement techniques further, for both research and clinical applications.

The use of fMRI to study infant noxious-evoked activity is even more geographically limited, with all five published studies conducted by just two research centres in the UK, from London [176] and Oxford [177, 178, 230, 231]. Whilst fMRI isn't viable for clinical pain assessment, as with EEG, it is still a valuable research tool to progress our understanding of infant pain experience and factors which modulate pain. There is potential to construct detailed assessment of infant noxious-evoked activity using fMRI data which could provide clinically translatable insights. fMRI could be used to identify and develop likely analgesic drug candidates, by comparing the action of a drug candidate on the central nervous system to that of known effective analgesics [238]. Proof-of-concept for this approach has been demonstrated in adult fMRI studies [238], which could accelerate the early-stages of drug development by pre-selection of likely candidates for clinical trials. This has the potential to benefit infant pain management, as any methods that could increase the likelihood of clinical trial success would reduce the investment risk for pharmaceutical companies, and cortical activity is a promising objective method to measure drug efficacy in a population who lack verbal report [66].

1.4.5 Neonatal Brain Imaging and Long-Term Health

Functional brain activity and brain structure imaged in early life, including the neonatal period, has been used to discriminate between typical and atypical development and has been implicated in long-term health outcomes. Studies relating neuroimaging data and health can follow longitudinal or cross-sectional designs, and results from different studies can be linked to identify influencing factors. For example, associations can be found between the effects of early-life stress on brain development in infancy and subsequent mental health issues in adulthood in different neuroimaging studies [239]. Cumulative pain exposure in preterm infants has been shown to disrupt typical development of DMN connectivity in infancy; connectivity which has been associated with poor mental health outcomes in adults [239]. Gao

et al. also demonstrated that infants with low socioeconomic status had disrupted FC of the DMN in the first year of life, which could contribute to differences in long-term cognition and health outcomes [169]. Gui et al. also showed that it was possible to predict neurodevelopmental outcomes from neonatal brain tissue volumes in a longitudinal study of preterm infants [240]. Lower growth rates of brain tissue from birth to TEA were associated with lower GA. Tissue volumes at birth and TEA predicted motor outcomes at 2-years, and tissue volumes at birth and TEA and lower tissue growth rates predicted cognitive outcomes at 5-years. However, parental socioeconomic status remained the most important factor in predicting cognitive outcomes at 2- and 5-years. Thus suggesting, unsurprisingly, that better support needs to be provided for families of lower socioeconomic status.

Infant fMRI research could be leveraged to identify infants at risk of adverse outcomes from early-life brain function. Although fMRI is expensive and impractical to be used as a routine biomarker of long-term health, even in vulnerable populations such as preterm infants, these data could be used to identify more suitable markers of at-risk infants. For example, if fMRI data are linked to associated physiological or behavioural differences in infancy and early development, or help to uncover environmental factors associated with atypical brain development which increase long-term health risks. This could be used to target long-term follow-up and interventions for those individuals most likely to need additional support. Neuroimaging studies of longitudinal brain development could also identify timescales over which at-risk children's development departs from healthy development, which can be used to target timely interventions and prevent these adverse outcomes. Early prevention and management strategies increase quality of life for targeted individuals and reduce the ultimate burden on healthcare systems.

Similarly, neonatal EEG has been used to study typical and atypical development, and its relation to long-term neurodevelopmental outcomes. The evolution of EEG characteristics are well-characterised with respect to GA, PMA, as is the effect of prematurity on typical EEG signals, as discussed in Section 1.4.1. One recent

example of research leveraging the powerful insight EEG characteristics provide into developmental (ab)normality was presented by Pillay et al. [153].

1.5 The Importance of Reproducibility in Neonatal Research

As much as research and its publication are beneficial to the progress of scientific knowledge, spreading information and enabling others to develop that knowledge in additional research or practical applications, there are well-known issues with the state of publishing and academic research culture. Academics live in a world where they are required to publish regularly to sustain funding and keep their jobs, often living on short fixed-term contracts, and this pressure can distort the field of published research. These issues amount to problems with reproducibility, publication bias and a number of questionable research practices. Several of these can be ameliorated by changing the way we typically conduct and publish our research.

Due to the flawed reward structure of academia, novel and innovative research is more likely to be published and valued for job security, whilst reproducibility studies are ignored or not attempted due to limited resources and no guarantee that the outcomes will be valued by publishers and funders [241, 242]. This conflicts with best scientific interest, which would be to confirm, or improve the precision, of the 'known truth' to achieve greater and more accurate understanding of our universe. This also conflicts with the clinical and public interest in providing medical care based on detailed and accurate information, rather than being misguided by spurious results. In any study there is always a statistical chance that published results are false negatives or false positives, even without considering human error or poor research practice, and only through repeated and independent research can we build confidence in any finding. Even if previous results are not a fault of statistical chance, further research with different methods or in different populations is also necessary to define the boundaries in which prior results hold. For example, medications

may have different effects depending on genetic factors, such as reduced effect of anaesthetics in female red-heads [243, 244] or differences in cardiovascular drug response in different ethnic groups [245], and presentations of medical conditions may also differ between groups, such as the difference in typical signs of heart attack in men and women [246], with implications for clinical evaluation and public health messaging. Many of these examples tie in to systemic issues of sexism and racism in medical science developments, and many examples of flawed technologies which perpetuate these harmful biases can be found in use in society, including technologies which do not work for darker skin tones [247], and gender and racially biased algorithms [248] used in judicial [249] and health care systems [250], which highlight just how damaging dismissing reproducibility research can be.

1.5.1 A Lexicon of Reproducibility and Replication

Testing the reproducibility or replication of research findings is fundamental to the honest and successful advancement of science. Given the relatively recent drive for reproducibility and replication studies in the social, biomedical, and neural sciences, there are mixed interpretations of the different meanings of replication and reproducibility, which are not consistent with the use of these terms in physical sciences such as physics and chemistry. Attempts have been made to standardise the use of these terms; to clarify and classify different meanings of reproducibility and replication. However, different recommendations have recently been published with this aim [251, 252]. Goodman, Fanelli and Ioannidis (2016) suggest in Science Translational Medicine that *reproducibility* should be used as an umbrella term within which fall the more specific descriptors of methods reproducibility, results reproducibility, and inferential reproducibility [251]. In their proposed definitions, one common understanding of 'replication' as a corroborating study which follows similar methods and produces the same results as an original study, using independent data, is equivalent to 'results reproducibility'. 'Methods reproducibility' is the ability to replicate original results exactly, using the same data and available tools, by following the published methods as exactly as possible given the completeness

of the published information. Meanwhile, 'inferential reproducibility' concerns whether inferences made from independent analysis of the same data as an earlier study, but not necessarily using the same analysis methods, are in agreement with the original interpretation(s). The authors also distinguish from these terms that generalisability is "the persistence of an effect in settings different from and outside of an experimental framework" [251]. However, Nosek and Errington (2020) provide a different perspective on terminology in PLOS Biology [252]. These authors suggest that *replication* should be used in a more general sense than strictly describing the process of repeating a prior study's methods and testing whether results match the previous finding(s). They argue that a precise definition such as this unnecessarily restricts research design; distracting from the underlying purpose of replication to advance knowledge. It is a counterproductive waste of time and energy if researchers are concerned with whether their methods are similar enough to an original study to qualify under a specific definition of replication, whether to describe it in terms such as 'direct replication' or another precise category, rather than whether their study outcomes will provide helpful information to advance scientific knowledge [252]. They suggest applying a general meaning of a replication study when "outcomes consistent with a prior claim would increase confidence in the claim, and outcomes inconsistent with a prior claim would decrease confidence in the claim" [252]. In this sense, Nosek and Errington's suggestion for general 'replication' includes the umbrella term of 'reproducibility' suggested by Goodman, Fanelli, and Ioannidis, since all the subtypes of methods, results, and inferential reproducibility would fall under the Nosek and Errington ideology of replication [251, 252]. However, Nosek and Errington's replication definition is even more inclusive, also considering as replication studies those which neither use the same methods nor data as the original study, but test the same hypotheses, test outcomes predicted by an original model, or could otherwise change the confidence in the original claims. In order to qualify as a replication, the outcomes of the study must be able to both increase or decrease confidence in the original claims, depending on the outcomes. In a definition consistent with Goodman et al.'s description of generalisability, all replication tests

are generalisability tests, since all replication studies will differ from the unique conditions of the original, and thus demonstrate new conditions in which the original results do or do not hold. All replication studies help to improve a framework of understanding by discovering the breadth of conditions for which an existing theory holds. However, not all generalisability tests are necessarily replication tests. A study would only be a generalisability test, but not a replication test, if a 'failure to replicate' would be interpreted only as defining boundary conditions for a model rather than questioning the original claims themselves.

Following both of these publications, I will use 'replication' and 'reproducibility' in their broadest senses, as defined by each of these authors [251, 252]. This thesis will not regard replication as restricted to any commonly-understood exclusive definition, which involves using exact data or exact analysis methods to an original study. Replication and reproducibility are overlapping but non-equivalent terms, where reproducibility research is a subset of replication where at least one factor has been kept constant (as much as practically possible) between the original study and the replication study. To summarise, in a reproducibility study, either the same methods/analyses are applied to new data (results reproducibility), the same data is studied from a new angle which tests the original theory (inferential reproducibility), or both the same methods and same data are studied but by new analysts (methods reproducibility). A replication study can include any reproducibility study as defined here, or a replication which tests a theory using both new data and new analysis methods. In the last case, this would not test the reproducibility of the original study as defined by Goodman, Fanelli and Ioannidis [251], but would test generalisability and replication according to Nosek and Errington [252]. All of these reproducibility and replication studies will be, intrinsically, generalisability tests, as even the most faithfully-matched replication study achievable will include some difference to the original study, however minor. Indeed, Nosek and Errington may argue that deliberating whether a study fits within Goodman et al.'s precise definitions for reproducibility is another counterproductive activity. Nosek and Errington suggested the use of replication with an inclusive definition to remind

researchers of the purpose of replication studies to improve our world-models and seek truth for the benefit of society, rather than worrying about minor semantic details which disincentivise scientists and hold back research.

1.5.2 An Outlook on Replication in Science: Problems and Solutions

There is a general lack of replication research and insufficient studies to evaluate the reproducibility or generalisability of individual results in science. The reproducibility movement is not a criticism of the integrity or misconduct of researchers, it is not an accusation that there are a majority of scientists publishing falsified or knowingly misleading results that skew our understanding, but a criticism of the accepted approaches to research, the ways in which research is valued and rewarded, and practices of the publishing industry which amplify biases whilst concealing their presence. It is also worth acknowledging the importance of a positive approach to addressing these concerns without spreading panic and distrust by naming current affairs as a 'reproducibility crisis' [253, 254], which has the potential to create drama and pointed accusations rather than addressing the core scientific issues and encouraging mutually beneficial and collaborative efforts to improve the status-quo [254, 255]. There are four key issues behind irreproducibility; publication bias [242, 253], *p*-hacking [256, 257], hypothesising after the results are known (HARKing) [258], and underpowered research [259–261]. Each of these issues tend to co-exist, and each will be elucidated below.

Publication bias refers to preferential publication of new or 'exciting' results, compared to null effects, or even blaming researchers for 'failing to replicate' an experiment rather than questioning the validity of the original result [242, 253]. The collection of research which remains unpublished, whether by authors' choice or by journal rejection, because they show null or negative results is known as the 'file-drawer problem' [242, 256]. Publication biases suppress the spread of information when exploratory results are not replicated, as negative findings are deemed less interesting or less important by academic journals and peer reviewers [242, 253].

Research demonstrates that these biases arise from a combination of increased likelihood of journal rejection and researchers not investing the time to write-up null results, even with high-quality studies [262]. Expectation that a study demonstrating null results will not be publishable may explain why some research is not even written-up nor submitted before being left to the 'file-drawer'. Researchers are pressured into pursuing positive results in order to meet publication goals to acquire funding and sustain their research career. The emphasis is too often on novel and innovative exploratory research, which has an important place in science, but which becomes problematic when it is pitched as a-priori hypothesis testing, at the expense of statistically powered and tested hypotheses. Fanelli showed that publication of negative findings has decreased between 1990 and 2007, whilst the proportion of papers with an explicit hypothesis test reporting positive findings has increased by over 20%, and that the evidence supports that publication bias is to blame [263]. Over time, these publication biases have presented a distorted view of reality, and propagate a waste of resources when scientists independently try and fail to replicate results in their own lab without a respected and efficient way to share these failed attempts, compared to positive results shared in scientific journals. These scientists could be spending their time and money pursuing more hopeful causes, given a better knowledge of the reproducibility of past results. A lack of reproducibility and resource wastage is not necessarily down to flawed research, but other factors such as insufficient communication, where researchers may not share methods in enough detail to identify the key steps in reproducing their findings. Full transparency of research methods, and mapping the cases where a result may or may not be reproduced, also builds a more complete and nuanced picture of the generalisability of results, including additional factors which influence the outcome. In this way, negative reproducibility is not simply a dismissal of the original findings, but by considering different methods and samples involved, it improves the breadth and depth of knowledge in that area, e.g., by understanding interaction effects and cultural or socioeconomic differences between populations that influence the results.

The remaining issues of p-hacking, HARKing and low statistical power are linked to the traditional pipeline of research and publication. Typically, scientists collect the data, analyse the data, and interpret the data and then write-up their analyses, results and conclusions for publication. This creates room for p-hacking, whereby if the first analysis method did not produce significant results, more analysis methods are tried until a significant result is found [257]. This is an abuse of statistics, since each test performed should include correction for multiple comparisons, but if only one test is reported by the authors, then the plethora of insignificant results compared to significant findings will not be reported or known by the reviewers and readership. This common practice of flexible study design and flexible statistical analyses contaminates the literature with false positives, misleading results and conclusions [256]. Training in many scientific disciplines does not sufficiently cover statistics [264], and therefore many do not understand the problem with this flexible research approach in reducing the value of a so-called 'significant' result determined by p-values.

HARKing is also enabled by typical research pipelines and the order of public distribution of research outcomes. HARKing refers to when scientists falsely increase confidence in their results by presenting hypotheses that match the outcomes of their research as if they were predictions, but which were actually conceived after analysis [258]. Hypotheses should be generated first, analyses should be designed to test these hypotheses, and the results will either confirm the hypotheses or not. HARKing is a case of circular reasoning, where hypotheses are generated from observing a set of data, but the validity of these hypotheses are also evaluated from the same observations of the same data [258, 265]. If hypotheses are generated post-hoc, then pretending they were a-priori distorts the truth. HARKing can hide random effects by presenting outcomes as if they are reasonably testing an expected result, for example by characterising a set of inclusion/exclusion criteria as relevant and justified to a post-hoc hypothesis, when these criteria are actually arbitrary, to present a positive finding [256]. These approaches do not have to be intentionally deceitful on the part of the researchers to occur. Human reasoning is imperfect,

and we are all subject to cognitive biases and imperfect memory which distort our own understanding and recollection of our methods [266–268]. Researchers may truly believe that they would have predicted the observed outcomes even if they did not originally design a study to question that outcome, overestimating their prior expectations [266], or fall victim to confirmation bias, where we are more likely to acknowledge observations which suit our world-view and dismiss evidence which challenges it [267, 268]. Nevertheless, this practice skews confidence in published research findings, and increases the number of false positives in the literature.

Statistical power refers to the probability that a test will reject the null hypothesis when the null hypothesis is false [269]. Underpowered research describes when sample sizes in a study are not sufficient to test the hypotheses due to low statistical power. Insufficient data, or low power, results in high rates of false negatives and false positive rates [270]. More people are familiar with the fact that low power can reduce the chance of detecting a true effect (false negative), than the fact that low power also reduces the chance that a statistically significant result represents a true effect (false positive) [260, 271]. The latter is due to the inter-relatedness of a study's probability to reject the null hypothesis correctly, and the positive predictive value of a study, which refers to the probability that a positive result reflects a true effect [260, 271]. Even studies which do identify true positive results are likely to lead to over-inflated estimations of the effect size when the sample size is small, a phenomenon known as the 'winner's curse' [259, 260]. Replication studies need to account for this effect, and should include more data points than the original study; a replication is not necessarily well-powered to detect a previously-reported finding by just matching the original sample size [272]. All of these phenomena serve to demonstrate the under-appreciated importance of statistical power in research design, and results interpretation, and the ways in which under-powered studies can skew our understanding of significant effects. Low statistical power, p-hacking and HARKing often co-occur, since it is difficult to report a 'significant' effect in an under-powered study without resorting to p-hacking, and using HARKing to justify the chosen methods [273, 274]. This does not mean that small sample size

studies are worthless, they can be equally as informative and important as large-scale studies when conducted with statistical fluency and integrity, particularly in fields where data collection is challenging, or collecting more samples is impossible or unethical [275–278]. The interactions between sample size, results, and interpretation simply need to be considered by researchers when discussing research and planning future studies.

If the analyses are planned carefully from the beginning, including appropriate sample size planning for the hypothesis tests and expected effect sizes, then many of these issues can be avoided. The intention of these discussions around the potential harm of these research practices is not to perform a 'witch-hunt' to expose or criticise offenders, since the likelihood is that most, if not all, researchers have engaged in these practices at some point, unaware of the significance of a flexible approach to analysis [255, 274]. Instead, the purpose of the reproducibility movement was, and is, to spread awareness of these issues so that we can all work together to improve research practices now and for the future [274].

1.5.3 A Case for Pre-Registration

Pre-registration is one process which aims to combat each of the aforementioned biases in publication, reducing questionable research practices and errors arising from flexible analysis methods. Pre-registration differs from the standard publishing pipeline by cementing methods including data collection and analysis plans in a published protocol, prior to carrying out any of the planned research (see Figure 1.1). Pre-registration is similar to the registration of clinical trial designs in advance. Authors design their study in detail, including data collection methods, inclusion and exclusion criteria, sample size planning and the analysis methods they will use. For confirmatory analysis, they will need to state their hypotheses clearly in advance and design their study to appropriately test these hypotheses. Exploratory research can also be pre-registered, where advance registration of decisions, including inclusion and exclusion criteria, demonstrate that these decisions were decided a-priori rather than contributing to post-hoc over-manipulation of the data to suit a narrative.

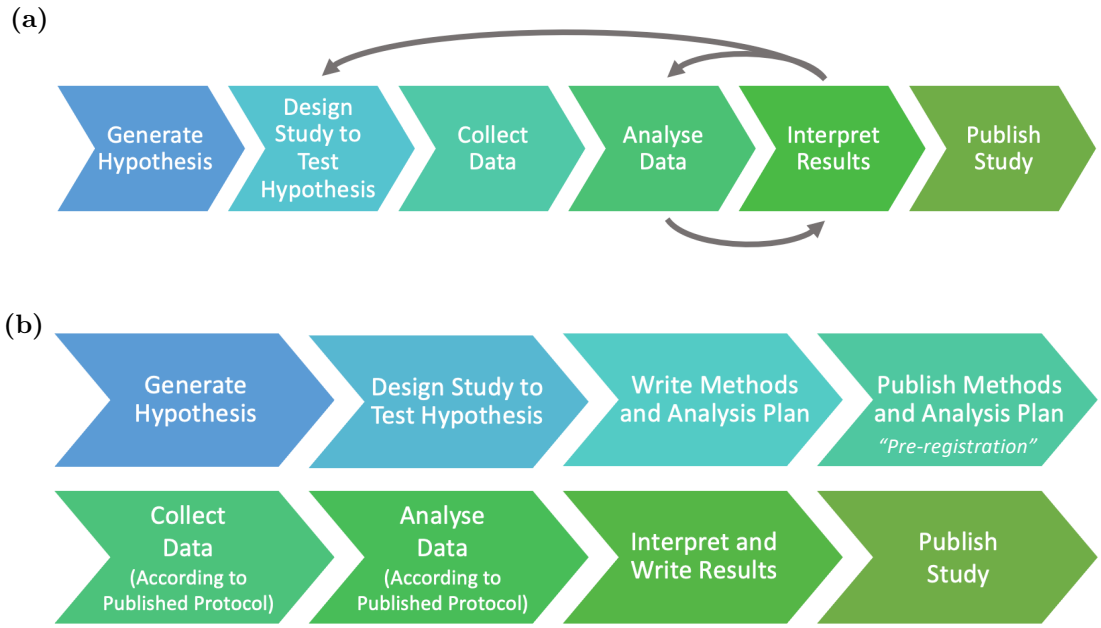


Figure 1.1: (a) Standard publication pipeline. The arrows indicate some of the opportunities to revise analysis methods prior to publication. (b) Publication pipeline involving pre-registration (top row is the pre-registration stage, the bottom row is from pre-registration until final publication). In the pre-registration pipeline, methods and analysis plans are time-stamped and circulated prior to the analysis for transparency and to avoid flexible study designs. In the subset of pre-registrations known as Registered Reports, stage 1 peer review occurs prior to publication of the methods and analysis plan (top row, final box). After stage 1 peer review, the study is accepted for in-principle publication by the reviewing journal. The final study is also subject to (stage 2) peer review before publication in the journal (bottom row, final box), but this will be a lighter review based only on concordance with the published protocol, and valid interpretation of results, rather than chosen methods or direction of results.

Pre-registration does not seek to stifle exploratory research however, and interesting additional exploratory analyses may also become clear only after acquiring and analysing data [273]. Publishing additional analyses is not prohibited by the act of pre-registration, it merely provides full transparency and clarity over which analyses are confirmatory or exploratory, and which analyses were pre-registered and which were not. However, to fulfil all benefits of the process, at a minimum, published results should include all pre-registered analyses, regardless of their outcome. Pre-registration offers a number of potential benefits to the research community, as well as to individual researchers.

From a community perspective, pre-registration combats the file-drawer problem,

prevents HARKing, p-hacking, and underpowered research by encouraging sample size planning for each hypothesis test. By providing a time-stamped record of the research design prior to analysis, the community have awareness into when decisions were made, and can discern more easily whether issues of HARKing or p-hacking occurred. By comparing the final publication to the pre-registered plan, they can be confident that these questionable research practices were avoided, and the results can be viewed with increased credibility. If readers wish to replicate the results, or to design a new research study based on an honest pre-registered study's findings, their efforts are more likely to be meaningful. This process can therefore lead to increased efficiency and productivity in science by reducing the avoidable questionable research practices which misdirect research and waste resources. It also increases trust between researchers and credibility of results, whether they meet prior expectations, are controversial, or otherwise. Exploratory analyses are clear from the outset, since pre-registration sets a distinction between confirmatory and exploratory research. There is also a clear distinction between un-registered and pre-registered exploratory analyses. Even if pre-registered studies are ultimately never completed or published, their existence helps to provide insight into the extent of the file-drawer problem. Pre-registration provides an avenue for public discovery of ongoing and unpublished research, which can be used to assess the extent of topic-specific selective publication biases and their impact on meta-analytical studies [259, 264]. The pre-registration process therefore helps to re-align research with the evolution of knowledge towards ultimate truth, rather than reputation-driven efforts to 'prove' your own predictions are correct, or presenting an appealing narrative regardless of its true likelihood.

From an individual perspective, even if you are fully aware of all common research pitfalls and manage to avoid them, including flexible study designs and even recall bias, you can still benefit from pre-registration. Pre-registration not only prevents falling victim to questionable research practices, but it also provides clear evidence to others that you have followed best practice. This reduces suspicion in your methods, where others may not trust your decision process [273]. Pre-registration avoids

unnecessary and challenging efforts to convince others that you did not partake in *p*-hacking or HARKing, since the pre-registration exists as a solid defence. Therefore, this process builds credibility in all your results, and confirmatory analyses are given their due credit [273]. Thorough planning in advance of data collection is also likely to increase the quality of your research and improve your workflow, since you will have detailed materials to refer to when it comes to the analysis stage. This can be particularly beneficial if you are involved in multiple ongoing studies, or where the time from study inception to execution is long. Pre-registration also offers an opportunity to receive credit for existing pre-analysis plans, such as may be written for grant funding applications. In this case, placing a time-stamped protocol in a repository requires minimal investment, and many platforms are designed with an optional embargo period if there is a reason to keep the pre-registration private for a time. This relationship is two-way, as you could also use pre-registrations as a basis for funding applications. Pre-registration can also be used to combat issues of 'scooping', protecting your work by evidencing that your ideas pre-dated another publication.

You do not have to rebuild the wheel when it comes to pre-registration. There are several popular and respected platforms for sharing pre-registrations, and many guides and templates to ensure adequate detail is included for different research purposes. For example, the Open Science Framework (OSF: `https://osf.io`) hosts a growing repository of pre-registered research, which can be assigned permanent digital object identifiers (DOIs), and there are many bespoke templates available (examples of curated templates at `https://osf.io/zab38/wiki/home/`). The OSF provides its own comprehensive pre-registration template (`https://osf.io/jea94`), whilst other examples include pre-registering secondary analysis (`https://osf.io/jqxfz`), or replication studies (`https://osf.io/4jd46`). Another particular version of pre-registration, is known as a Registered Report [279]. In this form, the pre-registration manuscript undergoes peer review at a journal, but the final report is also guaranteed to be published at that journal if the pre-registration is accepted, provided authors follow their pre-registered protocol. Registered Reports are a more

involved and lengthy process, and journals will have their own requirements, as with a standard publication, but the benefit of guaranteed publication insures against publication bias and the file drawer problem. At the peer review of a Registered Report, errors or improvements in methods can be highlighted before data collection and analysis have begun, preventing problems that would have arisen later when time and resource has already been wasted. Registered Reports also prevent issues that may arise in peer review of a completed study, such as post-hoc critique of methods, and 'moving the goalposts' where peer review suggests additional work which may be unrealistic or inappropriate. Since all the methods have been agreed upon in advance in a Registered Report, the goalposts cannot be moved for the final publication, and any such suggestions can easily be discussed as future work. Registered Reports therefore require more investment and time than un-reviewed pre-registration, but provide additional benefits which can be worthwhile for research quality, research dissemination, and ultimate publication experience.

In summary, pre-registration and Registered Reports are increasingly popular [268] and effective methods to reduce issues surrounding irreproducible research. However, as always, readers and reviewers must remain vigilant. Just because a study has been pre-registered, does not mean that authors necessarily followed their pre-registered plan accurately, included results of all planned metrics, or reported any methodological deviations clearly [280]. However, studies should not receive automatic credibility simply for having been pre-registered; comparisons to assess deviations and selective reporting between pre-registered and final protocols need to be made before research outcomes can earn increased trust. Since pre-registration does not have a standardised set of approved rules, the system can also be used inadequately. For example, if a pre-registration is vague about hypotheses and methods, this adds little value to reducing issues of p-hacking and other questionable research practices. Similarly, pre-registration does not improve a study which is fundamentally based on flawed methodology or a flawed research question. Registered Reports do have certain standards and guidelines that must be met, and is subject to peer review. Therefore, assuming the peer review process was performed correctly,

studies published via the Registered Report route should meet readers' expectations of truthful execution of the pre-registered protocol. Peer review, when performed correctly, should also ensure methods and research questions are not flawed.

1.5.4 Reproducibility and Neonatal Research

In neonatal research in particular, there is a lack of published reproducibility research. The difficulty conducting research in this vulnerable population is no doubt a contributing factor; there are a multitude of research questions which remain unanswered and most neonatal research relies on small sample sizes due to the difficulty in administration, ethical approval, and recruitment for this population.

Neonatal research would benefit from reproducibility studies, where conclusions and interpretations of findings could be skewed by the typically small sample sizes involved. Additionally, there is understandably a lack of immediate trust in research findings from the clinical perspective, and a slowness to translate research findings into clinical practice, which could be assisted by increased confidence in results attained from positive replication studies. Similarly, despite negative replication, there is resistance to changing existing medical practice which may not be beneficial to the infant, and additional larger-scale independent replication studies could encourage change away from irrelevant, or even potentially harmful, practices. One such example, is the controversy around providing sucrose as 'pain relief' discussed previously. It is possible that sucrose might hide the behavioural expression of pain or discomfort in infants, whilst not reducing pain felt nor preventing the adverse outcomes of pain, and perhaps even leading to additional or worsened adverse outcomes. There are psychological and emotional factors at play in clinical decision making, and it is difficult for medical practitioners to accept that a reduction in behavioural signs of distress in infants from sucrose may be harmful. This case demonstrates the importance of well-designed reproducibility studies to confirm or counter previous findings, as well as improving the perception of the value placed on these studies, which can shape the future of medical practice and quality of care.

Nevertheless, there are a number of barriers to conducting reproducibility studies in academia which I have discussed, and it is not an easy task to undertake.

1.6 Chapter Summary

Neonatal pain treatment has been neglected in recent history, and many unknowns remain. Early-life pain experience has been linked to short- and long-term adverse outcomes including cognitive and motor deficits, yet appropriate assessment and treatment of neonatal and infant pain is not well-understood. We lack a gold-standard objective pain assessment tool for this non-verbal population, and pain treatment regimes are not standard on either a local or global scale. Difficulty validating pain relief interventions in infants, in part due to subjective and insensitive pain measurement tools, has led to a dearth of licensed pain medication for infant pain and painful procedures. Dose-related drug efficacy, absorption, and metabolism profiles in infants also lack research. Age is an important factor when considering infant health and neurodevelopment, and factors into infants' pain experience. Premature infants are the most vulnerable, with increased early life pain exposure and subsequent adverse outcomes. Neuroimaging techniques provide non-invasive methods to study infant brain health and functional activity. They can be used to measure noxious-evoked cortical activity, and have the potential to provide a more objective pain-assessment metric in infants compared to existing behavioural tools. Neuroimaging metrics of noxious-evoked activity could therefore be used as a surrogate outcome measure of pain with improved sensitivity, to assess the efficacy of infant pain relief interventions in clinical trials.

Neonatal neuroimaging studies are difficult to conduct, resource-expensive, and limited by ethical considerations in neonatal research, which often leads to small sample sizes. These small sample sizes may lead to underpowered research and biases in the literature. One initiative to combat these issues has been the advent of large open neuroimaging datasets, such as the developing Human Connectome Project. These projects have pooled resources, including funding, equipment, and

expertise across multiple centres to provide high-quality data of a greater magnitude than would be possible with each centre working individually. The open science movement has also led to increased sharing of datasets from individual institutions and publications. Another initiative to combat biases published research has been the growing support for replication studies. Tools have been developed to help researchers to design and publish replication research, including pre-registration and registered report initiatives. Replication studies are lacking in infant research, likely due to the recruitment difficulties and resource limitations common in this field, but these studies are vitally important for establishing scientific truth and developing generalisable measurement tools for both research and clinical purposes.

1.7 Thesis Outline

The aims and motivation of this thesis are to address unknowns in neonatal pain and neurodevelopment research using neuroimaging data, and implementing methodological best-practices to do so. I achieve this through (1) adopting a pre-registered approach to reduce bias and improve the quality of my replication study, (2) designing a robust machine learning pipeline, avoiding common issues with bias; (3) using large independent datasets to explore hypotheses without wasting the valuable resource of difficult-to-acquire high quality neonatal neuroimaging data.

In Chapter 2, I present a pre-registered replication study which evaluates the efficacy of an EEG assessment of infant noxious response across different hospitals and infant populations. In Chapter 3, I describe how I developed and validated a machine learning pipeline to predict individual infant noxious-evoked responses from resting-state brain activity recorded with fMRI. I was integral in the development and application of this model to the data published in [231]. In Chapter 4, I apply this machine learning pipeline to an independent dHCP dataset to explore links between predicted noxious-evoked activity and MRI markers of structural and functional development and neurodevelopmental outcome at 2-year follow-up.

Finally, in Chapter 5, I provide a thesis summary, critique of my research, and outlook on future work.

2

A Pre-Registered Replication Study of an EEG Template of Noxious-Evoked Activity

The methods and analysis plans in this Chapter are presented in a Registered Report, of which I am lead author, accepted by Cortex and hosted on OSF: M. Aspbury, L. Baxter, M. M. Cobo, M. van der Vaart, et al. Establishing a standardised approach for the measurement of neonatal noxious-evoked brain activity in response to an acute somatic nociceptive heel lance stimulus. *Cortex*, 2022. DOI: 10.17605/OSF.IO/ZY9MS

2.1 Introduction

EEG can be used in neonates to measure brain activity changes that are evoked by noxious events, such as clinically-required immunisations, cannulation and heel lancing. EEG provides an alternative approach to infer pain experience in infants compared with more widely used behavioural and physiological pain assessments. Establishing the generalisability and construct validity of these measures will help corroborate the use of brain-derived outcomes to evaluate the efficacy of new or existing drugs to treat neonatal pain. Following a Registered Report format, I aimed to replicate three previously published findings where EEG-derived measures have been used to quantify changes in noxious-evoked brain activity in neonates [219, 228].

Firstly, I assessed whether an infant EEG template of noxious-evoked brain activity discriminates between noxious and non-noxious procedures. Secondly, I tested whether the scale of the average infant noxious response measured with the template is equivalent across different hospital datasets. Thirdly, I assessed whether the magnitude of noxious response measured with the template is positively correlated with age in premature infants. These hypotheses were tested using retrospective data that has been independently collected by another research group and will also be tested using prospectively collected data that is being collected at another research site.

2.1.1 Research Background and Motivation

The study of neonatal pain management would benefit from reproducibility research. Currently, pain management presents a major challenge in neonatal care and analgesic provision is infrequently and inconsistently provided during essential medical interventions [42, 94]. In part, this is due to difficulties in measuring pain and assessing the efficacy of analgesics in the non-verbal neonatal population [91]. While numerous clinical tools have been developed to calculate pain scores in neonates [282, 283], these measures are often reliant on behavioural and physiological observations that may not be sensitive or specific enough to (i) distinguish pain from distress [62], or (ii) to be primary outcome measures in clinical trials that aim to objectively assess the efficacy of analgesic interventions [91]. An alternative approach has been to consider changes in brain activity evoked by noxious clinical procedures [49, 106].

Over the past 10 years, brain-derived approaches have been used to quantify how the infant brain responds to acute noxious medical procedures, such as heel lancing and venepuncture [211, 221]. These approaches have been used as the primary outcome measure to test the efficacy of analgesics in clinical trials [49, 106]. In addition, further multicentre trials are underway to identify optimal pain management practices that may relieve pain in hospitalised infants [284]. Current and ongoing discussions with the FDA Division of Pediatrics and Maternal Health in collaboration with the Division of Anesthesiology, Addiction Medicine and Pain

Medicine [93, https://www.fda.gov/drugs/news-events-human-drugs/fda-m-cersi-analgesic-clinical-trial-designs-extrapolation-and-endpoints-patients-birth-less-two] are focussed on the use of brain-derived metrics in infants to improve analgesic drug development, with a specific focus on analgesic medications with known mechanisms of action such as NSAIDs, acetaminophen, local anaesthetics and opioids. It is therefore imperative that proxy pain measures that rely on brain-derived metrics are valid, reliable, and appropriate for use in large-scale clinical trials if they are to be a standard by which the efficacy of new or existing drugs can be evaluated in the neonatal population. It is critical to ensure the reproducibility of brain-based measures of noxious response given their potential to be used in clinical decision making.

An EEG-derived pattern of noxious-evoked activity has been observed in multiple settings, including in the UK at University College London Hospital [e.g., 211, 212, 227] and the John Radcliffe Hospital, Oxford [215, 217, 219], in Israel at Soroka Medical Centre in Beer Sheva [237], in Switzerland at the University Basel Hospital [223] and in the USA at the Nationwide Children's Hospital [236] – and is the most commonly used neuroimaging approach to study pain in infants [285]. Across multiple settings, a range of biological factors have been identified that modulate noxious-evoked brain activity, including premature birth [213, 217, 228], mode of delivery [223], stress [227], inflammation [68] and sex [226]. However, noxious-evoked EEG activity can be quantified in various ways, which hampers the comparison of effect sizes across different studies. To quantify and standardise changes in brain activity that arise following noxious clinical procedures, an EEG-based "template" measure that quantifies the magnitude of the noxious-evoked activity at a single central electrode site has been developed [219]. This measure constitutes a waveform which can be scaled to fit evoked activity in an EEG recording. We have previously demonstrated that the template magnitude is (i) larger following noxious stimuli compared to arousing non-noxious sensory events [219], is (ii) sensitive to modulation by pain management strategies [68, 215, 219] and (iii) is moderately correlated with pain-relevant behaviour [219]. However, we have not previously investigated the

consistency of these observations at different research sites. I pre-registered a report describing a multicentre study to explore the reproducibility and generalisability of the noxious-evoked EEG activity and to establish if reported biological observations that use this approach are generalisable across different centres [281]. The Registered Report design and preliminary results following this design are presented in this Chapter.

2.1.2 Research Aims

The EEG template of noxious-evoked brain activity was derived and validated in term and preterm infants aged $34.0 - 43.0$ weeks' PMA and is scaled to give an average magnitude of 1 in response to a heel lance (HL) performed in newborn term-aged infants [219]. This pre-registered reproducibility study has two aims. Firstly, we want to establish cross-centre generalisability by identifying whether the template of noxious-evoked brain activity can be reliably projected onto independent datasets. COSMIN produces a set of properties which should be upheld by medical instrument measurements. We used these guidelines as a basis to create testable hypotheses for two key criteria outlined by COSMIN; construct validity and reliability [69]. To this end, we will use retrospective and prospective datasets collected at two different sites to (i) identify whether the magnitude of the template of noxious-evoked brain activity is significantly larger in response to noxious compared to innocuous procedures, which is an assessment of the template measure's construct validity, and (ii) identify whether the template magnitude to a noxious procedure is consistent across sites, which is an assessment of inter-site reliability. This multi-level testing will form a robust construct validity and inter-site reliability assessment, and therefore generalisability, of the noxious-evoked EEG response template to discriminate between noxious and innocuous procedures in infants $34.0 - 43.0$ weeks' PMA. Secondly, there is value in exploring whether previous research findings relating to biological factors that are reported to modulate the magnitude of the EEG noxious event-related potentials are reproducible [217, 228]. Hence, we tested whether the magnitude of noxious-evoked brain activity as quantified by the template increases

with PMA during the period of prematurity up to 37 weeks in infants born at less than 36 weeks' GA, as reported by Schmidt Mellado and colleagues in 2022 [228]. The study populations, sample size planning, measurement, EEG processing and planned analytical approaches are described in detail in the Methods, and the precise hypotheses are stated below. It is important to note that while the proposed tests assess the reproducibility of a measure of noxious-evoked neonatal brain activity, this work does not aim to provide evidence that these activity patterns are a direct measure of neonatal pain. In the absence of language, we cannot know whether another person is in pain and therefore the brain-derived activity characterised here reflects only a pattern of brain activity that is known to be evoked by a noxious stimulus.

To plan this study and to conduct appropriate sample size planning analyses, I initially tested the generalisability of the template in a large Oxford dataset that was independent of the data included in the original template derivation [219]. Results from these initial analyses were used for sample size planning for the subsequent studies. Our hypotheses are tested in two datasets collected at independent sites: (1) the "UCL dataset" is retrospective data collected and published by an independent research group at University College London Hospital [286], and (2) the "Exeter dataset" is a prospective dataset that will be collected at The Royal Devon & Exeter Hospital after in-principle acceptance of our pre-registered report.

2.1.3 Pre-Registration

There are many benefits that led me to design and publish this study as a registered report. The general arguments for pre-registration and registered reports were previously outlined in the Introduction (Section 1.5.3). Results which this research aimed to replicate largely stem from studies conducted by the Paediatric Neuroimaging Research Group, of which I am also a member. Therefore, it would be reasonable for critics to doubt the authenticity of any positive replication outcomes published from within the research group, where they could have greater trust in replication studies which are conducted independently to the original research group(s). In order

to build trust in the outcomes of these replication tests, to demonstrate research integrity, as well as to avoid bias pitfalls which we may be more vulnerable to in testing the replicability of our own results, we decided to pre-register the research methods and protocol. We went one step further than pre-registration, to submit detailed plans as a Registered Report for peer review, to provide opportunity for our methods to be improved prior to the study based on feedback from external review, and ensure that all our methods were checked and approved by independent parties in advance. Peer review of the methods prior to the study commencing means that there is still an opportunity to improve the study methodology before it is too late. We decided that this process was the optimal pipeline to maximise the robustness and quality of our replication study. We also felt that this was a good opportunity to trial the novel research and publication practices around pre-registration, which should ultimately become more commonplace.

All necessary support and approvals were already in place prior to the study pre-registration. Data collection and analysis according to the protocol commenced upon in-principle acceptance of our study as a Registered Report at Cortex. When our submission was accepted by Cortex, the approved protocol was immediately registered on OSF (`https://osf.io/zy9ms`). We agreed to share all figure source data and analysis code with publication of the final results.

2.1.4 Chapter Outline

This Chapter tests whether a measure of noxious-evoked EEG activity, that was originally derived in a sample of term-aged infants at the Oxford John Radcliffe Hospital in 2017, can reliably distinguish noxious from non-noxious events in two independent datasets (i) retrospectively collected at University College London Hospital and (ii) prospectively collected at Royal Devon & Exeter Hospital. This research design was pre-registered and approved as a Registered Report to ensure research integrity and credibility. I aimed to reproduce three published results that use this EEG measure to quantify noxious-evoked changes in brain activity: (i) whether significantly larger noxious-evoked activity is recorded in response to

a clinical HL compared to a non-noxious control heel lance (CL) procedure; (ii) whether the magnitude of the activity evoked by a noxious HL is equivalent in independent cohorts of infants; and (iii) whether the magnitude of the noxious-evoked brain activity increases with age in premature infants up to 37 weeks PMA. Each hypothesis is outlined in detail in Section 2.2.1. Positive replication of these studies would build confidence in the use of brain-derived signals as valid and reliable pain-related outcomes which could be used to evaluate analgesic efficacy in neonates. Positive replication could also help towards establishing a standard method of analysing infant noxious-evoked responses using EEG.

Retrospective pilot data from Oxford were selected to qualitatively demonstrate internal reproducibility of each prior result being tested, and for sample size planning for each hypothesis test. The Oxford pilot data were independent samples to the original publications used to generate each hypothesis. Pilot data were used to estimate effect sizes, which were used in power calculations to determine appropriate sample sizes required from Exeter and UCL datasets to test each hypothesis. These Pilot results and analysis plans were pre-registered [281]. In this Chapter, these methods were applied to UCL and Exeter Datasets to test each pre-registered hypothesis. These results are presented for the first time in this thesis Chapter.

2.2 Methods

2.2.1 The Hypotheses

Hypothesis 1: Template magnitude discriminates noxious from non-noxious stimuli

The magnitude of noxious-evoked brain activity measured using the template should be greater following a noxious heel lance compared with a non-noxious control heel lance procedure in infants $34.0 - 43.0$ weeks PMA (described in Section 2.2.6). Thus, we hypothesise a significantly larger evoked template magnitude in response to the heel lance relative to the non-noxious control heel lance in infants $34.0 - 43.0$ weeks PMA, assessed using a paired t-test.

Hypothesis 2: Template magnitude for heel lance scale reproducibility

The group average magnitude of the noxious-evoked brain activity measured with the template should have a magnitude of 1 in response to a heel lance in newborn term-aged infants. Thus, we hypothesise that the average evoked template magnitude in response to the heel lance is equivalent to one, assessed using the magnitude, confidence interval and pre-defined equivalence bounds.

Hypothesis 3: Template magnitude for heel lance correlation with age

The magnitude of the noxious-evoked brain activity measured with an EEG template approach in response to a heel lance has been reported to increase in premature infants with PMA up to 37.0 weeks. Thus, we hypothesise a statistically significant positive correlation between PMA and the template magnitude in response to the heel lance, assessed using a Pearson correlation test.

2.2.2 Study Design

For hypothesis 1, the study uses a within-participant cross-sectional design, comparing outcomes of an observed clinical procedure to an experimental control procedure in the same participants. For hypotheses 2 and 3, the studies use a between-participant cross-sectional design, measuring outcomes of an observed clinical procedure across participants which are each studied only once. There are no additional interventions for any participants, all infants receive standard clinical care throughout in accordance with local practice unit guidelines, and therefore no randomisation procedure is required.

2.2.3 Recruitment and Research Protocols

Ethical approval was obtained from the NHS Research Ethics Committee (REC) for all infants in this study (references 11/LO/0350, 12/SC/0447 and 19/LO/1085). Further details of recruitment and study protocols are provided in sample consent

forms and patient information leaflets (PILs) for infant EEG studies, and standard operating procedures (SOPs) for withdrawal and safety reporting, in Appendix A.1.

2.2.4 Oxford Pilot Datasets Extracted from Oxford Database for Sample Size Planning

The first step in our replication study was to establish the internal validity of our EEG template results in all available Oxford data, and use these data to perform power calculations and sample size planning for each of the hypotheses I aimed to test in this replication study. Ethical approval for these data has been obtained from the NHS REC as part of ongoing studies (references 11/LO/0350, 12/SC/0447 and 19/LO/1085). Oxford infant metadata are stored in the Oxford Paediatric Neuroimaging Group Database (hereby termed the Oxford Database) developed by MedSciNet, which conforms to all relevant FDA, NIH, and HL7 standards, guidelines, and recommendations for storing data (`https://medscinet.com/about.aspx`, last accessed November 2021).

I extracted relevant retrospective data for this study from the database, according to inclusion and exclusion criteria listed below. There are two separate data extractions due to the different infant demographic groups required to test hypotheses 1 and 2 (Oxford Pilot Dataset A) compared to hypothesis 3 (Oxford Pilot Dataset B). Demographics of both datasets compared to the entire Oxford Database on MedSciNet are shown in Figure 2.1. I did not extract data for infants who were recruited in Oxford as part of an ongoing blinded clinical trial. Otherwise, I extracted all data from this database for infants recruited at the John Radcliffe Hospital, Oxford, who had EEG recordings taken as part of their study, according to criteria outlined in the following sections for each pilot dataset.

Oxford Pilot Dataset A: For Hypotheses 1 and 2

I extracted all data from the secure database corresponding to infants who were $31-43$ weeks GA and $34-43$ weeks PMA at time of study (inclusive ranges). The age ranges were chosen to encompass the ages of all infant datasets included in

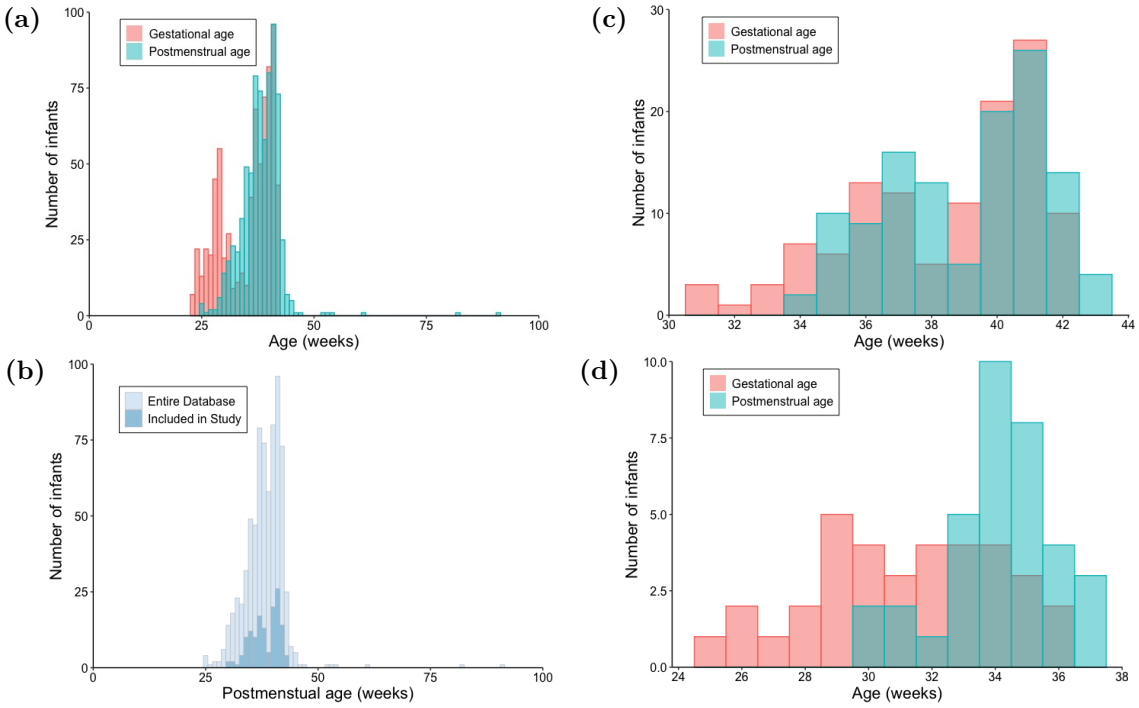


Figure 2.1: This figure illustrates properties of the entire Oxford Database and the Pilot Datasets selected to plan each hypothesis test. (a) The entirety of Oxford infants in our database ($n = 724$) on the date of extraction (30/10/2021), showing the distribution of their GA (range $= 23.0 - 42.4$; mean $= 35.2$ weeks) and PMA (range $= 24.7 - 90.7$; mean $= 38.1$ weeks). (b) The PMA distribution of the infant datasets selected from the database (both Pilot Dataset A and B) is shown as the shaded portion of the entire database's PMA distribution. The total number of selected (shaded) infants is $n = 141$, since 13 infants met the selection criteria for both Oxford Pilot Dataset A and Dataset B. (c) Oxford Pilot Dataset A's ($n = 119$) distribution of GA (range $= 31.0 - 42.4$; mean $= 38.4$ weeks) and PMA (range $= 34.1 - 43.0$; mean $= 39.0$ weeks) (d) Oxford Pilot Dataset B's ($n = 35$) distribution of GA (range $= 25.0 - 35.9$; mean $= 31.1$ weeks) and PMA (range $= 30.1 - 37.0$; mean $= 34.1$ weeks). Abbreviations: GA = gestational age; PMA = postmenstrual age.

Hartley et al. to (a) derive the template, and (b) prove the validity of the template in pre-term infants [219]. These parameters yielded 391 EEG recording test occasions covering 371 unique infants on the date of export (30th October 2021). There are more test occasions than infants since some infants were studied multiple times. The extracted EEG dataset was subset to $n = 211$ test occasions with both a HL and CL studied with EEG ($n = 180$ excluded). The following exclusion process is also illustrated as a flow diagram in Appendix A.5.

I excluded infants based on the following criteria: infants whose data were used

	Oxford Pilot Dataset A	Oxford Pilot Dataset B	Original Template Derivation Group
Number of infants	119	35	18
Gestational age (weeks)	39.4 (36.1, 40.7)	31.3 (29.1, 33.1)	39.3 (37.5, 41.4)
Postmenstrual age (weeks)	39.7 (36.8, 41.0)	34.1 (33.4, 35.0)	39.9 (37.9, 41.7)
Postnatal age (days)	3 (1, 5)	24 (6.5, 29.5)	3 (2, 5)
Birthweight (grams)	3510 (2770, 3923)	1626 (1045, 2060)	3437 (3236, 3638)
Sex			
Male	71 (60%)	20 (57%)	11 (61%)
Female	48 (40%)	15 (43%)	7 (39%)
Mode of delivery			
Normal vaginal	42 (35.3%)	15 (42.9%)	9 (50.0%)
Assisted vaginal / forceps	34 (28.6%)	3 (8.6%)	1 (5.6%)
Elective C-section	10 (8.4%)	3 (8.6%)	3 (16.7%)
Emergency C-section	33 (27.7%)	14 (40.0%)	5 (27.8%)
Apgar score			
1 minute	9 (7, 9.8)*	8 (4.5, 9)	9 (8.3, 10)
5 minute	10 (10, 10)	9 (8, 10)	10 (10, 10)
10 minute	10 (10, 10)	10 (9, 10)*	10 (10, 10)
Ventilation at time of study			
Self-ventilating in air	109 (91.6%)*	22 (62.9%)	18 (100%)
Low flow	5 (4.2%)	2 (5.7%)	0 (0%)
High flow	4 (3.4%)	10 (28.6%)*	0 (0%)
Estimated number of prior painful procedures	4 (2, 8)***	20 (8, 42)**	1 (0, 1)****

Table 2.1: Demographic variables for selected Oxford datasets. Values presented as median (LQR, UQR), or number (percentage) where appropriate. Asterisks indicate missing data in the database for * 1 infant, ** 3 infants, *** 8 infants; **** 11 infants. Abbreviations: LQR = lower quartile range, UQR = upper quartile range

in the original template creation ($n = 8$) [219], infants who were studied during a brushing or stroking intervention ($n = 48$) [215], infants who were included in a study investigating a kangaroo care intervention ($n = 3$), infants with neurological conditions such as (but not limited to) hypoxic-ischaemic encephalopathy (HIE) and intraventricular haemorrhage (IVH) of any grade ($n = 6$). I excluded test occasions where: either the HL or CL stimulus annotation was missing from the EEG record ($n = 5$), the EEG eventlist was incorrect or missing in storage ($n = 4$), the Cz electrode was not recorded ($n = 4$), there were repeated samples on the database ($n = 2$), and where infants were studied twice, I excluded one of their two test occasions ($n = 4$). For the latter exclusion, I always included the first test occasion and excluded the second, except one infant which had multiple CL on the first test occasion, and so I used the second test occasion which had exactly one CL and one HL measured. This left $n = 127$ infants whose EEG data I pre-processed.

Following EEG pre-processing rejection steps to remove artefactual epochs (see Section 2.2.7.2 for details), I excluded $n = 8$ infants who did not have both one HL and one CL epoch passing EEG rejection, to enable paired statistics. Another $n = 8$ infants had multiple HL and/or CL events after EEG rejection and, in these cases, I kept the first artefact-free recording for each of the HL and CL measurements for each infant.

This gave a final sample size of $n = 119$ infants with exactly one HL and one CL EEG epoch each (see Appendix A.5). This dataset comprised infants with a GA of $31.0 - 42.4$ weeks (mean = 38.4 weeks), PMA at study of $34.1 - 43.0$ weeks (mean = 39.1 weeks) and PNA at study of $0 - 44$ days (mean = 4.5 days). The dataset age distribution is illustrated in Figure 2.3c and full demographics are provided in Table 2.1.

Oxford Pilot Dataset B: For Hypothesis 3

For hypothesis 3, a separate sample of infants were selected from the Oxford Database due to the different age range under study, compared to hypotheses 1 and 2. I extracted all infant data from the database including EEG recordings for a HL

procedure ($n = 305$ test occasions), on the date of export of 1st November 2021. The following inclusion and exclusion criteria are also illustrated with a flow diagram in Appendix A.6.

I included infants who were born at less than 36.0 weeks GA and studied at an age up to and including 37.0 weeks PMA to match the upper age limit in Schmidt Mellado et al., 2022 [228] ($n = 125$ test occasions). Following this, I excluded infants included in the original Schmidt Mellado study ($n = 52$ exclusions), infants with neurological conditions including HIE and IVH of grade 2 or higher ($n = 10$), without EEG data ($n = 6$), with a missing eventlist ($n = 2$) or with a problem in the eventlist annotation for the HL such as the automatic trigger missing ($n = 3$). I also excluded duplicate data ($n = 2$), infants with an intervention aside from standard-care to reduce noxious response to the HL ($n = 2$), and test occasions beyond the first study for infants studied more than once ($n = 7$). This resulted in a sample of $n = 41$ infants with one test occasion, with EEG recordings for a clinical HL procedure. After EEG pre-processing and rejection (Section 2.2.7.1), $n = 6$ infant HL traces were rejected. In $n = 3$ infants there were multiple HL measurements after EEG rejection and, in these cases, I kept the first artefact-free recording for each of the HL measurements for each infant.

This left $n = 35$ infants each with one HL recording which were included for sample size planning for hypothesis 3 (see Appendix A.6). This dataset comprised infants with a GA of $25.0 - 35.9$ weeks (mean $= 31.1$ weeks), PMA at study of $30.1 - 37.0$ weeks (mean $= 34.1$ weeks), and PNA at study of $1 - 59$ days (mean $= 21.2$ days). The dataset age distribution is illustrated in Figure 2.3d and full demographics are provided in Table 2.1.

2.2.5 Pre-Registered Inclusion and Exclusion Criteria for External Datasets for Hypothesis Testing

Data inclusion and exclusion follow the same criteria for each external dataset, although inclusion criteria depend on the hypothesis in question. Both hypotheses 1 and 2 share the same inclusion criteria, however hypothesis 3 requires a different

demographic sample. Inclusion and exclusion criteria closely match key variables in the prior studies that our reproducibility study is based on, such as the age range of the infants in the original studies. In this section, shared exclusion criteria for all datasets are listed first, followed by specific details for each external dataset in their relevant sections.

2.2.5.1 Exclusion Criteria for All Datasets

Infants were excluded if they had any known or suspected neurological condition such as (but not limited to) HIE or IVH of grade 2 or higher, if they received analgesics in the 24 hours preceding the study, if there was known maternal substance abuse, if they received an intervention other than standard clinical care which was being assessed for efficacy in reducing their noxious-evoked response, or if they were a repeat study of an infant which was already included. EEG traces for remaining infants were then pre-processed and analysed, following EEG rejection procedures to remove artefactual epochs as outlined in Section 2.2.7.2. Furthermore, infants were excluded after pre-processing if they did not have required EEG recordings after rejection (a HL and CL for hypotheses 1 and 2; only a HL for hypothesis 3), or if there were any other technical problems with the data for that infant (e.g., missing required data, or unable to access recordings due to data corruption).

2.2.5.2 UCL Dataset

The UCL dataset is available via ReShare. Authorised access is given to users due to the sensitivity of the data ([287], UK Data Service, `https://doi.org/10.5255/UKDA-SN-853204`). I did not request data access and analyse any UCL data until after in-principle acceptance of the Registered Report by Cortex. From UCL, I included all available data from this retrospective dataset to test each hypothesis. The UCL dataset contains $n = 112$ infants from which we selected infants to test each hypothesis according to the relevant inclusion criteria.

UCL Dataset A: Data Inclusion Criteria for Hypotheses 1 and 2

For both these hypotheses I included healthy infants who were $31.0 - 43.0$ weeks GA and $34.0 - 43.0$ weeks PMA at study, without restriction on postnatal age. These

	UCL Dataset A	**UCL Dataset B**	**Exeter Dataset A**
Number of infants	74	51	23
Gestational age (weeks)	36.6 (35.3, 39.3)	33.1 (29.9, 34.9)	38.9 (36.6, 41.0)
Postmenstrual age (weeks)	37.4 (35.9, 39.9)	34.6 (32.8, 35.8)	39.0 (37.1, 41.2)
Postnatal age (days)	5 (3, 6)	6 (5, 14.5)	2 (1, 5)
Birthweight (grams)	2625 (2182, 3445)	1760 (1320, 2243)	3339 (2650, 3740)
Sex			
Male	35 (47%)	24 (47%)	10 (43%)
Female	39 (53%)	27 (53%)	13 (56%)
Mode of delivery			
Normal vaginal	23 (31.1%)	12 (23.5%)	9 (39%)
Assisted vaginal / forceps	13 (17.6%)	12 (23.5%)	4 (17%)
Elective C-section	14 (18.9%)	8 (15.7%)	3 (13%)
Emergency C-section	23 (31.1%)	19 (37.3%)	7 (30%)
Apgar score			
1 minute	9 (8, 9)	8 (6, 9)	8 (7, 9)
5 minute	9 (9, 10)	9 (9, 10)	9 (8, 9)
10 minute	Not avail.	Not avail.	9 (9, 9)
Ventilation at time of study			
Self-ventilating in air	73 (100%)	38 (75%)	22 (96%)
Low flow	0	1 (2%)	1 (4%)
High flow	0	12 (24%)	0 (0%)
Estimated number of prior painful procedures	8 (4, 13)	16 (12, 26)	3 (1, 5)

Table 2.2: Demographic variables for selected UCL and Exeter datasets. Values presented as median (LQR, UQR), or number (percentage) where appropriate. Not avail. indicates data which were not provided in UCL database. Abbreviations: LQR = lower quartile range, UQR = upper quartile range

demographics are matched to datasets in the original Hartley validation study [219], as well as the Oxford pilot data used to power these hypotheses (Section 2.2.4: Oxford dataset A). Infants with both a HL and a CL procedure recorded with EEG were included.

The age inclusion criteria resulted in a sample of $n = 79$ infants with GA between $31.9 - 41.9$ weeks (mean = 37.1 weeks) and PMA at study between $34.0 - 42.9$ weeks (mean = 37.9 weeks) in UCL Dataset A. The EEG data were then checked for these 79 infants to ensure there was both one HL and one CL annotation available for these infants. After EEG pre-processing, there were $n = 74$ infants with a HL passing EEG rejection criteria, and $n = 61$ infants with both one HL and one CL epoch passing EEG rejection criteria. Demographics are summarised in Table 2.2.

UCL Dataset B: Data Inclusion Criteria for Hypothesis 3

We included infants who were born at less than 36.0 weeks GA and studied up to and including 37.0 weeks PMA, without restriction on PNA. This age range matches those studied in Schmidt Mellado et al. [228], who found that noxious-evoked response correlates with PMA in premature infants in this age range. Infants only required an EEG recording during a HL procedure which passes EEG rejection criteria to be included in this dataset; data quality of CL epochs were not assessed for inclusion.

The age inclusion criteria resulted in a sample of $n = 57$ infants with GA of $23.2 - 35.9$ weeks (mean = 31.7 weeks) and PMA at study of $29.4 - 37.0$ weeks (mean = 33.9 weeks) in UCL Dataset B. The EEG data were then checked for these 57 infants to ensure there was one HL annotation available for these infants. After EEG pre-processing, there were $n = 51$ infants with one HL annotation passing EEG rejection criteria. Demographics are summarised in Table 2.2.

2.2.5.3 Exeter Dataset

Ethical approval for the Exeter study is in place from the NHS REC as part of an ongoing study (reference 12/SC/0447). Written parental consent is obtained from

parents prior to each study and studies will conform with the declaration of Helsinki. Pilot data collection began in June 2021 for training purposes, to ensure adequate familiarity with EEG recording equipment and protocol in advance of the planned study. However, data analysis for this study did not begin until after in-principle acceptance of the Registered Report by Cortex.

Individual infants, prospectively recruited in Exeter, may be studied on more than one occasion, however only data from the first usable test occasion was used in this study. Infants received standard clinical care throughout the procedures in accordance with local practice guidelines. We planned to recruit at least 10% more infants than the calculated sample sizes to account for data artefacts and participant withdrawal. Full participant inclusion and exclusion criteria are presented here.

Exeter Dataset A: Data Inclusion Criteria for Hypothesis 1

In Exeter, we planned to prospectively recruit and study the minimum required sample of $n = 40$ infants to test hypothesis 1 (power calculation in Section 2.3.2.1). We recruited and included infants who were $31.0 - 43.0$ weeks GA and $34.0 - 43.0$ weeks PMA at study, without restriction on PNA. This reflects the demographics supported in the original Hartley study [219], as well as the sample of pilot data used to power this hypothesis (Section 2.2.4: Oxford dataset A). Infants needed to have both a HL and a CL procedure recorded with EEG, and passing rejection criteria outlined in Section 2.2.7.2, to be included.

Due to time constraints, at time of study, recruitment was in progress with $n = 27$ eligible infants recruited so far. These infants are part of an ongoing blinded RCT for a pain-reliving HL intervention for which the data are still blinded, and therefore the Exeter analysis is only at a pilot stage. EEG data were checked for these 27 infants to ensure there was both one HL and one CL annotation available for these infants. After EEG pre-processing, there were $n = 23$ infants with both one HL and one CL epoch passing EEG rejection criteria. Demographics are summarised in Table 2.2.

Exeter Dataset B: Data Inclusion for Hypothesis 3

We did not plan to specifically recruit infants in Exeter to test this hypothesis. Instead, we planned to include the subset of infants from our total recruited sample of 40 infants in Exeter, whose demographics are described above, who met the inclusion criteria for this hypothesis test. For this hypothesis, inclusion criteria are infants born prematurely; with a GA less than 36.0 weeks and with PMA at study of up to and including 37.0 weeks, to reflect the hypothesis that noxious-evoked response correlates with age at study in premature infants in this range, from Schmidt Mellado's study [228].

These infants only need an EEG recording during a HL procedure which passed EEG rejection criteria (Section 2.2.7.2); data quality of CL was not assessed for inclusion. The majority of data used to test hypothesis 3 should be sourced from the retrospective UCL dataset, and we only planned to include additional data from the prospective Exeter dataset to maximise the available sample size. Data from UCL alone were expected to meet the minimum calculated sample size requirement of $n = 38$ infants (Section 2.3.2.2).

Of the $n = 27$ infants eligible for hypotheses 1 recruited in Exeter at time of study, only $n = 4$ matched the demographic inclusion criteria for hypothesis 3. These infants are also part of an ongoing blinded RCT for a pain-reliving HL intervention for which the data are still blinded. Therefore, due to the very small numbers and noise due to unknown intervention status for the HLs, no Exeter data was included to test hypothesis 3 at this time.

2.2.6 Noxious and Non-noxious Stimuli

Participants included in this study have received a HL and CL stimulus to the foot (hypotheses 1 and 2) or one HL to the foot (hypothesis 3). No extra HLs for blood tests were taken for these infants solely for research purposes.

2.2.6.1 Heel Lance

The noxious stimulus used throughout is the heel lance (HL), which is a clinically required procedure that involves puncturing the skin on the heel in order to obtain blood for analysis, and which is one of the most common invasive procedures performed in the NICU [45]. In all cases, the HLs are applied due to clinical need, and not solely for research purposes. If the clinician performed multiple blood tests to collect enough blood, data is included from only the first HL where data is useable.

2.2.6.2 Control Heel Lance

As a control to the noxious stimulus, the control heel lance (CL) is used, which is designed to replicate the non-noxious aspects of the HL procedure. The CL is performed in the same conditions as the HL, however the control stimulus involves rotating the lance by 90 degrees, so that the blade is released into the air and not the infant's foot. This removes the noxious skin-break whilst maintaining the auditory and tactile stimulation of the procedure.

2.2.7 EEG Methods

2.2.7.1 EEG Recording

In all datasets, the HL and CL stimuli were linked electronically to the EEG recording equipment to precisely time each stimulus and mark the concurrent EEG at the time of each stimulus. This improves accuracy and precision of the marked stimuli, reduces human error, and avoids disrupting clinical practice. This time-locking method has been used and described in previous studies [49, 211, 288] and was also used and described by researchers for the UCL data collection [286]. EEG recording equipment remained connected throughout the clinical procedure to avoid disruption to the infant's state.

Prospective Exeter Data Collection: Electrophysiological activity was acquired with a Compumedics Grael V2 EEG System with a bandwidth from DC to 400 Hz and a sampling rate of 2048 Hz. CURRYscan8 neuroimaging suite (Compumedics

Neuroscan) was used to record the activity. All equipment conformed to the electrical safety standard for medical devices, IEC 60601-1. Eight EEG recording electrodes were positioned on the scalp at Cz, CPz, C3, C4, FCz, T3, T4 and Oz according to the modified international 10-20 System. Reference and ground electrodes were placed at Fz and the forehead respectively. EEG conductive paste were used to optimise contact with the scalp. Electrode/skin contact impedances were reduced to approximately 5 kΩ by rubbing the skin with EEG preparation gel prior to electrode placement. An ECG electrode was placed on the left clavicle to record heart rate.

Oxford Pilot Data: The EEG data collection procedures were identical to the prospective Exeter data collection, apart from a few technical differences. Oxford data were collected with a SynAmps RT 64-channel headbox and amplifiers (Compumedics Neuroscan), with a bandwidth from DC to 400 Hz and a sampling rate of 2000 Hz and recorded with CURRYscan7 suite (Compumedics Neuroscan). Infants on the Oxford database either had EEG recorded with Fz as the reference electrode, or where FPz was used as the reference electrode these data were re-referenced to Fz during pre-processing.

UCL Data: EEG activity was acquired using the Neuroscan SynAmps2 EEG/EP recording system with a bandwidth from DC to $\geq$ 500 Hz [286]. Signals were digitised with a sampling rate of 2 kHz and a resolution of 24 bit. EEG was recorded from up to 20 electrodes, in addition to ground and reference electrodes. Recording electrodes were positioned according to a modified international 10/10 electrode placement system, with high density central-parietal and posterior temporal coverage, overlying primary visual (O1, O2), primary auditory (T7, T8), association (F7, F3, F4, FCz, F8, P7, P8, TP9, TP10, POz), and somatosensory cortices (C3, Cz, C4, CP3, CPz, CP4). The reference electrode was placed at Fz and the ground electrode at either FC1 or FC2 depending on the infant's position. Electrode/skin contact impedances were minimised by gently rubbing the skin with a prepping gel before applying electrodes with a conductive paste. ECG was recorded and respiratory movements were monitored. EEG was assessed by an experienced clinical

scientist, and all infants had normal EEG activity for their PMA. EEG recordings were converted into EEGLAB data structures and segmented into 4 s epochs (2 s pre- and post-stimulus). Each file contains data from the EEG electrodes, ECG, and respiration transducer, with a trigger indicating the stimulus at 0 s. All methods reported in Jones et al. [286].

2.2.7.2 EEG Analysis

EEGlab [289], Brainstorm [290] and custom-made scripts were implemented in MATLAB (Mathworks) to filter, epoch, and reject EEG traces, as well as fit the noxious-evoked EEG response template. The EEG processing procedures were identical for all datasets, with the exception of downsampling the data from 2048 Hz to 2000 Hz for the Exeter dataset. The procedures and parameters in the semi-automated EEG processing pipeline are as follows:

Automated filtering and epoching: raw EEG data from the Cz electrode (continuous data for the Oxford and Exeter datasets, and 4-second epochs for the UCL dataset) were loaded into MATLAB. If not already recorded at 2000 Hz, data was downsampled to 2000 Hz using EEGlab [289] to allow compatibility with the template. Next, data was filtered between 0.50 Hz and 30 Hz, and epoched with 500 ms before and 1000 ms after each stimulus annotation of interest. Epochs were baseline-corrected to the pre-stimulus mean.

Manual rejection procedure: EEG traces were assessed in a time-window of -500 ms to +1000 ms around the stimulus marker. Traces with sustained activity greater than +150 µV or smaller than -150 µV are rejected, as well as traces where the baseline amplitude spans more than 100 µV, and traces with visually apparent artefact falling within these pre-defined amplitude bounds (e.g., gross movement artefact, muscle artefact, line noise or other technical artefact).

Automated template projection: The EEG template was projected onto all Cz EEG traces passing quality criteria in a time-window $400 - 700$ ms after the stimulus. Each individual trial was first Woody filtered to the template with a maximum

jitter of ± 100 ms in the template fit time-window to allow for individual differences in stimulus response latency. The result of this template projection is a scaling factor which represents the magnitude of noxious-evoked response for each trial, also referred to as the template magnitude, or template coefficient.

2.2.8 Blinding Procedure

I was the sole analyst to pre-process the EEG data and analyse results, and I was not involved in any of the prior published research whose results were tested for reproducibility in this study. I did not access, view, or analyse the UCL data prior to designing and submitting the study for pre-registration approval, sharing the submitted pre-registered protocol on OSF (`https://osf.io/zy9ms`) and receiving in-principle acceptance at Cortex. This included sample size planning, choice of hypotheses and analyses, and all methods published in the Registered Report [281]. The UCL metadata are available to view as supplementary material, under isa-tab metadata, with the publication (`https://doi.org/10.1038/sdata.2018.248`), which I last accessed on October 12th 2021 to estimate sample sizes for the planned analyses outlined here and in the Registered Report. Equally, participants were being recruited in Exeter for an ongoing blinded RCT during the Registered Report peer review process, and I did not view or analyse any of the Exeter data included in this analysis prior to designing and submitting the study for pre-registration approval, or sharing the submitted pre-registered protocol on OSF (`https://osf.io/zy9ms`).

Some of my colleagues in the Oxford Paediatric Neuroimaging Group had previously observed and analysed aspects of the UCL data for separate research concerning premature infants' discriminable behavioural, physiological, and brain responses to noxious and non-noxious stimuli [218]. This publication included 1) visualisation of raw EEG for HL and CL across a subset of the dataset (infants aged 28-40 weeks PMA) 2) extraction of the main waveforms in the EEG response data; 3) projection of the template to the EEG data for inclusion in a machine learning model. This did not compromise the blinding procedure, as I was not involved in this prior research and did not have access to UCL data before pre-registration.

2.2.9 Statistical Analysis Plans

The statistical analysis plans to test each hypothesis under examination for reproducibility in this study are presented individually below. Analysis methods are also described for quality control checks for the included data, termed outcome neutral criteria. These methods were detailed in advance of conducting any data analysis to test the hypotheses.

2.2.9.1 Outcome Neutral Criteria

Outcome neutral criteria are quality control checks which are designed to confirm that data meet hypothesis-independent expectations, to avoid false conclusions and misinterpretation of the data due to unrecognised faults, arising from technical error or otherwise. We decided that appropriate outcome neutral criteria would comprise confirming the presence of stimulus-evoked activity in the infants' EEG traces for both stimuli. These criteria are based on the premise that if EEG data quality is sufficient, evoked activity should be present in response to both the HL and the CL.

We would expect both the non-noxious (auditory and tactile) control stimulus and the noxious stimulus to evoke electrophysiological activity. Therefore, we designed our outcome neutral criteria to check for significant differences between post-stimulus (stimulus-evoked) and pre-stimulus activity on average, for both non-noxious and noxious stimuli. For outcome neutral criteria involving the CL stimulus data, we only performed these checks for the included CL data which was used to test hypothesis 1, as the other hypotheses only rely on HL data. The Oxford pilot dataset was tested first against these outcome neutral criteria, since it was also used to pilot the outcome neutral criteria tests. All included HL data that was used to test the hypotheses, from Exeter and UCL datasets, was combined to check outcome neutral criteria across datasets.

I performed a within-trials non-parametric cluster-based permutation analysis, in all included EEG traces recorded at the Cz electrode, to test for significant differences between the period 1-second pre-stimulus and 1-second post-stimulus [291]. I used

pre- and post-stimulus EEG epochs of equal length (1 s each) in order to use paired statistics. These EEG traces had been processed up to and including the filtering stage described in Section 2.2.7.2, without Woody filtering. For each infant, the difference between the pre-stimulus and the post-stimulus data was computed, and t-statistics were calculated at each timepoint. Clusters of timepoints with t-statistics greater than the 97.5th percentile of the t-distribution were identified. Cluster mass (defined as the sum of the t-statistics in the cluster) was calculated and was assessed for significance using 10,000 random permutations of the data (significance threshold $p < 0.05$). To pass this quality control stage, there should be significant clusters of difference between pre- and post-stimulus epochs for both CL and HL stimuli separately, which would demonstrate that there are stimulus-evoked signals in the EEG in each stimulus condition.

2.2.9.2 Analysis to test Hypothesis 1: Assessing the template's generalisability to discriminate noxious and innocuous activity

Hypothesis 1: The magnitude of noxious-evoked brain activity measured using the template should be greater following a noxious HL compared with a non-noxious CL procedure in infants $34.0 - 43.0$ weeks PMA. Thus, we hypothesise a significantly larger evoked template magnitude in response to the HL relative to the non-noxious CL in infants $34.0 - 43.0$ weeks PMA, assessed using a paired t-test.

To test Hypothesis 1, I planned two studies:

1. To retrospectively analyse all data available in the UCL Dataset for the inclusion criteria for this hypothesis (outlined in Section 2.2.5.2: UCL Dataset A) which I estimated to be $n \approx 88$ from the published metadata [286].

2. To prospectively recruit $n = 40$ infants* from The Royal Devon & Exeter Hospital, with recruitment and inclusion criteria outlined in Section 2.2.5.3: Exeter Dataset A. *Following sample size planning in Section 2.3.2.1.

If there is a significant increase for HL greater than CL in each dataset ($p < 0.05$) then the hypothesis would be confirmed, and the result that the EEG template

discriminates between a noxious HL and a non-noxious CL would be reproduced across multiple data centres. Otherwise, I would report that we are unable to reject the null hypothesis and a failure to reproduce this result. To test this hypothesis, data were first selected according to data inclusion (Section 2.2.5) and EEG processing procedures (Section 2.2.7.2). Then, for each dataset, I performed a paired one-tailed t-test for a HL response greater than CL response, as measured by the noxious-evoked EEG template.

2.2.9.3 Analysis to test Hypothesis 2: Assessing the consistency of the template magnitude

Hypothesis 2: The group average magnitude of the noxious-evoked brain activity in response to a HL measured with the template should have a magnitude of 1 in term-aged neonates. Thus, we hypothesised that the average evoked template magnitude in response to the HL is equivalent to one, assessed using the template magnitude confidence interval and pre-defined equivalence bounds.

To test hypothesis 2, I calculated the mean and 90% confidence interval (CI) for HL response in the UCL dataset (Section 2.2.5.2: UCL dataset A). The CI was calculated by bootstrap resampling using the boot library in R [292, 293]. If the mean and 90% CI are within the bounds $0.8 - 1.2$, which is $\pm$ 20% around 1, then the mean HL response can be considered equivalent to 1 in an independent dataset according to FDA guidelines (see Section 2.3.2.2), and we would confirm this hypothesis. If the 90% CI spread outside either of these bounds (less than 0.8 or greater than 1.2) then we would be unable to conclude whether the mean HL response is equivalent across sites, and this hypothesis will be unconfirmed.

Accuracy In Parameter Estimation (AIPE) sample size planning described in Section 2.3.2.2 suggests a sample of $n = 73$ infants is required to test the consistency of template magnitude across sites. This was only achieved in the retrospective UCL dataset ($n \approx 88$, data inclusion in Section 2.2.5.2: UCL dataset A). It was not feasible to recruit this sample size at the Exeter site during the study period, but I will also report results of the mean and 90% CI for HL response in the smaller

($n \leq 40$) Exeter dataset for transparency and completeness. These results were not expected to be conclusive due to the smaller sample size.

2.2.9.4 Analysis to test Hypothesis 3: Assessing whether the template measure of noxious response increases with PMA in premature infants

Hypothesis 3: The magnitude of the noxious-evoked brain activity measured with an EEG template approach in response to a HL has been reported to increase in premature infants with PMA up to 37.0 weeks. Thus, we hypothesise a statistically significant positive correlation between PMA and the template magnitude in response to the HL, assessed using a Pearson correlation test.

If hypothesis 1 is unconfirmed then we will lack the basis to test hypothesis 3, as we first assume that we are measuring the magnitude of noxious-evoked brain activity to a HL. If hypothesis 1 is confirmed, then we will proceed with the analysis plan to test this third hypothesis.

Hypothesis 3 is designed to assess the reproducibility of a biological finding of age-related change in noxious-evoked EEG activity using the template measure, as opposed to the generalisability of template characteristics across multiple sites. Therefore, to test hypothesis 3 we combined all data in the retrospective UCL Dataset ($n \approx 56$) and the prospective Exeter Dataset ($n \leq 40$), which matched inclusion criteria as outlined in Section 2.2.5.2: UCL Dataset B and Section 2.2.5.3: Exeter Dataset B. The majority of this data is sourced from the retrospective UCL dataset, as it is a larger dataset than the prospectively collected Exeter sample. In contrast to hypotheses 1 and 2, we did not recruit additional infants in Exeter to specifically test hypothesis 3. We did however plan to include any available data from the prospective Exeter dataset that matched the criteria to test this third hypothesis. We expected that data from UCL alone should at least meet the calculated sample size requirement of $n = 38$ infants to test this hypothesis (see Section 2.3.2.3). We therefore expected to exceed the required sample size.

To test hypothesis 3, we planned to perform a one-sided Pearson correlation analysis, testing for a positive correlation between PMA and HL magnitude. We planned to combine retrospective UCL and prospective Exeter data for this test, and include a confound variable for the data site (UCL or Exeter) to account for any differences in response magnitude between different sites. If the Pearson correlation is positive between PMA at study and HL response with p-value < 0.05, accounting for the site confound variable, then this hypothesis will be confirmed. Otherwise, we will report that we are unable to reject the null hypothesis and a failure to reproduce this result.

2.3 Pilot Data Analysis

2.3.1 Pilot Validation Procedures

2.3.1.1 Reproducing Original Template Validation Result

Since the publication of the template and its validation, our EEG processing pipeline has been updated in line with recommended processes. Since the revised pipeline has the potential to alter the morphology of the EEG data, we determined it necessary to confirm that the revised pipeline did not change the results and interpretation of the template validation in Hartley et al., using the original infant dataset including noxious-evoked, non-noxious-evoked, and background EEG data ($n = 18$, demographic details in Table 2.1). We tested the updated EEG processing pipeline using identical raw data to those used in the original study. The template was scaled so that a value of 1 represents the mean response evoked by a clinically-required HL in a term-aged neonate, but these original data include a combination of experimental noxious stimuli and HL stimuli, therefore the mean response to all noxious stimuli is less than 1 (Figure 2.2). Non-noxious stimuli comprise CLs and gentle touch stimuli, and background activity is activity prior to any stimulation which was manually event-marked at time of recording [219].

Using the EEG template and updated pipeline, the noxious response was still significantly greater than non-noxious response and background activity, whilst

non-noxious response was not significantly different from background (Figure 2.2). The results confirm that the updated EEG data analysis pipeline did not change the original conclusions reported. Therefore, we start the replication study on solid ground.

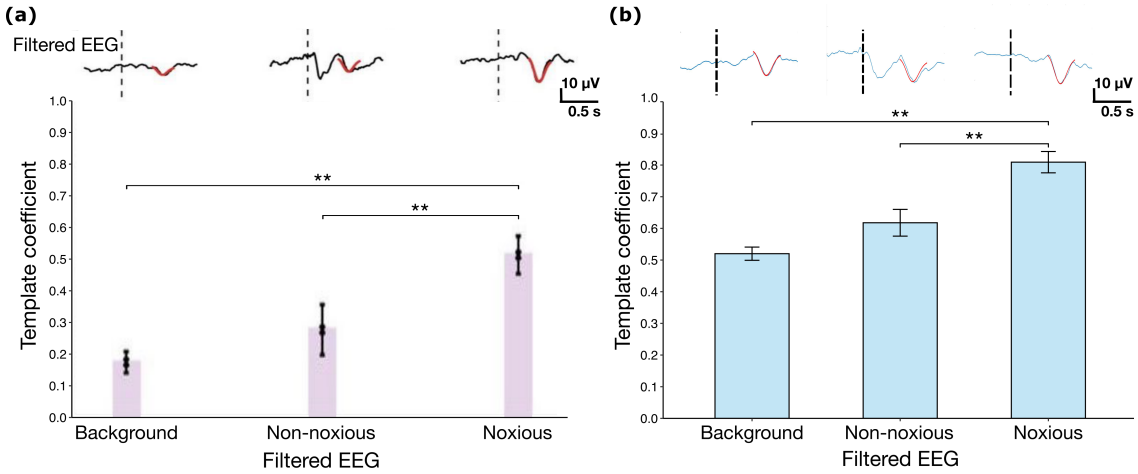


Figure 2.2: (a) Original template derivation in [219]. Figure from C. Hartley, E. P. Duff, G. Green, G. S. Mellado, et al. Nociceptive brain activity as a measure of analgesic efficacy in infants. *Science Translational Medicine*, 9(388), 2017. DOI: `10.1126/scitranslmed.aah6122`, reprinted with permission from AAAS, and adapted with permission from the lead author. (b) Replication of template results with the dataset ($n = 18$) used in the original template derivation [219]. Top: Mean Woody-filtered EEG traces for all participant trials for background, non-noxious, and noxious stimuli, with the noxious-evoked response template fit overlayed in red. The vertical dashed black line for each trace shows the time marker for the stimulus (time = 0 seconds). Bottom: Mean noxious-evoked template coefficient over all trials for background, and non-noxious and noxious stimuli with bars showing standard error of the mean, and asterisks indicating different levels of significance between measures. Noxious response is significantly greater than non-noxious ($t = 2.9$, $p = 0.006$) and background ($t = 2.8$, $p = 0.007$) whilst non-noxious response is not significantly different from background ($t = -0.9$, $p = 0.2$) using a one-sided partially overlapping samples t-test, as described in Derrick et al. to account for data containing both paired and independent samples [294–296]. These comparisons match those shown by Hartley et al., 2017, with the same data, despite differences in the EEG pre-processing. The original template derivation included responses to low intensity experimental noxious stimuli and therefore, as expected, the re-analysed data had a mean response magnitude that was less than 1 (mean noxious response = 0.81; 95% CI: $0.71 - 0.93$).

2.3.1.2 Outcome Neutral Criteria in Pilot Data

All pilot data were combined to test for outcome neutral criteria. There are $n = 141$ unique HL traces from both Oxford datasets A and B, since some participants are

present in both datasets. There are $n = 119$ CL traces, which are all from Oxford dataset A, since CL are not required for the third hypothesis test, which uses Oxford dataset B. Results for the outcome neutral criteria for each stimulus are shown in Figure 2.3.

There are significant clusters in the post-stimulus EEG compared to the pre-stimulus EEG for each stimulus separately, using non-parametric cluster-based analysis (Figure 2.3, bottom). For the HL stimuli ($n = 141$), there is one significant cluster identified after correction, representing a noxious-evoked potential, from $435 - 1000$ ms ($p < 0.0001$). For the CL stimuli ($n = 119$), one significant cluster, representing a non-noxious early potential, is identified at $159 - 375$ ms ($p = 0.005$). These cluster timepoints match those reported in the literature for both the noxious potential and non-noxious early potentials [212, 219, 297]. These data pass the outcome neutral criteria tests since there is significant post-stimulus evoked activity, not present in the pre-stimulus baseline activity, for both HL and CL EEG traces.

2.3.2 Power Analyses and Sample Size Planning with Pilot Data

In the above, I have demonstrated consistency of the original publication findings [219, 228] in our pilot data using a large locally-acquired independent Oxford dataset (Section 2.3.1.1), and determined that Oxford Pilot datasets meet quality control criteria relating to outcome neutral expectations (Section 2.3.1.2). Following the quality control checks, I used these independent data to estimate effect sizes for power calculations to plan sample sizes to test each hypothesis, which I present in the following sections. These power calculations match the relevant statistical analyses used to test each hypothesis, as described in Section 2.2.9.

2.3.2.1 Sample Size Planning for Hypothesis 1

In the original study [219], the noxious-evoked response template was shown to discriminate between noxious and non-noxious stimuli in term and premature infants across multiple experiments. This included a significantly larger magnitude of evoked

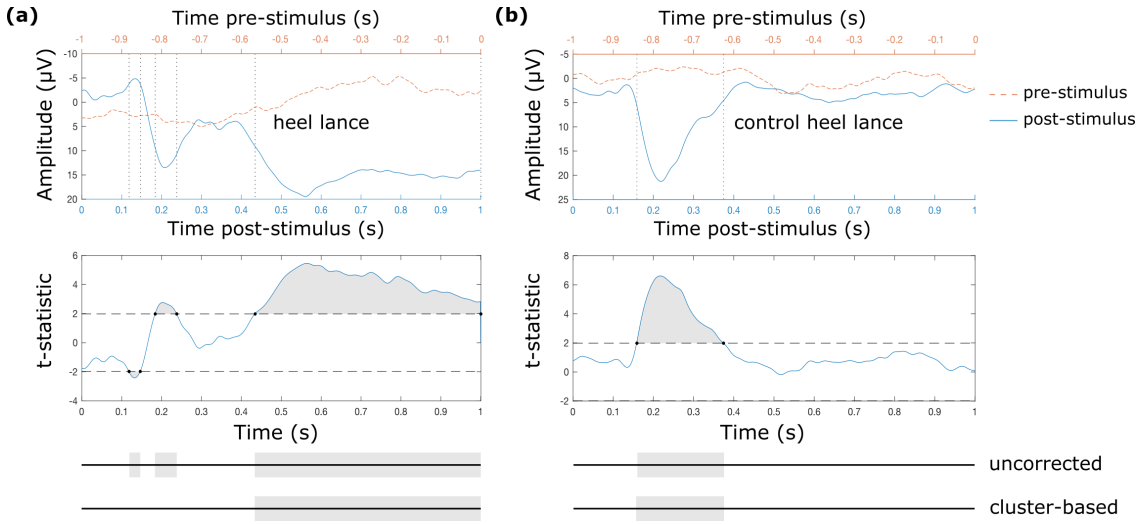


Figure 2.3: Outcome neutral criteria for all unique Oxford Pilot Data (a) heel lance stimulus EEG ($n = 141$) and non-parametric cluster analysis results (b) control heel lance stimulus EEG ($n = 119$) and non-parametric cluster analysis results. Top: stimulus EEG trial averages overlayed for the 1-second post-stimulus (blue) and 1-second pre-stimulus (orange), where stimuli are at 0 seconds. Bottom: non-parametric cluster analysis was performed between all participant post- and pre-stimulus EEG traces, and the shaded regions are identified regions of difference before the cluster-based correction. Clusters which are significantly different after cluster-based correction are shaded at the bottom of the figure ($p < 0.05$). Both heel lance and control heel lance stimuli have a significant cluster of evoked post-stimulus activity after cluster-based correction compared to pre-stimulus baseline activity.

response to a noxious HL compared with a non-noxious CL procedure, assessed with the template of noxious-evoked activity. Using a locally-acquired dataset ($n = 119$, described in Section 2.2.4: Oxford Dataset A, and in Table 2.1) which is independent to the data used in the original template derivation (Table 2.1), we demonstrated that the EEG template's ability to discriminate between noxious HL and innocuous CL stimuli can be observed in a larger independent dataset (Figure 2.4; Table 2.3). We can use this larger dataset for more reliable effect size estimation and sample size planning to test hypothesis 1. The effect size is calculated by the standardised mean difference effect size for this within-participant design, i.e., the mean of differences of the HL and CL responses divided by the standard deviation of the differences in these responses, also known as Cohen's d_z [298]. The effect size in this dataset is 0.589 (95% CI: 0.39-0.77).

To ascertain the sample size required to test whether there is a significant

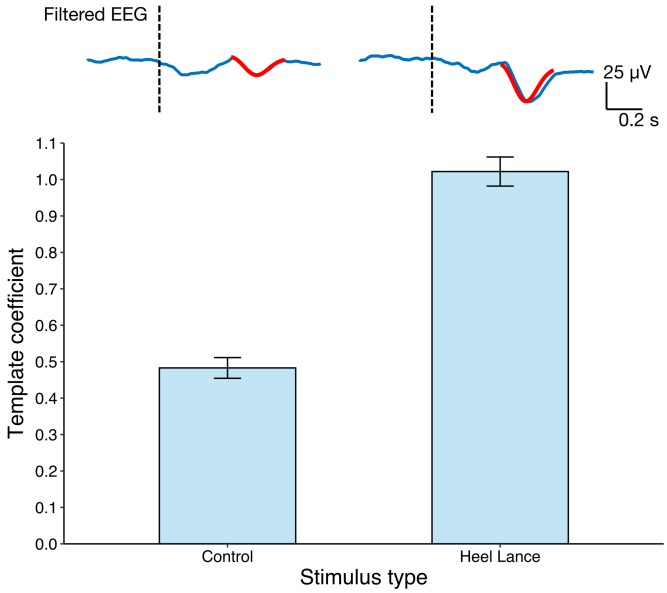


Figure 2.4: Top: Mean Woody-filtered EEG traces for all participant trials in the Oxford Pilot Dataset A ($n = 119$) for control heel lance (left) and heel lance stimuli (right) in blue, with the noxious-evoked response template fit shown in red. The vertical dashed black line for each trace shows the time marker for the stimulus (time = 0 seconds). Bottom: Mean noxious-evoked template coefficient over all trials for control heel lance ($n = 119$, mean = 0.48, std = 0.54) and heel lance ($n = 119$, mean = 1.02, std = 0.75) stimuli, with error bars showing standard error of the mean. The effect size for the difference between heel lance and control heel lance response in this sample is Cohen's $d_z = 0.589$. Abbreviations: std = standard deviation.

Measure	Mean or Mean Difference (95% CI)	Effect Size (95% CI)
Heel lance	1.02 (0.90 – 1.12)	1.35 (1.12 – 1.56)
Control heel lance	0.483 (0.41 – 0.61)	0.90 (0.69 – 1.18)
Heel lance minus control heel lance	0.539 (0.38 – 0.72)	0.589 (0.39 – 0.77)

Table 2.3: Summary metrics for heel lance, control heel lance, and comparisons between measures for an Oxford dataset of $n = 119$ infants. Confidence Intervals (CIs) are calculated by bootstrap resampling [293] using the boot library in R [292]. The effect size for the heel lance minus control heel lance is calculated as the standardised mean difference effect size for this within-participant design, i.e., the mean of the differences of the two response scores divided by the standard deviation of the differences, also known as Cohen's d_z [298]. The effect size of the individual response scores (heel lance or control heel lance alone) is in comparison to a null response, thus these effect sizes are calculated as the mean of the responses divided by the standard deviation of the responses.

difference between HL and CL response in our within-participant design, we calculated our required sample size with the software G*Power for a one-tailed t-test of matched pairs [299]. Using the effect size of 0.589 calculated from the pilot data, a power of 90%, and a reduced alpha rate of 0.02 to account for positive bias in publication, we need **35 infants** to test this hypothesis. To compensate for data loss due to artefact and participant withdrawal in the prospective study, we increased the sample size for recruitment by 10%, resulting in a total sample size of **40 infants** which would be required for recruitment for the prospective study. The UCL dataset had already been screened for artefactual EEG traces before the authors shared this dataset, and therefore the exact calculated sample size ($n = 35$) is sufficient from the UCL dataset [286].

2.3.2.2 Sample Size Planning for Hypothesis 2

The original derivation scaled the template such that the mean response to a HL in newborns at term age was equal to 1 [219]. To test the equivalence of the HL template response to 1, we calculated the mean magnitude and 90% CI of the HL response in an independent dataset and determined whether this sits within the equivalence bounds of $0.8 - 1.2$. According to FDA guidelines [300, 301] for planning bioequivalence studies, two measures which are within ± 20% of each other can be considered equivalent. In tests of equivalence, one can consider whether the 90% CI around the result is within the acceptable boundary, which in this case is a ± 20% difference of 1, and this is equivalent to carrying out two one-sided t-tests at a 5% significance level [300–302]. Therefore, if the 90% CI for a group mean HL response template magnitude sits within $0.80 - 1.20$ then this group mean value can be considered equivalent to a value of one.

The CI was calculated by bootstrap resampling [293] using the boot library in R [292]. In our Oxford dataset of 119 infants, the magnitude of the mean HL response is 1.02 with 95% CI: $0.90 - 1.12$, and the standard deviation of HL response in this sample is 0.75 (see Table 2.3; Figure 2.4). The 95% CI in these 119 infants demonstrates that the template measure of the mean HL response is equivalent to a

magnitude of one in infants aged $34.0 - 43.0$ weeks PMA (Table 2.1). This sets a broader age range to test the scale equivalence of a template measurement of mean HL response to one in independent datasets.

To plan the sample size for this hypothesis testing, I used AIPE analysis, which allows one to calculate the sample size that should result in a specified confidence interval width for a mean result [303, 304]. The inputs of AIPE sample size planning are the population mean (1.02) and standard deviation (0.75) of the HL response in the pilot data, the desired accuracy width (0.4), and confidence interval level (90%). In this case, we planned for a 90% confidence interval width no greater than 0.4; i.e., $\pm$ 20% around 1, to confirm our hypothesis. I calculated the required sample size using AIPE with the MBESS library in R [304, 305] which suggested that we needed to include **73 infants** for a 90% CI of mean HL response with an accuracy width of 0.4. To compensate for data loss due to artefact and participant withdrawal in a prospective study, we increased the required sample size by 10%, resulting in a total sample size of **80 infants** which would be required for recruitment for the prospective study. The UCL dataset had already been screened for artefactual EEG traces before the authors shared this dataset, and therefore the exact calculated sample size ($n = 73$) is sufficient from the UCL dataset [286].

2.3.2.3 Sample Size Planning for Hypothesis 3

In 2022, Schmidt Mellado and colleagues reported that the magnitude of the noxious-evoked brain activity increased with PMA at the time of study in infants studied longitudinally [228]. Using the independent locally-acquired cross-sectional sample of infants ($n = 35$, described in Section 2.2.4: Oxford Dataset B, and Table 2.1), we observed a positive correlation between infant PMA and the magnitude of the noxious-evoked brain activity (Pearson $R = 0.51$) as shown in Figure 2.5. This indicates that the finding reported by Schmidt Mellado and colleagues, that HL response correlates with PMA in premature infants, is observable in this cross-sectional sample and when the template is applied to the data. To calculate the required sample size for a one-tailed Pearson correlation test, I used G*Power for a

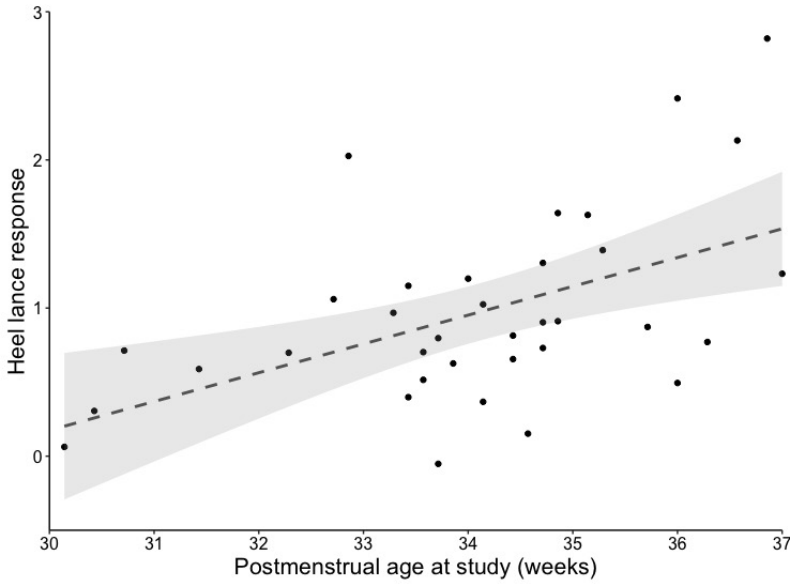


Figure 2.5: Noxious-EEG template measure of heel lance response compared to PMA for the included sample of $n = 36$ infants, where each infant is represented once. The grey dashed line shows the linear least squares regression fit to the data with 95% confidence interval shaded. There is a positive correlation between PMA and heel lance response in this sample (Pearson R = 0.51, heel lance response mean = 0.97; standard deviation = 0.65). The infants were aged $25.0 - 35.9$ weeks GA (mean = 31.1 weeks), and $30.1 - 37.0$ weeks PMA at study (mean = 34.1 weeks). Abbreviations: GA = gestational age; PMA = postmenstrual age.

calculation under the Exact test family and Correlation: Bivariate normal model statistical test [299]. Using the correlation between PMA and HL response in the pilot data, Pearson R = 0.51, a power of 90% and alpha = 0.02, with a null result represented by Pearson R = 0, the required sample size to test this hypothesis is **38 infants**. Inflating the sample size by 10% to compensate for data loss due to artefact and participant withdrawal, we find a total sample size of **42 infants** required for recruitment for the prospective study. The UCL dataset had already been screened for artefactual EEG traces before the authors shared this dataset, and therefore the exact calculated sample size ($n = 38$) is sufficient from the UCL dataset [286].

2.4 UCL and Exeter Datasets for Hypothesis Testing

In this section, I describe the final UCL Datasets for hypothesis testing which fit inclusion and exclusion criteria. Data inclusion and exclusion criteria followed the methods presented in Section 2.2.5 as published in the Registered Report [281]. The same criteria were applied for each dataset origin, where exclusion criteria were also shared across hypotheses, though inclusion criteria depended on the hypothesis (Section 2.2.5). Since the UCL Dataset is retrospective, all data were available at this stage. Preliminary Exeter Datasets are also presented, however due to the ongoing nature of the Exeter RCT into which these infants were recruited, these data are incomplete in size (prospective recruitment is unfinished) and information (data are blinded). Nevertheless, Exeter Datasets were selected and analysed at this stage for proof-of-principle of the analysis methods and to provide early indications of hypothesis test results in this dataset. All dataset demographics are summarised in Table 2.2. An overview of ethics and status of UCL and Exeter Datasets are also re-summarised here for clarity and completeness.

2.4.1 UCL Dataset

The UCL dataset is available via ReShare. Authorised access is given to users due to the sensitivity of the data ([286], UK Data Service, `https://doi.org/10.5255/UKDA-SN-853204`). I did not request data access and analyse any UCL data until after in-principle acceptance of the Registered Report by Cortex. From UCL, we included all available data from this retrospective dataset to test each hypothesis. The UCL dataset contains $n = 112$ infants from which we selected infants to test each hypothesis according to the relevant inclusion criteria.

2.4.1.1 UCL Dataset A: Data for Hypotheses 1 and 2

The demographic inclusion criteria resulted in a sample of $n = 79$ infants with GA between $31.9 - 41.9$ weeks (mean = 37.1 weeks) and PMA at study between $34.0 - 42.9$ weeks (mean = 37.9 weeks) in UCL Dataset A. The EEG data were then

checked for these 79 infants to ensure there was both one HL and one CL annotation available for these infants. After EEG pre-processing, there were $n = 74$ infants with a HL passing EEG rejection criteria, and $n = 61$ infants with both one HL and one CL epoch passing EEG rejection criteria. The dataset demographics are summarised in Table 2.2.

2.4.1.2 UCL Dataset B: Data for Hypothesis 3

Demographic inclusion criteria resulted in a sample of $n = 57$ infants with GA of $23.2 - 35.9$ weeks (mean = 31.7 weeks) and PMA at study of $29.4 - 37.0$ weeks (mean = 33.9 weeks) in UCL Dataset B. The EEG data were then checked for these 57 infants to ensure there was one HL annotation available for these infants. After EEG pre-processing, there were $n = 51$ infants with one HL annotation passing EEG rejection criteria. The dataset demographics are summarised in Table 2.2.

2.4.2 Preliminary Exeter Datasets

Ethical approval for the Exeter study is in place from the NHS REC as part of an ongoing study (reference 12/SC/0447). Written parental consent is obtained from parents prior to each study and studies will conform with the declaration of Helsinki. Pilot data collection began in June 2021 for training purposes, to ensure adequate familiarity with EEG recording equipment and protocol in advance of the planned study. However, data analysis for this study did not begin until after in-principle acceptance of the Registered Report by Cortex.

2.4.2.1 Exeter Dataset A: Data for Hypothesis 1

In Exeter, we planned to prospectively recruit and study the minimum required sample of $n = 40$ infants to test hypothesis 1 (power calculation in Section 2.3.2.1). Due to time constraints, so far only $n = 27$ eligible infants have been recruited at time of study. These infants are also part of an ongoing blinded RCT for a pain-reliving HL intervention for which the data are still blinded, and therefore the Exeter analysis is only at a pilot stage. EEG data were checked for these 27 infants to ensure there was both one HL and one CL annotation available for these infants.

After EEG pre-processing, there were $n = 23$ infants with both one HL and one CL epoch passing EEG rejection criteria. The dataset demographics are summarised in Table 2.2.

2.4.2.2 Exeter Dataset B: Data for Hypothesis 3

We did not plan to specifically recruit infants in Exeter to test this hypothesis, but to use recruited infants matching demographic criteria for this hypothesis test. Of the $n = 27$ infants recruited in Exeter at time of study already eligible for hypotheses 1, only $n = 4$ matched the demographic inclusion criteria for hypothesis 3. These infants are also part of an ongoing blinded RCT for a pain-reliving HL intervention for which the data are still blinded. Therefore, due to the very small numbers and noise due to unknown intervention status for the HLs, no Exeter data was included to test hypothesis 3 at this stage.

2.5 Results of Outcome Neutral Criteria in UCL and Exeter Data

2.5.1 Outcome Neutral Criteria in UCL Dataset

All UCL EEG datasets were combined to test for outcome neutral criteria after the EEG rejection stage. There are $n = 97$ unique HL traces from both UCL datasets A and B, since some participants are present in both datasets. There are $n = 61$ CL traces, which are entirely from UCL dataset A, since CL are not required for the third hypothesis test. Results for the outcome neutral criteria for each stimulus are shown in Figure 2.6. There are significant clusters in the post-stimulus EEG compared to the pre-stimulus EEG for each stimulus separately, using non-parametric cluster-based analysis (Figure 2.6, bottom). For the HL stimuli ($n = 97$), there is one significant cluster identified after correction, representing a noxious-evoked potential, from $502 - 1000$ ms ($p = 0.001$). For the CL stimuli ($n = 61$), one significant cluster, representing a non-noxious early potential, is identified at $179 - 455$ ms ($p = 0.009$). These clusters comprise the time-windows reported in the literature for both the noxious potential and non-noxious early potentials, respectively [212, 219,

297]. These data pass the outcome neutral criteria tests since there is significant post-stimulus evoked activity, not present in the pre-stimulus baseline activity, for both HL and CL EEG data.

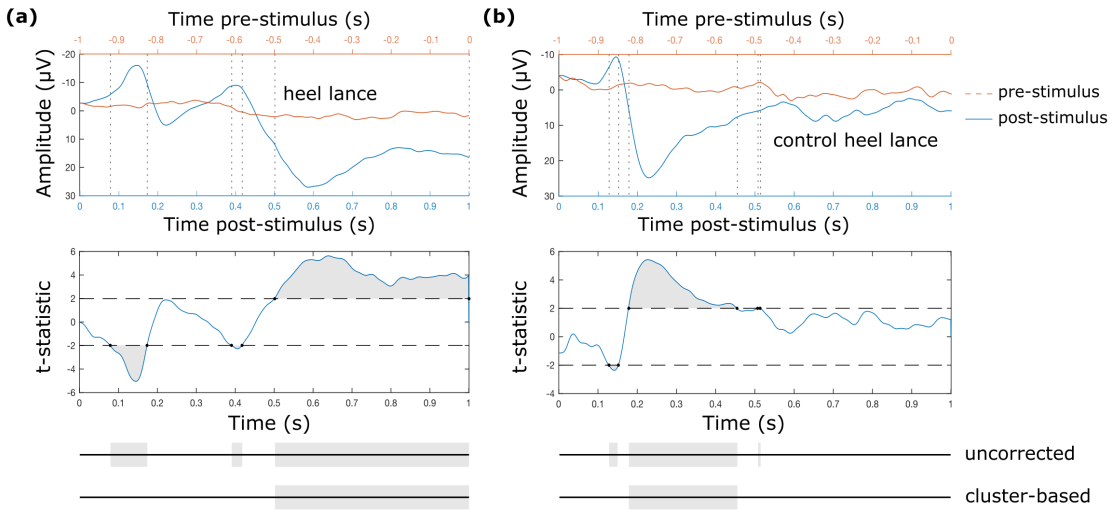


Figure 2.6: Outcome neutral criteria for all unique UCL data (a) heel lance stimulus EEG ($n = 97$) and non-parametric cluster analysis results (b) control heel lance stimulus EEG ($n = 61$) and non-parametric cluster analysis results. Top: stimulus EEG trial averages overlayed for the 1-second post-stimulus (blue) and 1-second pre-stimulus (orange), where stimuli are at 0 seconds. Bottom: non-parametric cluster analysis was performed between all participant post- and pre-stimulus EEG traces, and the shaded regions are identified regions of difference before the cluster-based correction. Clusters which are significantly different after cluster-based correction are shaded at the bottom of the figure ($p < 0.05$). Both heel lance and control-heel lance stimuli have a significant cluster of evoked post-stimulus activity after cluster-based correction compared to pre-stimulus baseline activity.

2.5.2 Outcome Neutral Criteria in Exeter Dataset

There is only one Exeter dataset at this stage of the study, Exeter Dataset A of $n = 23$ infants with CL and blinded-intervention HL EEG data. Since there are few infants in this group, no significant clusters were found for post-stimulus CL compared to pre-stimulus CL EEG data after correction. However, considering uncorrected data, one cluster was identified between $172 - 224$ ms (corrected $p = 0.2$). Furthermore, no significant clusters were found for the post-stimulus HL compared to pre-stimulus HL EEG data after correction. For uncorrected data, one non-significant cluster was identified between $567 - 756$ ms (corrected $p = 0.06$). Identifying no significant

clusters for the HL data is not unexpected, since these infants were recruited for a RCT investigating pain-relieving interventions for the HL, and the data are still blinded to investigators. Therefore, it is expected that half of the HL trials could have reduced intensity due to the intervention effect, which would reduce the signal in finding significant differences in post-stimulus activity for the HL traces compared to pre-stimulus activity. Once unblinded, we expect that splitting the infants into groups of RCT intervention and standard-care would reduce noise in comparing post- and pre-stimulus HL activity.

2.6 Hypothesis Test Results

2.6.1 Hypothesis 1 Results: Noxious-evoked activity is significantly greater than non-noxious activity measured using the EEG template

Hypothesis 1 stated that the magnitude of noxious-evoked activity for a HL should be significantly greater than the magnitude of noxious-evoked activity for a CL, for paired data measured with the template. The sample size estimated for 90% power to test this hypothesis with paired HL and CL data was $n = 35$ infants.

2.6.1.1 Hypothesis 1 Results in UCL Data

Hypothesis 1 was tested with a paired one-sided t-test of template magnitudes for $n = 61$ infants from UCL Dataset A with both one HL and one CL passing EEG rejection criteria. Noxious-evoked activity after HL was significantly greater than noxious-evoked activity after CL ($t = 6.8$, $p < 1 \times 10^{-8}$). The mean noxious-evoked template coefficient over for HL stimuli was 1.27 (95% CI: $1.08 - 1.50$) and for CL stimuli was 0.43 (95% CI: $0.35 - 0.52$), whilst the mean difference between HL and CL template coefficients was 0.84 (95% CI: $0.62 - 1.10$). A graphical representation of these results, including the EEG traces for HL and CL epochs with the template fit are shown in Figure 2.7a. The sample size of $n = 61$ is greater than the estimated sample size of $n = 35$ required for 90% power to test this hypothesis.

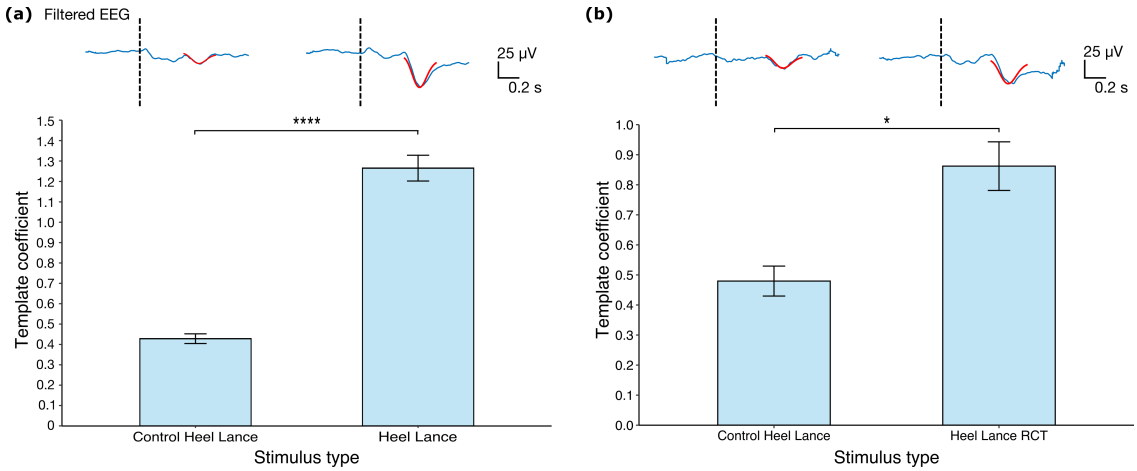


Figure 2.7: EEG results for Hypothesis 1 for (a) UCL Dataset A and (b) Exeter Dataset A. Asterisks indicate significant difference between groups (* $p < 0.05$; **** $p \leq 0.0001$). Top: Mean Woody-filtered EEG traces for control and heel lance trials are shown in blue, with the noxious-evoked response template fit overlayed in red. The vertical dashed black line for each trace shows the time marker for the stimulus (time = 0 seconds). Bottom: (a) UCL Dataset A ($n = 61$) mean noxious-evoked template coefficient over all trials for control heel lance ($n = 61$, mean = 0.43, std = 0.33) and heel lance ($n = 61$, mean = 1.27, std = 0.86) stimuli, with error bars showing standard error of the mean. One-sided t-test for noxious-evoked response for heel lance greater than control heel lance in UCL dataset is $t = 6.8$, $p < 1 \times 10^{-8}$. (b) Exeter Dataset A ($n = 23$) mean noxious-evoked template coefficient over all trials for control heel lance ($n = 23$, mean = 0.48, std = 41) and heel lance ($n = 23$, mean = 0.86, std = 67) stimuli, with error bars showing standard error of the mean. One-sided t-test for noxious-evoked response for heel lance greater than control heel lance in Exeter dataset is $t = 2.0$, $p < 0.03$. Abbreviations: std = standard deviation.

2.6.1.2 Hypothesis 1 Results in Exeter Data

Hypothesis 1 was tested with a paired one-sided t-test of template magnitudes for $n = 23$ infants in Exeter Dataset A with both one HL and one CL passing EEG rejection criteria. Noxious-evoked activity after HL was significantly greater than noxious-evoked activity after CL ($t = 2.0$, $p < 0.03$). The mean noxious-evoked template coefficient over for HL stimuli was 0.86 (95% CI: $0.67 - 1.25$) and for CL stimuli was 0.48 (95% CI: $0.34 - 0.68$), whilst the mean difference between HL and CL template coefficients was 0.38 (95% CI: $0.12 - 0.93$). A graphical representation of these results, including the EEG traces for HL and CL epochs with the template fit are shown in Figure 2.7b. The sample size of $n = 23$ is fewer than the estimated sample size of $n = 35$ required for 90% power to test this hypothesis, and the blinded HL data includes infants with either HL with or without additional intervention.

Therefore, the effect size is expected to increase in the complete Exeter dataset when including only infants with HL trials with standard-of-care without additional intervention.

2.6.2 Hypothesis 2 Results: Scale of heel lance response measured with EEG template is not site independent

Hypothesis 2 postulated that the mean noxious-evoked template magnitude for a heel lance stimulus would be equivalent to 1 in newborn term-aged infants. The sample size estimated for 90% power to test this hypothesis was $n = 73$ infants with HL data.

2.6.2.1 Hypothesis 2 Results in UCL Data

There are $n = 74$ UCL infants from UCL Dataset A with one HL passing EEG rejection criteria that can be used to test Hypothesis 2. Using bootstrap resampling with based on 10,000 bootstrap replicates to calculate CIs, the 90% CIs for mean HL template magnitude are $1.15 - 1.50$ for $n = 74$ infants from UCL Dataset A. This range extends beyond the upper bound of the pre-defined equivalence bounds of $0.8 - 1.2$, suggesting that the magnitude of noxious-evoked template response is not a measure which is independent of data collection site.

Hypothesis 2 is rejected using the UCL Dataset. The average noxious-evoked template magnitude in response to a HL in term-aged neonates is not equivalent to one across multiple sites, so this measure is not independent to site effects. The template magnitude could be influenced by site-to-site variable factors such as the EEG equipment hardware and software, EEG set-up protocol, or expertise and variable approach of technicians recording the EEG. The noxious-evoked response magnitude may also be influenced by differences in standard-of-care to manage pain in infants undergoing the HL procedure. Future studies comparing noxious-evoked template magnitude data across sites should include a confounding variable for independent recording sites.

2.6.2.2 Hypothesis 2 Results in Exeter Data

Using bootstrap resampling with based on 10,000 bootstrap replicates to calculate CIs, the 90% CIs for mean HL template magnitude are $0.69 - 1.17$ for $n = 23$ infants in the Exeter Dataset. This range extends below the lower bound of the pre-defined equivalence bounds of $0.8 - 1.2$, which would typically suggest the hypothesis is rejected. However, considering that some of the Exeter HL trials include an intervention to reduce noxious response, which would reduce the average noxious-evoked template magnitude measurement for intervention-arm HL if effective, we cannot confirm nor reject a hypothesis of equivalence of response magnitude across Oxford and Exeter sites using this blinded Exeter dataset. This test is also underpowered regardless of the presence of any intervention effects, since there are substantially fewer Exeter infants than the estimated sample size for 90% power ($n = 23$ vs $n = 73$ required).

2.6.3 Hypothesis 3 Results: Noxious-evoked response is not significantly correlated with PMA in premature infants from the UCL dataset

Hypothesis 3 postulated that the noxious-evoked activity in response to a HL would increase in premature infants with PMA up to 37.0 weeks. The sample size estimated with single HL trials for 90% power to test this hypothesis was $n = 42$ infants. Although the analysis plan for this hypothesis proposed combining UCL and Exeter data for this hypothesis test, no Exeter data was included in this hypothesis test.

Few ($n = 4$) Exeter infants matched age inclusion criteria and these data have HL stimuli blinded to intervention. Therefore, inclusion of Exeter data was more likely to add noise than provide value at this stage, where we are unable to exclude infants by intervention status. UCL dataset size was sufficient to meet sample size requirements alone ($n = 51$ infants). Hypothesis 3 is rejected since the positive correlation between HL response magnitude and PMA for UCL Dataset B infants, which matched the demographic inclusion criteria for this test, is not significant (Pearson $R = 0.19$, $p = 0.2$), as displayed in Figure 2.8.

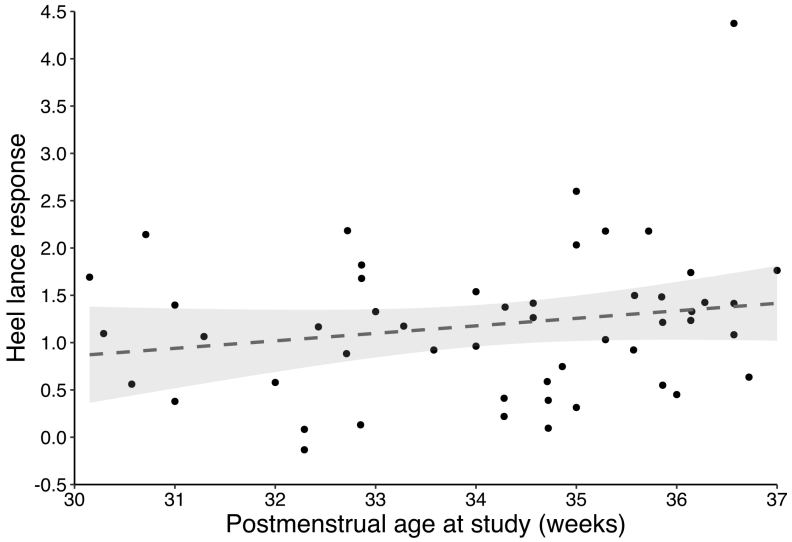


Figure 2.8: Noxious-evoked EEG template measure of heel lance response against PMA for UCL Dataset B ($n = 51$ unique infants) used to test Hypothesis 3. The grey dashed line shows the linear least squares regression fit to the data with 95% confidence interval shaded. There is no significant correlation between PMA and heel lance response in this dataset (Pearson R = 0.19, $p = 0.2$. Heel lance response mean = 1.19; standard deviation = 0.78). The infants were $23.3 - 35.9$ weeks GA (mean = 31.9 weeks), and $30.2 - 37.0$ weeks PMA (mean = 34.1 weeks). Abbreviations: GA = gestational age; PMA = postmenstrual age.

2.7 Hypothesis Test Conclusions

Whilst we are unable to finalise the hypothesis testing and complete all the pre-registered analyses to determine reproducibility in multiple datasets, due to the ongoing and blinded status of the Exeter RCT, we are able to draw some initial conclusions from the complete retrospective UCL Database.

Hypothesis 1 was confirmed in UCL data using a paired one-sided t-test, demonstrating that HL response magnitude was greater than CL response magnitude, measured using the noxious-evoked template. This supports the construct validity of the template, demonstrating reproducibility of the template measure to assess noxious-evoked response magnitude. However, Hypothesis 2 was not confirmed in the UCL data, indicating that the template measure is not consistent in scale across datasets collected in different conditions. Average template magnitude of noxious-evoked response to a clinical HL appears to be shifted greater than one in UCL Dataset A, demographically-matched to Pilot Oxford Dataset A which

had an average template magnitude of one to a clinical HL. This does not reflect on the reliability of the methods or applicability of the template to multi-site datasets, instead it indicates that researchers must account for site-site variation in EEG noxious-evoked response magnitudes, and include site as a confounding variable in any analyses using datasets across multiple sites. Many factors could influence these inter-site differences, including different approaches to standard-of-care in different hospitals, different EEG recording equipment, set-up, and personnel. Finally, Hypothesis 3 that magnitude of noxious-evoked response increases with PMA in a premature cohort under 37 weeks PMA was not supported by the UCL dataset. Therefore, we were unable to reproduce prior findings of positive correlation between noxious-response magnitude and PMA in premature in UCL data. Whether this result was not reproduced because the original data represented a false positive, or due to unknown effects arising from differences in data collection methods across sites, or differences in infant health and treatment approaches across sites, is unknown. Further investigations and future comparison to a larger Exeter dataset could help illuminate reasons behind the failure to reproduce this result.

We are greatly limited in our ability to interpret the preliminary results from Exeter Data, since some of the HL trials will include an intervention, which remains blinded until the RCT is complete. Conclusions of the preliminary Exeter analysis merely suggest that Hypothesis 1 is likely to be supported by these data, since there is a significant difference between HL and CL stimuli despite the blinded intervention status. We were unable to demonstrate significant results for the Outcome Neutral Criteria, due to low signal:noise in this limited dataset. Nevertheless, expected clusters of distinct post-stimulus activity were evident in the Exeter EEG data for both CL and HL traces, which lends promise to the quality of the final datasets. The preliminary Exeter analysis also demonstrates the capability and readiness of the EEG processing code and statistical analysis scripts to analyse data measured from this site to test the hypotheses when data collection is completed. When the Exeter data collection is completed, the RCT is complete and data unblinded for

analysis, it will be interesting to see whether hypothesis tests using Exeter data and the presented UCL results are in agreement.

2.8 Discussion

Classical methods of assessing pain, including vital signs and behavioural cues, are the current "gold standard" for assessing infant pain, despite the well-documented evidence of their practical and theoretical limitations [93, https://www.fda.gov/drugs/news-events-human-drugs/fda-m-cersi-analgesic-clinical-trial-designs-extrapolation-and-endpoints-patients-birth-less-two]. EEG approaches developed to quantify noxious-evoked changes in neonatal brain activity, initially developed by members of our research group, have been successfully used in both basic and applied infant pain research [213, 215, 219, 227], and as primary outcome measures in clinical trials to test efficacy of pain relieving interventions the analgesic efficacy of morphine in neonates ([106]; https://clinicaltrials.gov/ct2/show/NCT04901611). Despite the brain-based origin of the pain experience, and prior evidence for EEG measurement of noxious-evoked brain activity, this approach has not gained widespread use within the infant pain community. We believed that a pre-registered replication study involving both prospective data and independently collected retrospective data could help verify brain-based EEG methods for infant pain assessment and encourage their wider uptake.

This pre-registered research is highly valuable, comprising carefully planned methods and powered analyses to test three a prioi hypotheses. It is an unprecedented reproducibility study in infant brain-based pain assessment in multiple datasets, which allowed us to critically test the reproducibility of an EEG template measure of noxious-evoked brain activity to distinguish between noxious and non-noxious stimuli. The EEG template measure was shown to demonstrate significantly greater responses to noxious than non-noxious control stimuli in both retrospective and prospective datasets. This demonstrates the reproducibility and construct validity of the template as a measure of noxious response in infants. Scale equivalence of

the template measure of noxious response in infants measured in different sites was rejected. Future studies should account for site-to-site variability in infant noxious-evoked response, as magnitudes of noxious-evoked response to the same stimuli cannot be directly compared across sites. These results are valuable for designing and understanding future studies. For example, a multicentre Clinical Trial utilising the EEG template measure of noxious response should include a confound variable for centre origin to avoid erroneous conclusions. These results support that the EEG template measure is a valuable tool for distinguishing noxious from non-noxious stimuli in independent groups of infants, which can be used as an objective measure to assess the efficacy infant pain management interventions. However, this pre-registered research is not yet completed, and conclusions will be finalised once all additional evidence from the prospective Exeter data site is collected and unblinded.

2.8.1 Pre-Registration Discussion

Undertaking this research came with several unexpected challenges relating to the Registered Report process. Though peer review of a manuscript with only research methods should be more straightforward than peer review of a full manuscript with results and conclusions, the process took almost one year. These unexpected delays pushed back the start of data collection and analysis, since the study must be accepted as a Registered Report before data analysis and prospective data collection is allowed to commence. This also led to increased pressure, since the funding and resources for the study were in place and yet we were in stasis, and some resources, including myself, were time-limited. The inter-disciplinary nature of this research and the relative novelty of pre-registered manuscripts may have contributed to difficulties sourcing competent reviewers for this work. As pre-registration becomes more commonplace, researchers should be more comfortable and capable reviewing this alternative format.

Using alternative approaches to pre-registration which widen the net for peer review and maximise flexibility may improve efficiency. Platforms like PCI-RR

(Peer Community in Registered Reports: https://rr.peercommunityin.org/) allow you to schedule a peer review in advance of submission for a more efficient process, avoiding time wasted while the editor searches for reviewers after submission. Sometimes the optimal journal to disseminate information only becomes clear after results have been analysed. With some pre-registration processes you may have to commit to a journal before results are known, and not all journals offer Registered Reports. However, initiatives like PCI-RR allow for peer review and pre-approval for numerous journals without choosing the journal in advance.

Guidelines are also often unclear or incomplete for certain research designs, such as analysis of pre-existing data or replication studies, and particularly both together. It was difficult to determine the appropriate format and what is considered acceptable when combining both retrospective and prospective analysis, and particularly for a replication study, when there are lacking guidelines for both retrospective analyses and replication studies individually, let alone combined. Each journal presents their own guidelines to a different degree. Despite numerous resources from the Centre for Open Science directly (https://www.cos.io/initiatives/registered-reports) or hosted on OSF (e.g., https://osf.io/x4gzt/wiki/Preregistration%20Template/), these only presented guidelines for secondary analyses or replication separately, rather than considering both together. Similarly, there were no guidelines for pre-registration combining secondary and prospective analyses together.

Currently, there is no good pairing between Registered Reports or other peer reviewed peer registration formats, and funding models. Partially this is due to legal difficulties in collaboration between government funding bodies and corporate publishers. Having a collaborative process would be mutually beneficial in reducing duplication of pre-study review, and benefit researchers in having peer-reviewed approval of your research protocol which can garner trust and confidence from grant providers to fund your research. Combining initiatives, we could have a smoother, more open, and peer-led funding process.

Through this study, we learnt that pre-registration can be a difficult process during the early days of its adoption. Limitations of pre-registration can delay work too much to be viable in all cases, especially if there is a fixed data collection period, short-term contracts, or a collaboration with a limited time-window. It is best to decide on the registered report format early in planning stages of the study, to maximise available time and reduce pressure on the process. These experiences are based on one Registered Report submission to one journal only, and may be vastly different for other journals and research fields. In the future, I would choose PCI-RR for pre-registration over a journal-specific submission, which includes open peer review, improved efficiency, access to reviewers across multiple disciplines, and greater flexibility. Hopefully, with greater time and experience, more journals will offer a registered report pathway, guidelines will improve and turnaround times will decrease. Despite experiencing complications and delays, having an accepted Registered Report should lead to a smoother and more efficient publication process for our final results, as the methods have been pre-approved and lengthy additional analysis cannot be forced by reviewers at the final stage. Additionally, the quality of the completed reproducibility study should be stronger, with less bias and greater credibility, due to pre-registering the methods in detail.

2.9 Chapter Summary

We designed a Registered Report to test the reproducibility of an EEG template measure of noxious-evoked brain activity in infants. The research design included independent retrospective data collected at UCL and prospective data to be collected at Exeter sites. To date, only UCL data are complete and Exeter data collection is ongoing. Therefore current conclusions are based on UCL data only. These results support that the EEG template measure has construct validity, to distinguish noxious from non-noxious stimuli, but is not scale equivalent across sites. A cross-sectional dataset of premature UCL infants at less than 37 weeks PMA did not show an age relationship with noxious-evoked response. These results suggest that the EEG template can be a useful objective measure of noxious response in infants, provided

confounding factors are appropriately considered. There were numerous challenges associated with pursuing a Registered Report format, although it lends credibility to the research and ensured robust methods and powered analyses. Alternative approaches to pre-registration including PCI-RR may be more efficient and flexible for future research. Final conclusions of this replication study await completion of the Exeter data collection.

3

Developing a Pain Prediction Machine Learning Model using Infant MRI

The Machine Learning model I developed and have presented in this Chapter is included in the following publication, of which I am a co-author: L. Baxter, F. Moultrie, S. Fitzgibbon, M. Aspbury, et al. Functional and diffusion MRI reveal the neurophysiological basis of neonates' noxious-stimulus evoked brain activity. *Nature Communications*, 12(1), 2021. DOI: 10.1038/s41467-021-22960-0

3.1 Introduction

Lack of direct and objective information and overlapping signals for distress and pain make understanding the infant pain experience a difficult problem to untangle. This is not only a scientific conundrum, but is important for providing appropriate clinical care to reduce both immediate and long-term harm, for infants who experience pain. Studies have shown that repeated pain exposure in infancy leads to developmental deficiencies in later-life whilst, on the other hand, over-treatment of analgesia can also manifest with short- or long-term side-effects [66, 306]. There are large individual differences in pain sensitivity, some which can be explained by factors including sex [221], infection [68] and prematurity [217, 228]. In general however, there is always a level of inter-individual and trial-to-trial variation in infant pain response

which we are unable to explain or predict, and which affects the sensitivity of clinical trial outcome measures [122]. For the sake of infants' health, it is important to be able to understand and predict individual variation in infants' pain experiences, in order to provide care which neither over- nor under-treats pain.

3.1.1 Machine Learning in Neuroimaging and Pain Research

Machine Learning (ML) uses algorithms which take data as an input and output information in the form of predictions of continuous data or classifications, through learning patterns of association between features of the input data and the target output(s), without the explicit involvement of human intelligence in the learning process. ML models can be 'supervised' where data used to train the model is labelled, and models aim to find associations between input features of the data and different labels, or 'unsupervised' where models attempt to separate data into different categories or clusters with no prior information. ML model outputs can be used at face value to provide predictions, such as clinical outcomes or biomarkers of disease state or severity, as well as at a deeper level, to develop our scientific understanding through further investigation of the features and patterns which are important to the ML model's prediction or classification accuracy.

ML applications using neuroimaging data began in the early 2000s, and have increased rapidly in the past 10 years [307]. This includes studies which have focused on assessing and classifying pain in adults [308–310] and infants [235, 311, 312], and an exciting study using ML models of fMRI activity to investigate the efficacy of analgesic drug action [238]. However, infant studies have largely trained ML models using facial expression data and assessed accuracy by comparing to behavioural pain assessment ratings from clinicians [235, 312], disregarding the previously-outlined issues of subjectivity and non-specificity in facial expression ratings of pain in neonates. Adult ML studies, meanwhile, have included brain-based assessments of pain [229], although adult research often has the advantage of access to individuals' own verbal pain reports to use as a ground truth to train and test their models [229, 238, 308, 309]. Using neuroimaging data to assess and predict pain in neonates with

ML algorithms would provide objective improvement to this field of research, and unlock future potential to use neuroimaging-based pain predictions to study the efficacy of pain management interventions in the vulnerable infant population.

There are a variety of pitfalls to avoid when designing and implementing ML methods. Neuroscience and infant studies can be particularly vulnerable to some of these issues due to the small dataset sizes available to researchers, unlike fields which can harness big data to overcome insensitivity and unoptimised models. We can eradicate or minimise these issues through careful implementation of appropriate ML development practices, such as adequate cross-validation and preventing test-train data leakage, avoiding overfitting, valuing generalisability to additional datasets, and being cautious with conclusions and claims drawn from small-scale studies. I took care to implement a robust ML pipeline, seeking to avoid these biases.

3.1.2 Motivation and Background

The foundational component of this research focused on using functional MRI data to measure and predict neonatal noxious-evoked responses. fMRI scans can be taken under different circumstances: in rs-fMRI subjects lie in the MRI scanner at rest without any particular task; in stim-fMRI subjects complete an experimental task or have a stimulus applied, often repeatedly, during the scan. Data from both rs-fMRI and stim-fMRI can be analysed to decode stimulus-specific cortical activity, and to build models based on resting-state activity to predict stimulus response activity. The aim of this research was to develop an ML approach to predict the magnitude of stim-fMRI responses to experimental noxious stimuli in healthy infants from their rs-fMRI data. The ability to predict noxious stimulus response from rs-fMRI has potential value, as predicting an individual infant's sensitivity to pain might be used prior to painful procedures to inform clinical practice, such as implementing appropriate treatment strategies to provide adequate pain management. Additionally, whilst noxious stim-fMRI data is difficult to collect in infants due to ethical and practical reasons, it is much more straightforward and common to collect rs-fMRI in infants. Therefore, finding pain-relevant signals in rs-fMRI could allow us to

explore pain-related phenomena using larger available infant rs-fMRI datasets. This avenue could present a workaround to existing barriers in infant pain research around recruiting participants, acquiring sufficiently large and representative numbers of infants from different populations, and the challenges of time, cost and ethics in acquiring infant noxious MRI recordings. Instead, we could use pre-existing and available rs-fMRI data to more efficiently conduct exploratory research, and drive a focused and meaningful use of resources for future infant pain research.

Prior research provided central evidence to the feasibility of this work. Support came from adult research validating predictions of stim-fMRI from rs-fMRI in the verbal adult population [313], as well as research identifying similar cortical activity between adult and infant populations in response to noxious stimuli (see Figure 3.1) [177, 178]. We can derive specific RSNs from rs-fMRI data, which comprise spatially independent regions of the brain which display temporally synchronised activity at rest [314, 315]. RSNs have been shown to be consistent across subjects, and can be described by their functional relevance, such as auditory, visual, and motor networks, and are observed during their respective functionally-relevant tasks as well as at rest [313, 316]. With regards to pain research, studies have shown associations between RSNs and pain using fMRI in adults [317–319], and infant RSNs correspond to adult RSNs [160, 162]. In combination, these findings that (a) adult noxious-response activity is similar to infant noxious-response activity (b) adult RSNs correspond to infant RSNs and that (c) adult noxious-responses can be predicted from resting-state brain activity, suggested that it should also be possible to accurately predict individual infant noxious responses from RSN activity derived from fMRI.

I utilised a support vector regression machine trained in a cross-validated leave-one-out fashion on resting-state data to predict individual noxious-evoked brain activity in an fMRI dataset comprising 18 healthy infants. My specific aims were to reproduce the results from Luke Baxter's pilot ML model, then to develop and

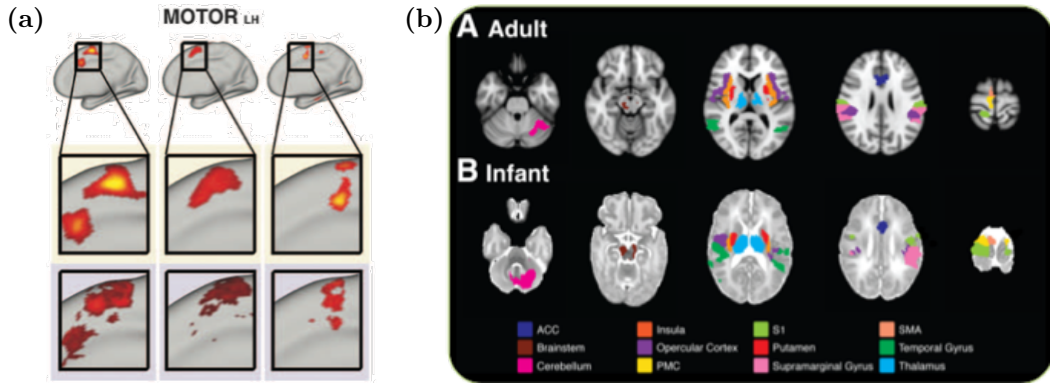


Figure 3.1: Left: Predictions of stim-fMRI patterns from rs-fMRI from a motor task in adults, from [313]. Reprinted with permission from AAAS. Right: Similarity of noxious-evoked response between adult and infant populations; 18 out of 20 possible adult-identified regions also identified in infants, from [177]. Reprinted with permission under a CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

improve the model. Firstly, I incorporated more thorough optimisation and confound-correction procedures for increased reliability and statistical robustness. The initial model contained errors including information leakage between the training and test data, due to using whole-dataset confound regression rather than robust cross-validated confound regression methods [320], and therefore I first had to consider bias in the pilot model and develop a more valid and robust pipeline. Secondly, I incorporated adult-validated pain-response signatures into the prediction model as response outcomes, rather than using only a noxious stimulus response template based on the training and test dataset itself. Using this independent signature of noxious-evoked brain activity was of additional value, since using independent signatures directly linked to pain in adults bolsters the interpretability and validity of the ML model compared to using within-dataset measures. Predictions of response measures from each of these two templates, the noxious-evoked stimulus response template (SRT, a within dataset measure) and Neural Pain Signature (NPS, a dataset-independent measure from [229]), were presented in the final publication as well as additional measures suggested by my co-authors to further demonstrate the specificity of the predictions to noxious-evoked brain activity [231].

3.1.3 Chapter Outline

This chapter outlines the development of a valid and robust ML pipeline to predict individual infant noxious-evoked responses from their resting-state brain activity. This research was conceptualised by my co-authors on Baxter et al. [231]. I developed and implemented an ML model to include key design principles to produce accurate, unbiased and generalisable predictions. The resulting ML pipeline, which I developed, was central to our research publication, Baxter et al. [231]. In the following, I will describe the development, methods, and validation of this ML pipeline, and provide results which demonstrate the accuracy of its predictions.

3.2 Methods

3.2.1 Recruitment and Research Protocols

Informed written consent for the studies were provided by a neonate's parent before the study commenced. Ethical approval was obtained from the NHS REC (reference 12/SC/0447), and research was conducted in accordance with NHS Good Clinical Practice guidelines and the Declaration of Helsinki. An overview of the on-the-day timeline and logistics for infant MRI scanning is provided in Figure 3.2. Further details of recruitment and study protocols are provided in sample consent forms and PILs for infant MRI studies in Appendix B. Detailed SOPs for MRI transport and safety, and MRI preparation and conduct, are also shared in Appendix B. General study withdrawal and safety reporting SOPs are provided in Appendix A.1.

3.2.2 The Dataset

The development of the ML Pipeline involved a retrospective dataset of $n = 18$ neonates which were recruited at the John Radcliffe Hospital, Oxford, and scanned in the adjacent Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB) at the Wellcome Centre for Neuroimaging. Neonates at the John Radcliffe Hospital were eligible for study recruitment if they were healthy, without history of neurological problems, self-ventilating in air and clinically stable at the time

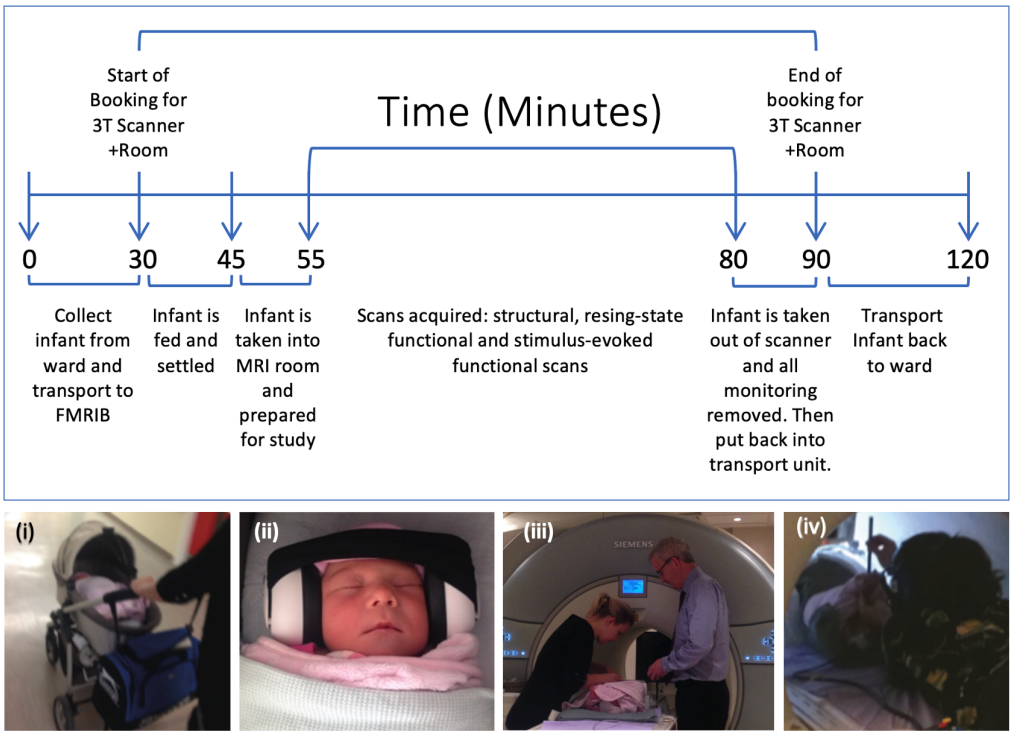


Figure 3.2: Overview of Infant MRI Scanning Protocol. The images show (i) transport of infant from the ward to the MRI scanner, (ii) infant comfortably swaddled and fitted with neonatal headphones to protect ears from the noise of the MRI scanner, (iii) infant prepped in the MRI room before scanning, including placement of head cushions and pulse-oximeter monitoring; (iv) application of stimulus to the foot during fMRI (PinPrick Stimulators, MRC Systems). Images printed with permission.

of study [177]. Included neonates ($n = 18$) were studied during the peri-term period (between $36 - 38$ weeks PMA [321]) or full-term period; they had a GA of $35 - 42$ weeks (mean = 38.3 weeks), PMA of $36 - 42$ weeks (mean = 38.7 weeks) and were scanned within the first 10 days after birth (mean PNA = 2.8 days). There were 10 male and 8 female neonates. Full demographics are presented in Table 4.1.

3.2.3 Noxious Stimulation

In the experimental design for these Oxford neonates, acute noxious (non-skin-breaking) stimulation was applied to the dorsum of each subject's left foot, using precision-force 128 mN nociceptive stimulators (PinPrick Stimulators, MRC Systems). These stimuli were repeated 10 times for a length of 1 s each, and with minimum inter-stimulus interval of 25 s. The stimuli were time-locked to the MRI recording using Neurobehavioural Systems software, version 20.1 (`https://www.neurobs.co`

m/) [177, 231]. This mild non-skin-breaking noxious stimulus is known to evoke brain activity closely resembling that evoked by skin-breaking clinical procedures associated with pain, including inoculation and heel lancing [219, 221]. However, the evoked brain activity is much lower in amplitude in response to the pinprick stimuli compared to clinical skin-breaking procedures and, unlike tissue-breaking procedures, the pinprick stimuli do not lead to significant increases in behavioural indicators of distress or clinical pain scores in neonates [225]. This pinprick stimulus is therefore a useful experimental tool to develop our understanding of nociceptive processing in neonates whilst alleviating ethical concerns of causing distress and pain in this population.

3.2.4 Functional MRI Recording

All MRI data from these infants were collected at FMRIB on a Siemens Prisma 3T scanner with an adult 32 channel receive coil. The functional image acquisition was as described in Baxter et al. [231]. The scan comprised T2* BOLD-weighted, GRE, EPI readout TE = 50 ms [232], TR = 1300 ms, mean TA = 6 min (approx.), multiband 4 [322, 323], with 56 slices, 2 mm isotropic voxels, and a single-band reference image acquired at the start.

3.2.5 Resting-State Networks

RSNs were inferred from rs-fMRI using PROFUMO (PRObabilistic FUnctional MOdes) [324, Codebase: `https://git.fmrib.ox.ac.uk/samh/profumo`]. We performed PROFUMO analysis with a customised neonatal haemodynamic response function (HRF) from Arichi et al. as a prior [233, 325], to extract biologically plausible RSNs from both $n = 18$ Oxford subjects and $n = 242$ subjects from the dHCP [155, 231]. PROFUMO was used for RSN selection, since this method decomposes fMRI data into spatial and temporal modes, but unlike other approaches such as Independent Component Analysis (ICA), PROFUMO uses a Bayesian approach which can incorporate the customised neonatal HRF, varying signal:noise, and can estimate subject-level as well as group-level networks. PROFUMO only

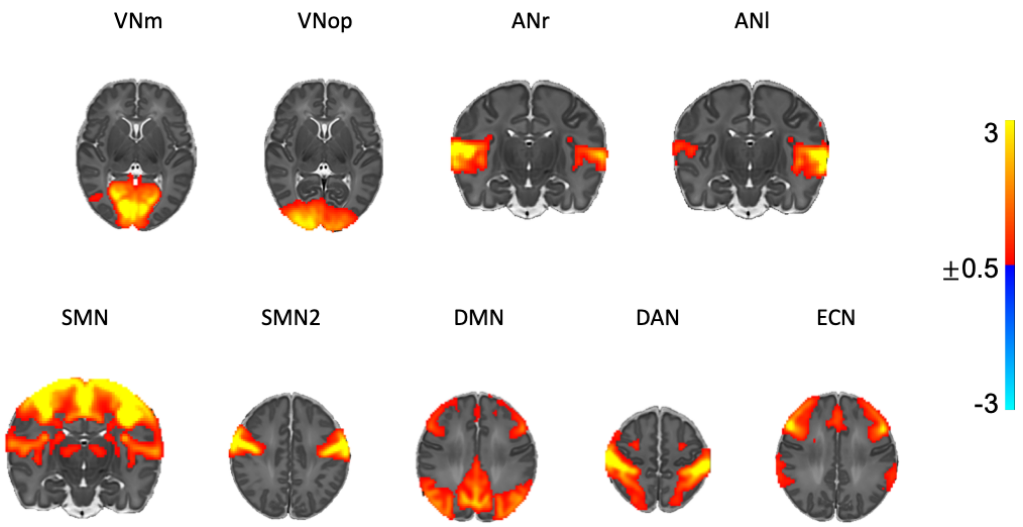


Figure 3.3: Nine thresholded group-level resting-state network maps from $n = 242$ dHCP subjects [231]. Arbitrary thresholding was for illustration purposes, and the colour bar represents posterior means. VNm = medial visual network, VNop = occipital pole visual network, ANr = right auditory network, ANl = left auditory network, SMN = somatomotor network, DMN = default mode network, DAN = dorsal attention network; ECN = executive control network. Adapted from [231] and printed with permission under a CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

extracts component RSNs which are supported by the data, therefore the number of extracted RSNs will be impacted by sample size and signal:noise, but we can have greater confidence in the results. Fitzgibbon et al. identified 16 RSNs in dHCP infants [155], and we identified 9 RSNs which were replicated across both Oxford and dHCP datasets [231]. The fact that these 9 RSNs were independently identified and replicated across two independent infant datasets, demonstrates that they are the more robust and reproducible components of neonatal resting-state activity in these datasets, and thus these 9 RSNs were used for subsequent analysis to maximise generalisability of the model [231]. Since they were calculated from a larger dataset with a better signal-to-noise ratio, the dHCP templates of these 9 selected RSNs were used to calculate RSN amplitudes in each neonate (Figure 3.3), rather than the Oxford-derived RSN templates.

Predictors for stim-fMRI response in the ML model were based on the amplitudes of these RSNs in each subject. Amplitudes were calculated from the median absolute deviation (MAD) of the general linear model (GLM) regression of subject rs-fMRI

timeseries with RSNs. Whilst network amplitudes from timeseries fMRI data are commonly calculated as using the standard deviation, MAD was chosen as a more robust metric [231], since it is less susceptible to noise and outliers of timeseries spikes, which are likely a result of infant head motion. All GLM calculations for this work relied on FSL (The FMRIB Software Library) [326, 327].

3.2.6 Confounds

Confounds are properties of the subject data considered to introduce noise or have their own nuisance trends with the rs- or stim-fMRI across the dataset. If the goals were purely driven by predictive accuracy, removing confound effects is not always necessary since the labelling of "confound" is a philosophical one dictated by what your hypotheses and research aims are [328]. However, to generalise application to additional datasets with different properties to the training data, including real-world datasets, understanding and removing confounding factors is important for model accuracy and reliability. Furthermore, removing confounds allows interpretation of underlying relationships in the predictions e.g. the specific significance of functional brain activity and neurological development on noxious responses.

Confounds we corrected for in these data were mean head motion, and amplitude measures of cerebrospinal fluid (CSF) and WM in the brain. These confounds impact quality and detail of fMRI scans so may account for variations in noise across subjects, which we do not want to inform the model instead of the noxious-evoked component of the BOLD fMRI signal. WM amplitude was chosen to capture between-subject variability in global signal amplitude, which is a known confounder of rs-fMRI data that was not removed in our fMRI pre-processing pipeline [329]. CSF amplitude was chosen to capture between-subject variability arising from residual cardiac pulsatility signals [330]. We did not correct for factors such as age, since these can have interesting and important biological implications for brain development and pain sensitivity. More precisely, noxious-evoked response amplitudes from stim-fMRI were corrected for (1) mean framewise displacement of the head, (2) stimulus-correlated head motion, and (3) CSF signal amplitude throughout the stim-fMRI scan. The

stim-fMRI CSF signal amplitude was calculated as the mean regression coefficient of the infant's noxious-response map within the CSF region of interest (ROI) according to the Schuh neonatal atlas [331]. Meanwhile, resting state network amplitudes from rs-fMRI were corrected for (1) mean framewise displacement of the head, (2) CSF signal amplitude, and (3) WM amplitude throughout the rs-fMRI scan [231]. WM and CSF ROIs were defined from the Schuh neonatal atlas [331], and in each case 'amplitude' was calculated as the MAD of the signal timeseries for each measure. We also demonstrated that a prediction model trained on these confounds does not predict infant noxious-evoked responses [231].

3.2.7 Within-Dataset Stimulus Response Template

One method used to measure the noxious stim-fMRI response relies on a "stimulus response template" (SRT) derived from Oxford infants' brain-activity during experimental noxious stimuli. This template was formed from the group pseudo t-statistic map of every Oxford subject's stim-fMRI contrast of parameter estimates (COPE) maps. "Pseudo" refers to the smoothing of subject variance, modifying the traditional t-statistic method [332]. We confirmed that this group-average response template included brain regions classically considered part of the adult nociception pain system [231, 333]. The SRT was derived by Luke Baxter, and its methodological development is discussed in detail in his DPhil thesis [334]. The other method of measuring noxious stim-fMRI response relies on the independent adult-derived NPS which is described further in Section 3.2.8 [229]. The SRT response score attempts to reflect an overall measure of the infant's brain activity in response to noxious stimuli, whilst the NPS is a signature of brain activity that has specifically predicted reported pain intensity in adults. Using these two different templates of noxious-evoked brain activity enables us to analyse the specificity and sensitivity of our measurements (in stim-fMRI) and predictions (from rs-fMRI) of pain response. Further details of these subject response measures and their use in the ML model are described in relevant sections in the text (Section 3.3).

3.2.7.1 Calculating Noxious Response Amplitude

The method to calculate an individual noxious-evoked response amplitude using the within-dataset SRT is as follows. The SRT is derived from all subjects' stim-fMRI COPE maps, and then each individual response score is derived from the GLM regression of the subject stim-fMRI COPE map against this group SRT [231]. This spatial regression coefficient is a scalar value representing the infant's noxious-response amplitude.

3.2.8 Neural Pain Signature (NPS)

Several pain-related signatures have been described in adults, including the NPS [229], Stimulus Intensity Independent Pain Signature [14], Picture-Induced Negative Emotion Signature [335], Heart Rate Stress [336] and Vicarious Pain [337]. For our infant study we are primarily interested in the NPS, an intensity-encoded adult signature for response to physical noxious stimuli which was validated using pain reports and non-noxious equivalent stimuli (e.g. noxious thermal, and non-noxious thermal stimuli). The NPS is specific to physical, personal, pain not including threat, vicarious or emotional pain, and has been validated for a broad range of noxious stimuli in adults including thermal, mechanical, and electrical. Figure 3.4 shows areas of fMRI-identified brain activity comprising the NPS [229]. These embody both positive predictive weights, regions of activity positively associated with intensity of reported pain response, and negative predictive weights; regions negatively associated with predicted pain response. It is important to note that the NPS doesn't explain the stimulus intensity-independent modulation effects in pain response, e.g., pain modulation due to placebo and psychological interventions. The NPS was transformed to infant space for this research by Eugene Duff [178, 231].

3.2.8.1 Calculating Noxious Response Amplitude

The method to calculate an individual noxious-evoked response amplitude using the NPS is described here. Using the signature validated by Tor Wager et al. [14, 229] required a different measure of similarity to the GLM used for regression of

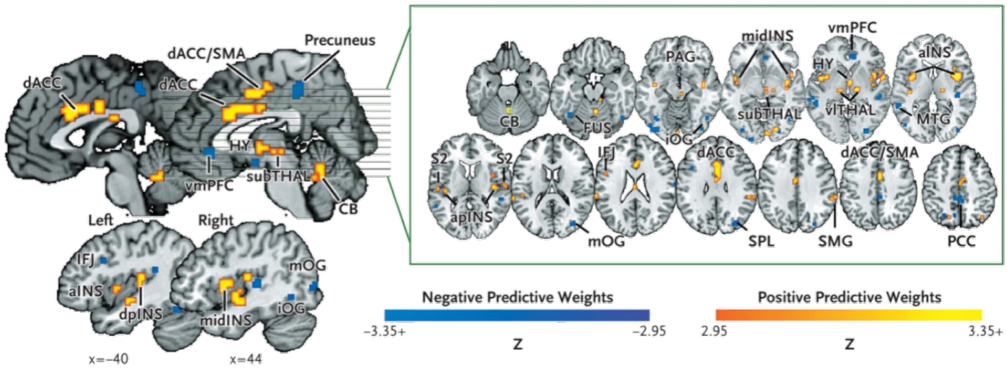


Figure 3.4: fMRI activity comprising the Neural Pain Signature (NPS), adapted from Wager et al. [229]. The overall NPS is represented, but it can also be split into components of positive predictive weights for pain response (yellow-red) and the negative predictive weights for pain response (blues). Reproduced with permission from [229], Copyright © 2013, Massachusetts Medical Society.

stimulus response COPE maps against the Oxford dataset derived stimulus response template mentioned previously. In preliminary analyses, a simple GLM to extract the signatures' contributions from stim-fMRI was ineffective since outliers dominated, but the pattern-based cosine similarity measure between the subject stimulus fMRI COPE maps and the NPS proved more effective. Cosine similarity is more robust to outliers since it is influenced less by differences in the strength of responses, and more by the pattern. Thus, a similar response pattern will give a strong positive similarity even if the relative response amplitude is weaker. This is also important since the NPS was derived in a very different dataset, of adults, and thus intensity magnitudes in the adult population may not be appropriate for quantitative comparison of similarity in infants, but pattern and direction of intensity in the signature is key. Deriving response scores from cosine-similarity to the adult-validated NPS template in infant-space [178] also reproduces the methodology of original authors of the NPS, who used cosine-similarity to quantify adult responses [229].

3.3 Machine Learning Pipeline Development

My contribution to the research publication by Baxter et al. [231] was in developing and implementing a robust ML pipeline, ensuring it was well-designed to avoid biases, and incorporating various improvements from a basic starting template proposed

by my colleague, Luke Baxter. Baxter initially proposed a basic linear support vector regression (SVR) model as part of his research plan to demonstrate that individual infant nociceptive sensitivity can be predicted from resting-state brain activity [231]. I built upon this work to eliminate issues with data leakage in the initial model, test alternative machine-learning models, and implement several optimisation procedures to reduce bias and increase generalisability, based on machine-learning best-practices [320].

In brief, I added a formal automated optimisation step, z-scaling, and replaced confound regression on the whole dataset, for both predictors and response scores, with cross-validated confound regression to remove the bias of data leakage between test and training datasets. In the following, I will describe the basis for each of my optimisations, along with the final ML pipeline and prediction results.

3.3.1 Choice of ML Algorithm

Linear SVR was initially chosen over ordinary least squares (OLS) linear regression since SVR is more robust to outliers (loss $\sim$ error) than OLS regression (loss $\sim$ error^2). In addition to linear SVR, I explored alternate algorithms of greater complexity, such as non-linear SVR using an *rbf* kernel. However, I found that alternate non-linear regression algorithms yielded no improvement or even degraded the prediction accuracy compared to linear SVR. Furthermore, research has shown that in small sample datasets in the range of tens to a few hundred, linear SVR regression is as accurate as more complex models including non-linear SVR and ridge regression, whilst performing significantly better then LASSO and OLS regression [338]. Given these results, I focussed on developing an optimised linear SVR ML pipeline.

3.3.2 Improvements and Steps to the ML Pipeline

The ML pipeline was optimised for a linear SVR model in which predictors for each subject were amplitudes of RSN activity, and responses were either of two different measures of subject noxious stim-fMRI activity as described in the methods (Sections 3.2.7; 3.2.8). The linear SVR model was implemented using scikit-learn

packages (version 0.21.3) in the Python™ programming-language (version 3.7.4) [339]. Early versions of the Python™ script were based on MATLAB® code by Luke Baxter, although later versions diverged from that implementation as the method was developed. The full ML pipeline including optimisation and fitting is summarised in Figure 3.5. The pipeline comprises pre-processing steps for input variables, and the subsequent fitting procedure gives the predicted responses. Permutation testing is then performed to evaluate the model.

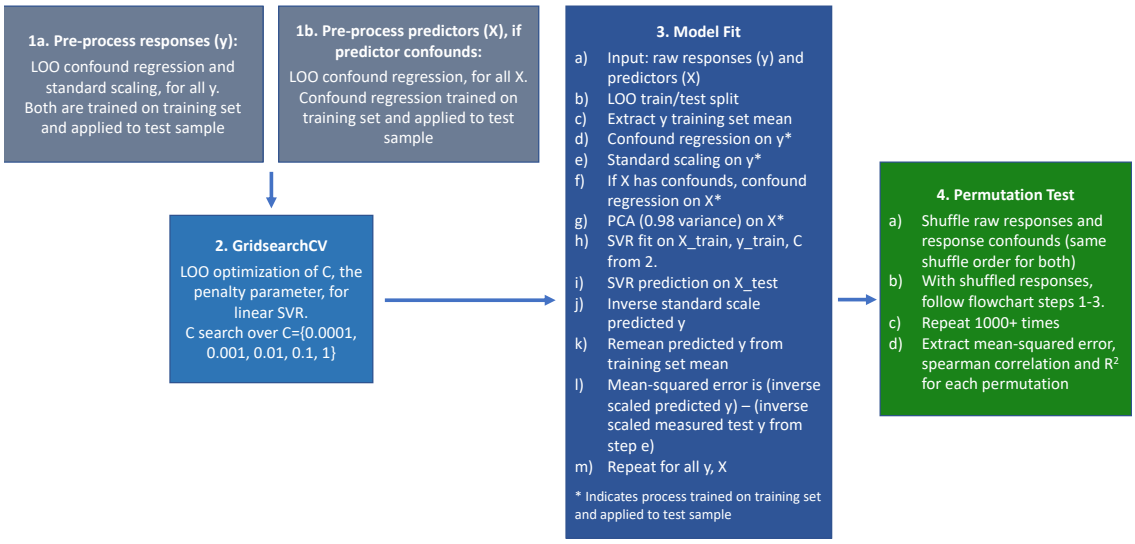


Figure 3.5: Flowchart of Machine Learning Pipeline. This pipeline was designed to optimise the prediction model, as discussed in detail in the text. Steps 1a and 1b can occur in parallel, but the following steps 2, 3 and 4 occur consecutively after each step is complete. Abbreviations: C = penalty parameter for the error term in the SVR model [339], CV = cross-validated, LOO = leave-one-out, PCA = principal components analysis; y = responses, X = predictors, SVR = support vector regression model.

3.3.2.1 Pre-processing Inputs

Some processing of raw responses and predictors is performed to improve model accuracy and generalisability. For both the responses and predictors, we include a confound correction step known as cross-validated confound regression (CVCR), in a leave-one-out fashion [328]. Leave-one-out CVCR involves fitting a least-squares linear regression model containing each confound variable for the training data (all but one held-out subject), then applying the linear regression to the test sample (held-out subject). New confound-regressed predictors/responses are the residuals of this

regression. The cross-validated leave-one-out (CV-LOO) nature of this processing ensures that the deconfounding procedure does not introduce signal into the training or test data that biases the ML predictions, and ensures that the ML model may be fed unknown subject data and make reasonable predictions rather than over-fitting to a specific, limited, dataset. Snoek et al. has shown that CVCR on both predictors and response variables is the least biased way to investigate relationships in neuroimaging data [328]. Code for CVCR was sourced from Snoek et al. and implemented in our CV-LOO pipeline [328, 340, Codebase: `https://github.com/lukassnoek/MVCA`].

For the responses there is an additional pre-processing step. This is applying 'z-scaling' via the scikit-learn StandardScalar function, which rescales each test subject's score by subtracting the mean of the training data scores and dividing by the standard deviation of the training data scores. Importantly, this z-scaling is also performed in a cross-validated leave-one-out fashion, since otherwise global variables of dataset mean and variance would be introduced into the responses and skew the model. This re-scaling improves the performance of the gridsearch step which optimises model parameters.

3.3.2.2 Optimisation

In essence, the optimisation procedure uses a CV-LOO gridsearch function to iteratively evaluate the mean squared error (MSE) of predicted responses against calculated responses, for a set of different SVR models with different parameter options. Whilst I initially tested alternative non-linear approaches as well, predominantly using the radial basis function kernel, they delivered inferior predictive accuracy and all subsequent analyses concern the linear SVR. For this linear SVR, the only parameter being searched is $C > 0$, the penalty parameter for the error term in the scikit-learn implementation of SVR [339, `https://scikit-learn.org/stable/modules/svm.html#svm-regression`]. Larger C means less regularisation, and the SVR puts greater weight to minimising error for any particular subject. Increasing C is not necessarily beneficial, since in cases of noisy data and/or outliers, a lower C is better to improve regularisation of the estimation. Furthermore,

it increases the time required for training and fitting, although negligible in this study, and the MSE from increasing C will meet a threshold and plateau. Hence, gridsearch ensures the best choice of C for our SVR model.

3.3.2.3 Model Fitting

Once the CV-LOO gridsearch step has been completed, and the optimum C has been identified, it is stored for implementation in the model fitting step(s). The fitting procedure of the ML pipeline is broadly similar to optimisation, with a few notable exceptions; principal components analysis (PCA) to reduce noise in the predictors, and producing the predictions themselves. The PCA step is performed across the nine RSNs to select the principal components (PCs) which explain 98% of the variance in the RSN data. Choosing PCs by 98% of explained data variance is intended to improve signal:noise without compromising model information. For the Oxford training dataset, this translates to selecting 7 or 8 PCs in each LOO training set.

Following all the appropriate processing steps described, and shown in Figure 3.5, the SVR model can finally be trained on the training set (all but one subject) of processed predictors and response measures, and a prediction made for the final held-out subject's response (whose rs-fMRI data has not been supplied in the training of the SVR). Each subject is iteratively held-out, keeping in all other subjects' data, and a predicted response stored for each subject after complete re-fitting of the SVR (all steps under 3. Model Fit in Figure 3.5). Once all subject predicted response scores have been collected, the prediction accuracy can be assessed since scores can be directly compared to measured stim-fMRI responses for the Oxford dataset. Primarily, root mean squared error (RMSE) between predicted and measured responses was used to assess accuracy.

3.3.2.4 Model Evaluation

Prediction model evaluation metrics were three-fold: RMSE, as noted in the model fit, as well as Spearman's rank correlation coefficient or "Spearman R", and the

coefficient of determination, R^2, which measures goodness-of-fit of a prediction model. These three metrics provide complementary information to evaluate model performance. RMSE is useful since it directly quantifies the error between predicted and true calculated response metrics. Since we are using two response metrics (the SRT GLM and NPS cosine-similarity) with different scales, we also report the Normalised RMSE (NRMSE) which divides the RMSE of the predicted responses by the difference between the maximum and minimum of the 'true' calculated responses. This enables us to more easily compare performance between models using different response measures. Spearman R is useful since it quantifies to what degree the model accurately ranks predictions from lowest to highest, reflecting the ability of the model to determine an individual infant's pain sensitivity on a relative scale within the group. The correct ordering of the predictions from lowest to highest calculated values does not necessarily compare with a low RMSE. Finally, R^2 is a standard measure of the goodness-of-fit of a prediction model which explains the proportion of variation of the response variable which is predicted by the model predictors. Typically, R^2 should range between 0 and 1, where 1 denotes perfect prediction (all but unachievable in scientific prediction models) and 0 denotes that the model consistently predicts the group mean of the response variable; a flat horizontal line on a plot of predicted vs calculated response. However, R^2 can be negative in cases where the prediction model performs worse than the mean (which gives an $R^2 = 0$), which can occur if using an inappropriate model (e.g. a non-linear model for a linear system), due to model-fitting to a mismatched system, extreme outliers, or mistakes in implementation [341]. Therefore, an increasingly positive value of R^2 indicates improved model prediction capability, whereas a negative R^2 is also a useful indicator of potential mistakes or inappropriate modelling to be fixed.

Permutation tests, where subject labels on responses were shuffled randomly before implementing the entire ML pipeline, were performed to calculate the statistical significance of the results as assessed by the three metrics of RMSE (or NRMSE), Spearman R, and R^2, and compute the p-values associated with each metric. Each permutation test comprised 10,000 permutations of the data which

were separately fed into the entire ML pipeline. The p-values were computed as a one-tailed significance test for each metric for the true order of data compared to each distribution of metrics from the permuted data.

3.4 Final Machine Learning Pipeline Results

The results comprise predictions using RSN amplitudes in the rs-fMRI data to predict two different response features calculated from two different templates applied to the stim-fMRI data. Whilst the predictors (RSN amplitudes) are the same in both models, the stim-fMRI-derived responses the model is trained on, and predicts, are different. Firstly, we use response scores derived from the GLM regression of subject stim-fMRI COPE maps against the stimulus response template derived from all subjects' stim-fMRI COPE maps [231]. Secondly, we use response scores derived from cosine-similarity to the adult-validated NPS in infant-space [178], which follows the similarity measure in Wager et al. [229] (see Section 3.2.8 for more information on the NPS). The first response score tries to produce an overall measure of the infant's brain activity in response to noxious stimuli, whilst the second is a signature of brain activity that has specifically predicted reported pain intensity in adults. Using these two different templates of noxious-evoked brain activity enables us to analyse the specificity and sensitivity of our measurements (in stim-fMRI) and predictions (from rs-fMRI) of pain response.

We found that both response measures can be robustly predicted from rs-fMRI activity (Figure 3.6). The stimulus-response template predictions have root mean-squared-error, RMSE = 1.57 ($p = 0.0005$), $R^2 = 0.64$ ($p = 0.0005$) and Spearman R = 0.79 ($p = 0.001$), whereas NPS predictions have RMSE = 0.025 ($p = 0.01$), $R^2 = 0.42$ ($p = 0.01$) and Spearman R = 0.62 ($p = 0.02$). The different magnitudes of RMSE relate to the different magnitudes of the two response measures, so we can compare across models using the Normalised RMSE. In this case, the NRMSE = 0.20 ($p = 0.0005$) for the SRT predictions and NRMSE = 0.25 ($p = 0.01$) for the NPS predictions. It is also worth noting that there is high correlation between

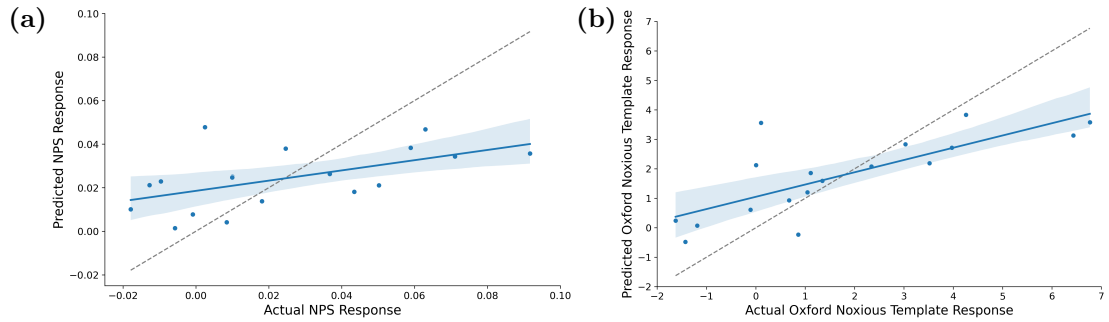


Figure 3.6: Final Machine Learning pipeline predictions for $n = 18$ Oxford infants. (a) using the adult-derived NPS (RMSE = 0.025 ($p = 0.01$), with NRMSE = 0.24, $R^2 = 0.42$ ($p = 0.01$) and Spearman R = 0.62 ($p = 0.02$)) (b) using the Oxford group-average SRT (RMSE = 1.57 ($p = 0.0005$), with NRMSE = 0.20, $R^2 = 0.64$ ($p = 0.0005$) and Spearman R = 0.79 ($p = 0.001$)). Abbreviations: NPS = Neural Pain Signature, (N)RMSE = (Normalised) Root Mean Squared Error; SRT = Stimulus Response Template.

the two measured response values (Pearson R = 0.81) as well as predicted response values (Pearson R = 0.87). Across all these metrics, the model using the within-dataset SRT response measure outperforms the model using the adult-validated NPS response measure. However, we are testing the model only with the dataset within which the SRT was derived, so we have no information on potential over-fitting or the generalisability of each measure to external datasets.

3.5 Discussion

These results support that rs-fMRI can be used to predict stim-fMRI responses to experimental noxious stimuli in infants, using a robust model. All steps were taken to prevent leakage of test data into training the model, including cross-validated correction procedures, to prevent bias in the prediction model. The prediction accuracies achieved suggest it may be possible to infer an infant's sensitivity to noxious-evoked stimuli, as a surrogate measure of their pain-sensitivity, from nociception-free RSN activity. The proven effectiveness of the model which incorporates the adult pain-validated NPS, along with prior research mapping infant noxious response activity to adult pain response, bolsters the interpretation that these noxious-evoked responses link to pain sensitivity. Whilst the within-dataset template measure gave greater model performance metrics on all counts, the use

of the adult-validated NPS is potentially more generalisable to external datasets. This hypothesis remains untested without access to adequate independent data. Additionally, the high correlation between the NPS and within-dataset template suggest that both are similarly measuring nociceptive sensitivity in this dataset.

Whilst we observed that the RSNs used as predictors in this model do not contain notable pain relevant regions from adult studies, namely the amygdala and orbitofrontal cortex, the fMRI research from Goksan et al. did not find these regions to be activated in response to the experimental noxious stimuli [177]. Although these regions may be activated in response to more painful stimuli in infants that we have not observed in prior studies, we do not have the evidence to suggest that is the case, nor that it could improve the accuracy or generalisability of this prediction model in neonates. This study acts as proof-of-concept for modelling and predicting individual infant pain sensitivity from nociception-free resting-state activity, and there are many steps yet to take in order to develop a clinically viable tool to pre-emptively manage pain in neonates. For example, translation to EEG methods or other cot-side imaging techniques would increase the clinical viability and potential impact of this approach.

Implementing MRI-based biomarkers of pain and learning from pain-prediction models is an exciting future possibility, particularly in clinical trial applications which have controlled environments and access to scanners [342]. However, a prediction model of pain based on rs-fMRI is not currently clinically practical, and future work could convert this model into more realistically implementable systems. In particular, if this framework could be applied to EEG data, which is cheaper and easier to acquire even cot-side, it would be much more accessible to clinical practice and clinical trial research. Another direction for this research is developing our understanding of the underlying relationships that allow us to predict noxious-evoked responses from resting-state activity. Understanding individual infants' pain sensitivity, from more fundamental and accessible measures than fMRI activity, could help towards developing better treatment plans tailored to infants. It is evident

from the measured noxious-response data that there is a high degree of variability in individual noxious sensitivity, and therefore individual infants may require different levels of pain management in a clinical setting. Explanations for this variability can be a mixture of genetic and environmental factors [343], as well as infants being at different stages of the rapid development of their brain in early life which impacts noxious signal processing [212, 217, 228]. Given the nature of the underlying signals used in this accurate prediction model, where predictors are RSN activities, the amplitudes of major functional networks and the sensitivity of an infant to noxious stimuli may reflect the infant's developmental stage. Baxter et al. explored and uncovered correlates between the underlying functional RSN activity in these infants and measures of their structural brain development [231]. We hypothesised that noxious responses were positively correlated with brain maturation, motivating my research in the next Chapter, which aimed to test the maturation hypothesis.

Given the small sample size limitations of this study, and the expense and difficulties associated with running larger scale infant MRI studies, we relied on data from the pioneering dHCP to explore these relationships to a greater extent. Further to the work presented in Baxter et al., on how individual nociceptive sensitivity relates to both structural and functional development in dHCP and Oxford data, in the following Chapter I will explore the relationships between predicted noxious response and measures of neurodevelopment, particularly in relation to prematurity, in dHCP datasets.

3.6 Chapter Summary

I developed an ML model to predict infants' noxious-evoked response magnitudes from their rs-fMRI activity. I optimised a linear support vector regression ML model, including best-practices such as nested cross-validation to prevent data leakage and scale normalisation to reduce bias in the predictions. I tested the ML model with two template measures of noxious-evoked activity: the Stimulus Response Template derived from stim-fMRI data of the Oxford dataset of 18 infants, and

a sample-independent Neural Pain Signature. Prediction accuracy was evaluated with three complementary statistics in a dataset of 18 infants; RMSE, Spearman R, and R^2, using permutation testing to evaluate statistical significance. ML models were able to accurately predict noxious-response magnitudes from rs-fMRI activity using each stim-fMRI noxious-evoked response template measure. These results demonstrated the validity of predicting noxious-evoked response from resting-state activity in infants.

4

Predicted Pain in Infants, Early and Long-Term Neurodevelopment

The Machine Learning model I developed and applied in this Chapter is presented in the following publication, of which I am a co-author: L. Baxter, F. Moultrie, S. Fitzgibbon, M. Aspbury, et al. Functional and diffusion MRI reveal the neurophysiological basis of neonates' noxious-stimulus evoked brain activity. *Nature Communications*, 12(1), 2021. DOI: 10.1038/s41467-021-22960-0

4.1 Introduction

There are two over-arching motivations for this Chapter, which focusses on exploratory research utilising the large dHCP neuroimaging dataset. The overall aim was to investigate factors associated with pain predictions in dHCP data, using the pain prediction model outlined in the previous Chapter. The primary motivation was to explore relationships of brain structure and function underlying inter-individual infant pain sensitivity. This involved developing the exploratory findings presented in Baxter et al. that linked predicted noxious-evoked response intensity to features of structural and functional development in the Oxford dataset and age-matched dHCP infants [231]. The second motivation was to explore whether predicted pain sensitivity is associated with health-related metrics which have been correlated to

early pain exposure, such as prematurity and neurodevelopmental assessment at 2-year follow-up. This latter motivation concerns whether relationships between early pain exposure, prematurity, and neurodevelopmental outcome could be modulated by nociceptive sensitivity. Would increased nociceptive sensitivity lead to reduced neurodevelopmental outcomes (which have been related to pain exposure), due to greater sensitivity to painful procedures? Are infants with reduced nociceptive-sensitivity protected from adverse neurodevelopmental outcomes? Is prematurity associated with heightened or reduced nociceptive sensitivity that could relate to increased pain exposure or to developmental differences compared to term neonates? Each of these questions could be exploratorily probed using longitudinal dHCP data.

4.1.1 Structural and Functional Correlates of Nociceptive Sensitivity

Our analysis in Baxter et al. showed that increased noxious-evoked responses correlated with metrics indicating greater brain maturation, implying that noxious sensitivity in early life increases with age and brain development [231]. This was based on observations that neonatal within-dataset SRT noxious-evoked responses increased with increasing RSN activity, increasing FA and decreasing MD in white matter tracts [231]. White matter MD typically decreases with age in the developing infant brain, since MD is a measure of overall diffusivity in all directions, which decreases due to decreasing water fraction compared to tissue content with brain development [344]. White matter FA measures the directionality of water diffusion which increases with infant brain development due to increased oligodendrocytes and myelination and greater organisation of brain tissue [344].

Prior exploratory research identified five nociception-relevant tracts which are coupled to brain regions key to pain and nociception processing [231]. The five tracts, displayed in Figure 4.1, are the superior thalamic radiations (*str*), corticospinal tracts (*cst*), anterior thalamic radiations (*atr*), uncinate fasciculi (*unc*) and forceps minor (*fmi*). Lower MD and higher FA in these tracts were correlated with increased noxious-evoked response [231]. These tracts are implicated in nociceptive function by

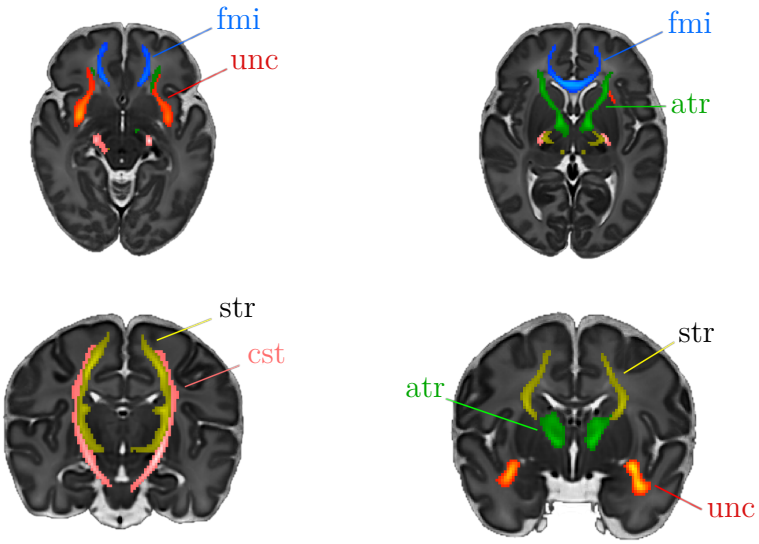


Figure 4.1: Annotated nociception-relevant white matter tracts (according to Baxter et al. [231]) overlayed on the dHCP standard T2w 40 week reference template, also known as the Schuh atlas [155, 331]. Displayed with radiological convention (left brain on the right side). Abbreviations: *atr* = anterior thalamic radiations, *cst* = corticospinal tracts, *fmi* = forceps minor, *str* = superior thalamic radiations; *unc* = uncinate fasciculi.

relaying signals through connections with key brain networks. The coupled pain and nociception-relevant cortical regions are: the somatomotor network involving sensory processing and nocifensive motor signals (connected with *str*; *cst*), limbic regions for emotional aspects of pain (connected with *atr*; *unc*), and medial and lateral frontal lobes for aspects of higher-order integration and modulation of pain signalling (connected with *fmi*) [231, 345, 346]. In the following, we test whether the nociception relationships with structure-function and brain maturation, observed in Oxford and dHCP term infants, are reflected in comparisons between development-poor preterm infants and age-matched term infants.

There have been numerous studies measuring macrostructural, microstructural, and functional development of the infant brain in isolation, and yet there has been very little work combining measures of structural, diffusion, and functional MRI data together to provide more comprehensive analyses. Whilst several infant studies have attempted to interpret diffusion and structural MRI together, or functional and structural MRI together, there are very few studies analysing a combination of

diffusion and functional MRI [205, 347]. Therefore, I chose to analyse both dMRI and fMRI in the same dataset in order to fill a gap in the literature.

4.1.2 Prematurity and Neurodevelopment

If the same trends are observed in premature infants as we initially observed in term-born infants [231], then the less mature premature-born infant brains should exhibit lesser noxious evoked responses than term-born infants matched in PMA. It would be interesting to observe whether we predict differences in pain sensitivity between premature and term-born dHCP cohorts based on RSN activity, and whether the influence of maturity on brain activity and noxious-evoked responses is generalisable to premature infants. For these functional analyses to be useful, it is vital to understand how these changes occur in the context of large-scale changes in the structural parameters of the brain. These have been widely studied in the newborn infant brain in separation, but the relationship between functional network and structural features of the brain have not been extensively characterised [347].

Early life complications, including prematurity, are associated with later life pain and developmental delay [77]. Whilst prematurity can accelerate aspects of functional development in newborns as a result of stimulus exposure, such as increased visual acuity and visual cortex activity relative to term newborns [348, 349], this is a short-term acceleration of development which is only present in infancy. However, there are sustained discrepancies in development in later life which disadvantage premature-born infants. For example, developmental delays and disabilities pertaining to cognitive and motor function are more prevalent throughout childhood and adulthood in those born prematurely compared to those born at term [350]. In this work we study the development and impact of prematurity on the structural and functional elements of the neonatal pain system, exploring the relationship of these elements to each other and to subsequently recorded Bayley Scores of Infant Development (BSID). Stimulus- and behaviour-evoked and spontaneous functional activity can be detected in the infant from the early gestation, providing an early window into the development of co-ordinated brain activity. In

infants, resting-state and stimulus-evoked activity is seen to be rapidly developing as the infant develops.

Prematurity is associated with higher risk for persistent developmental delay in later life [350–352], neurodevelopmental injury [353, 354], mental illness in adulthood [355], as well as both hypo and hypersensitivity to pain in later life [75, 77, 356]. There are several features which increase adverse outcomes in individuals born prematurely [344, 357–359]. For example, structural features including early cortical maturation at TEA have been shown to predict development at 2 years in very preterm infants [359], when including sex, gestational age, and socioeconomic status and global injury scores from structural MRI in the model. Other research has shown the significance of sex [358], gestational age [358–360] and socioeconomic status [359, 361–363] on developmental risks in preterm infants, although the latter is equally important for term-born infants [361].

4.1.3 The Brain Maturation Hypothesis

In Baxter et al., we hypothesised that observed and predicted noxious-evoked responses were correlated with metrics of brain maturity, such that infants with indicators of greater overall brain maturation had greater noxious-evoked responses. The dHCP dataset provides the opportunity to explore this hypothesis further, by investigating how predicted noxious-evoked responses differ in a premature-born set of infants compared to term-borns. I aimed to test three hypotheses in this premature and term dHCP dataset: (1) that premature infants' predicted noxious-evoked responses would be greater at TEA than during the preterm period, since the premature infants' brains will have matured significantly with increasing PMA from the early premature period to TEA; (2) that there will be a difference in predicted noxious-evoked responses between term-born and premature-born infants at TEA, owing to differences in brain maturation, since pre-term birth is associated with altered brain development driven by differences in the uterine and extrauterine environments, pre-term care and other biological and environmental factors [350, 364, 365]. (3) According to the results of our original study presented in Baxter

et al., we would also expect that differences observed in predicted noxious-evoked response between groups would be positively correlated with metrics of greater brain maturation, in particular group RSN activity and structural development of five WM tracts, identified as significant in the original study, measured with MD and FA.

4.1.4 Chapter Outline

In this Chapter, I will describe the results of applying the ML model which I presented in the prior Chapter 3 to the dHCP dataset, to test hypotheses regarding the links between predicted noxious-evoked activity and underlying functional and structural brain development. These explorations are based on background literature and initial research described in our publication [231]. For this research, I focused on characterising predicted noxious-evoked activity in the newborn in the context of functional and microstructural changes. I determined the effects of prematurity on predicted noxious-evoked responses, and how these intersect with structural developmental features, functional resting-state activity, and long-term neurodevelopmental outcomes. I aimed to gain insight into nociceptive sensitivity by characterising predictions derived from noxious-evoked responses in this additional dataset.

4.2 Methods

The ML pipeline is as described in Chapter 3. I used the NPS measure for the following analyses, as it is a validated sample-independent measure which has been shown to directly map to verbal pain ratings in adults (see Chapter 3, Section 3.2.8 for more detail). In general, prediction models trained on valid sample-independent measures are more likely to be generalisable, and I expected that a prediction model using the NPS measure was more likely to generalise to the dHCP dataset than a model using the within-Oxford SRT measure (Chapter 3, Section 3.2.7), which may over-fit to the training set and be less representative of a wider population.

The Oxford dataset and MRI recordings are as described in the previous Chapter (Section 3.2). Therefore, here, I will describe only the dHCP database, sample selection, and MRI recording details which differ from the Oxford dataset.

4.2.1 dHCP Data

The available dHCP data comprised 558 MRI scanning sessions from 505 unique infants in the second dHCP data release (`http://www.developingconnectome.org/data-release/second-data-release/release-notes/`). Of these sessions, there were 512 fMRI scans passing the dHCP fMRI quality control schema [155], and 490 dMRI scans passing the dHCP dMRI quality control schema [366]. Metadata are available for these infants, including variables such as age, sex, 18-month follow-up assessments, neonatal neurological conditions, mode of delivery at birth, and parental details such as occupation and mental illness. Not all data are recorded and complete for all infants. For example, enrolment details are marked as incomplete for 30 of the 464 unique infants with fMRI scans passing quality control.

4.2.1.1 dHCP Datasets: Sample Selection and Demographics

For the purposes of my research questions, I selected two unique cohorts of infants from the dHCP database. The first was a dataset of infants which were born and had MRI scans acquired at term, which I describe as the "Term infant" dataset. The second was a dataset of infants who were born premature and had scans acquired (a) during the premature period; (b) at term-equivalent age. This second dataset I describe as the "Premature infant" dataset, and the two sub-groups as (a) "Premature infants at prem." and (b) "Premature infants at TEA", referring to the PMA when the MRI scans were acquired. The PMA distributions of each dataset are shown in Figure 4.2a. Table 4.1 provides an overview of key demographic variables for each of the dHCP datasets and premature sub-groups, as well as the Oxford dataset used to develop the ML model for comparative ease.

	dHCP Premature Infants	dHCP Term Infants	Oxford Dataset
Number of infants	16	46	18
Gestational age (weeks)	31.9 (29.8, 34.4)	40.0 (38.5, 40.7)	38.3 (37.1, 39.1)
Postmenstrual age (weeks)	**Prem:** 33.4 (31.1, 35.2) **TEA:** 40.4 (39.9, 41.3)	40.6 (39.0, 41.3)	38.5 (37.5, 40.2)
Postnatal age (days)	**Prem:** 9.0 (7.0, 13) **TEA:** 58 (50, 75)	4.0 (3.0, 5.0)	2.0 (2.0, 2.8)
Birthweight (grams)	1620 (1254, 2126)	3335 (2639, 3630)	3400 (3008, 3769)
Sex			
Male	12 (75%)	25 (54%)	10 (56%)
Female	4 (25%)	21 (46%)	8 (44%)
Mode of delivery			
Normal vaginal	4 (25%)	12 (46%)	6 (33%)
Assisted vaginal / forceps	0 (0%)	14 (30%)	5 (28%)
Elective C-section	7 (44%)	7 (15%)	2 (11%)
Emergency C-section	9 (56%)	13 (28%)	5 (28%)
Apgar score			
1 minute	7.5 (5, 9)	9 (8, 9)	9 (8, 9.8)
5 minute	9 (8, 9.3)	10 (9, 10)	10 (10, 10)
10 minute	9 (9, 10)*	10 (10, 10)**	10 (10, 10)
Estimated number of prior painful procedures	N/A	N/A	6 (4, 13)***

Table 4.1: Demographic variables for selected dHCP datasets compared to the Oxford dataset. Values presented as median (LQR, UQR), or number (percentage) where appropriate. Asterisks indicate missing data. * marked unknown for 2 infants, ** marked as not measured (15) or unknown (13) for 28 infants, *** missing data for 8 infants. Abbreviations: LQR = lower quartile range, UQR = upper quartile range, N/A = not available, TEA = term-equivalent age.

Term Infant Dataset Sample Selection

Infants were included in the Term infant dataset if they had fMRI and dMRI scans passing dHCP quality control pipelines [155, 366], both scan sessions were completed fully, acquiring the full number of scan volumes, and the vertex of the cerebral cortex remained within field of view for at least 95% of each scan. The field of view assessments were manually completed by Luke Baxter. Where infants may have been scanned on multiple occasions, only the earliest scan was considered for inclusion. Infants were included if they had GA of $37-42$ weeks (inclusive) and PMA at scan of $37-42$ weeks (inclusive), meaning that infants were both born and scanned at term. Infants were only included if PNA at scan was $3-7$ days (inclusive). Infants scanned at 0, 1, or 2 days PNA were excluded to reduce confounding variation due to the presence of maternal hormones in neonates during these first few days of postnatal life. They were excluded if scanned beyond 7 days PNA to limit effects of postnatal life experience on the results. Finally, infants were only included if they had an 18-month follow-up assessment recorded on the database which included BSID scores. This resulted in the Term infant dataset of $n = 46$ infants, with one recording of fMRI, dMRI, and 18-month assessment data each. None of these infants had presence of neurological condition, including HIE or IVH grade ≥ 2, or sepsis recorded in the database.

Premature Infant Dataset Sample Selection

Infants were included in the Premature infant dataset if they had two MRI scanning sessions, once during the premature period (age at scan ≤ 36 weeks) and once at TEA (age at scan $37-42$ weeks inclusive). Infants were included if both their fMRI scans, at each timepoint, passed the dHCP quality control pipeline [155] and each fMRI scan was completed fully, acquiring the full number of scan volumes. Infants' dMRI scans were separately assessed for whether dMRI scans passed the dHCP quality control pipeline [366] and were completed with the full number of scan volumes, in order to maximise the number of infants included in this smaller starting pool of

premature infants. The age criteria for premature infants was GA ≤ 35 weeks with their first scan at ≤ 36 weeks PMA, in order to differentiate this cohort sufficiently from the Term infant cohort. PNA of the premature infants' first scan was also restricted to < 21 days to limit ex-utero life experience and remove outliers. This is a much broader restriction than the Term infants since more premature infants need longer between birth and scanning to ensure the infant is in a stable and healthy state to be scanned, and to avoid risk. The lowest PNA in the premature cohort was already 4 days, without placing any lower limit. There were 4 sets of twins fitting these criteria, twins are arbitrarily labelled as "A" or "B" in the database, and we removed all "B" infants in the twin pairs. Infants were only included if they had an 18-month follow-up assessment recorded on the database which included BSID scores. This resulted in a Premature infant dataset of $n = 16$ infants with two scanning sessions each, adequate quality-controlled fMRI data at both sessions, and one set of 18-month follow-up assessment data each. For dMRI analyses, this dataset reduces to $n = 13$ infants who also have adequate quality-controlled dMRI data for both scanning sessions. None of these infants had presence of neurological condition, including HIE or IVH grade ≥ 2, or sepsis recorded in the database.

Dataset Comparisons

Before starting any analysis, I checked that the Premature and Term infant datasets were well age-matched on PMA (Figure 4.2a). The histograms in Figure 4.2 demonstrate that the selected dHCP datasets of Term infants and Premature infants are well age-matched on PMA for the second scan of Premature infants at TEA. I also checked that the PMA distributions are comparable for the Oxford and Term infant datasets, and Oxford and Prem at TEA datasets. Since the dHCP Term and Prem at TEA datasets are age-matched, I only plot the former comparison for clarity (Figure 4.2b). The Oxford sample included infants between $35 - 37$ weeks, which is a lower cut-off than term age, and therefore the term-scan dHCP datasets have higher average PMA. Demonstrating that the dHCP datasets are well-matched on PMA between the Term infants and Prem infants for their second scan at TEA

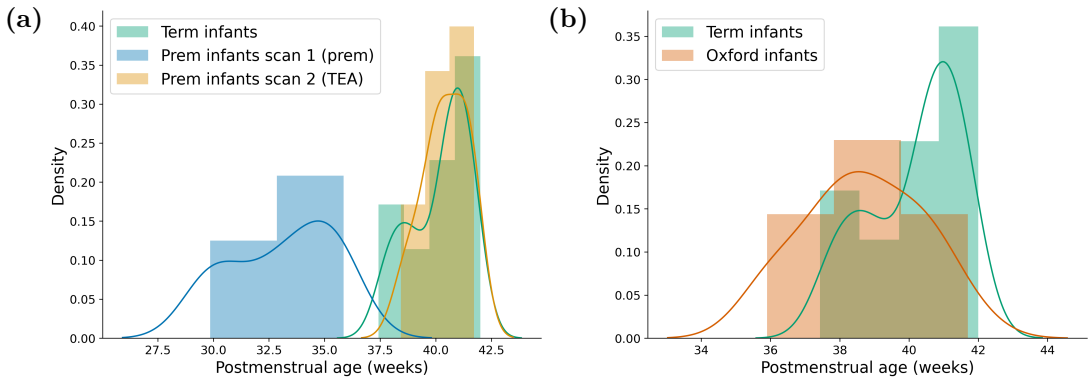


Figure 4.2: Density plot of demographics for dHCP datasets. (a) Each dHCP dataset PMA distribution demonstrating PMA-matching between Term infant and Prem infant at TEA groups. Term infants have PMA range $37.4 - 42.0$ weeks (mean $= 40.2$ weeks), Prem infants at prem have PMA range $29.9 - 35.9$ weeks (mean $= 33.1$ weeks), and Prem infants at TEA have PMA range $38.4 - 41.7$ weeks (mean $= 40.4$ weeks). (b) How the Term dHCP infant dataset's PMA distribution compares to the Oxford training dataset. The Oxford infants have PMA range $35.9 - 41.7$ weeks (mean $= 38.7$ weeks). Abbreviations: PMA = postmenstrual age, Prem = premature; TEA = term-equivalent age.

means that comparisons between these groups should reflect effects of prematurity without other age-related confounds. Table 4.1 also enables comparison of key demographic data for each dataset.

4.2.2 Neurodevelopmental Assessment

4.2.2.1 Bayley Scores of Infant Development Version III (BSID-III)

Bayley Scores of Infant Development version III (BSID-III) at a follow-up assessment at approximately 18 months corrected age were recorded for 364/464 (78%) of the unique infants with fMRI recordings passing quality control in the second dHCP data release. For the infants we planned to study, the BSID-III data availability is 46/52 (88%) of available Term infants and 16/19 (84%) of available Premature infants within our other data inclusion criteria, including quality-control and demographic constraints.

The BSID-III scores include three different composite metrics: assessments of Cognitive, Language, and Motor development [367]. There are two more scores called the "Social-Emotional" and "Adaptive" behaviour scales, but these are often not studied as they rely on the report of a parent or caregiver rather than

direct assessment by a health professional [368]. The dHCP did not measure these scores. Each of the Composite Cognitive, Language and Motor scores are intended to be standardised such that the population mean is 100 with a standard deviation of 15 [369]. Normal or delayed development is assessed by score bounds in these categories, although there are inconsistencies in the literature about which are the best cut-offs to use to determine the presence and severity of developmental delay [370–374]. For example, Lin et al. use < 70 for severe delay, $70 - 79$ for moderate delay, $80 - 89$ as borderline, and > 90 for normal development [373]. Whereas Johnson et al. report that both BSID-III cognitive and language scores < 85 provides 99% agreement of 'moderate to severe' delay compared to the typical BSID-II Mental Development Index (MDI) cut-off of < 70, but still underestimates delay by 1.1% [371]. Appropriate cut-offs may also differ for the three different assessments [374], or using a combined measure of scores may be more appropriate [371].

This makes it more difficult to analyse BSID-III data than the previous version, BSID-II, which had clear accepted cut-offs which indicated neurodevelopmental impairment (MDI < 70) [368], although some researchers suggest that it is unclear whether BSID-III overestimate infant development or BSID-II underestimated development [372]. Evidence suggests the former is true [370, 375], especially when compared to the lack of detection and prediction of impairment, clinically assessed at 4 years, in preterm infants using BSID-III assessments [376, 377], as well as BSID-III giving higher developmental scores than expected even in term infants [372]. Research repeatedly suggests that using BSID-III alone is unreliable for predicting impairment, and may lead to missed diagnoses and lack of intervention for infants at risk of developmental impairment in later life [378]. This is a problem for research, but also for medical care, where vulnerable infants may miss out on necessary resources if governments and agencies rely on absolute scores to decide who is eligible for support [370]. However, BSID-II has its own issues, including incorporating different assessments based on age, and an assessors ability to select which item sets to use as criteria based on their subjective impressions of an infant's functional abilities,

leading to inconsistencies across sites and the opportunity to under or overestimate development based on the selected test criteria [368]. The adjustments in BSID-III were designed to remove these inconsistencies and room for subjective biases in the previous version, and aimed to be more representative of a normative population by including infants with known risk conditions such as Cerebral Palsy and prematurity which were excluded from BSID-I and -II. Therefore, it could be the case that as well as BSID-III overestimating development, BSID-II underestimates development, due to the different populations included in their standardisation process, and different issues with their design and implementation [368]. A fourth edition of the Bayley assessments are currently under development, and will hopefully correct the known issues with BSID-III [368].

In my analyses of the dHCP data, I will only investigate linear correlations with BSID-III scores at 18-month follow-up, rather than attempting to classify developmental delay with cut-offs, given the inconsistency of cut-offs suggested in the literature. Additionally, given the potential bias and insensitivity in infant BSID-III measures in predicting motor and cognitive impairments in preterm cohorts [376, 377], I will only investigate trends, since absolute BSID-III metrics may overestimate development.

4.2.3 MRI Data

There are structural MRI, rs-fMRI and dMRI scans available for dHCP infants. There are no stim-fMRI scans acquired in the dHCP. In the following, I analysed fMRI and dMRI data. MRI data were collected on a Philips Achieva 3T scanner with a neonatal sized 32 channel receive coil. Detailed protocols for fMRI and dMRI image acquisition are described in their respective QC pipeline publications [155, 366].

4.2.4 Statistical Tests

Different tests were used to determine significant effects, which are described in their relevant sections. Due to the number of exploratory analyses undertaken

here, it was appropriate to adjust for multiple comparisons. Correction for multiple comparisons followed adjusted Bonferroni correction, where the effective number of tests are calculated using the correlation between measures, and the standard p value threshold for significance ($p < 0.05$) is reduced by dividing by this effective number of tests, M_e [379]. For ease of comparison I actually present calculated p values multiplied by the M_e, and continue to use the cut-off of $p < 0.05$ for these multiplied p. For clarity, I will state in the text when corrections for multiple comparisons have been used and the method followed.

For the RSN analysis, the effective number of tests was calculated from correlation between each RSN amplitude and the group mean amplitude ($n = 10$ total) for each group, and the maximum M_e was chosen for the multiple comparisons correction. The calculated M_e following Li et al. [379] were 4.61, 3.86 and 4.00 for the premature infants at prem, at term, and term infants respectively. I used the maximum $M_e = 4.6$ to be conservative, and multiplied this by 2, since there are two t-tests comparing different groups for each RSN, giving a final $M_e = 9.2$ to correct for. This means significance limits are adjusted by a factor of 9.2.

For the WM tract analysis, we similarly calculate the effective number of tests starting with individual MD and FA for each tract as well as mean MD and FA across tracts. This is a starting number of $n = 12$ tests. The calculated M_e were 3.56, 4.00 and 4.25 for the premature infants at prem, at term, and term infants respectively. We used the maximum $M_e = 4.25$ to be conservative, and multiplied this by 2, since there are two t-tests comparing different groups for each tract comparison, giving a final $M_e = 8.5$ to correct for. This means significance limits are adjusted by a factor of 8.5.

For the PCs of the RSNs, of which there are $n = 7$, we also separately calculate the effective number of tests. The calculated M_e were 4.64, 4.43 and 5.10 for the premature infants at prem, at term, and term infants respectively. We used the maximum $M_e = 5.10$ to be conservative, and multiplied this by 2, since there are

two t-tests comparing different groups for each PC, giving a final $M_e = 9.2$ to correct for. This means significance limits are adjusted by a factor of 9.2.

4.3 Results

4.3.1 Predicted Noxious-Evoked Response Magnitudes are Different Between Infant Groups

Predicted noxious-evoked responses were significantly higher in the premature infants from the scan at term-equivalent age than the same infants scanned within the premature period (paired t-test, $t = 3.6$, $p = 0.002$); see Figure 4.3. Predicted noxious-evoked responses were also significantly higher in premature infants scanned at TEA than their PMA-matched term-born counterparts (independent-samples two-sided t-test, $t = 2.2$, $p = 0.03$). According to a general interpretation of our hypothesis of brain maturation, which posits that features of brain maturation are positively correlated with noxious-evoked brain activity, given that the median noxious-evoked responses for premature infants studied at prem is the smallest, the median for term-born infants is intermediate, and median for premature infants at TEA is the greatest (Figure 4.3), then under the maturation hypothesis this would be interpreted as the premature infants at prem being the least functionally mature, the term-born infants being intermediate in functional maturity, and the premature infants at TEA being most functionally mature. Following the brain maturation hypothesis, increased noxious response in premature infants at TEA compared to the same infants scanned during the premature period is expected, due to increased maturation of their brain with increasing age. Furthermore, the greater predicted noxious-evoked responses for premature infants at TEA compared to age-matched term-borns would suggest that the premature infants have greater brain maturation at TEA than their term-born counterparts.

It is also notable that for premature infants, there is a significant trend for predicted noxious responses to increase with PMA using repeated measures correlation (R_{rm}) for the combined premature group [380, 381] ($R_{rm} = 0.71$, 95% CI: $0.55 - 0.85$,

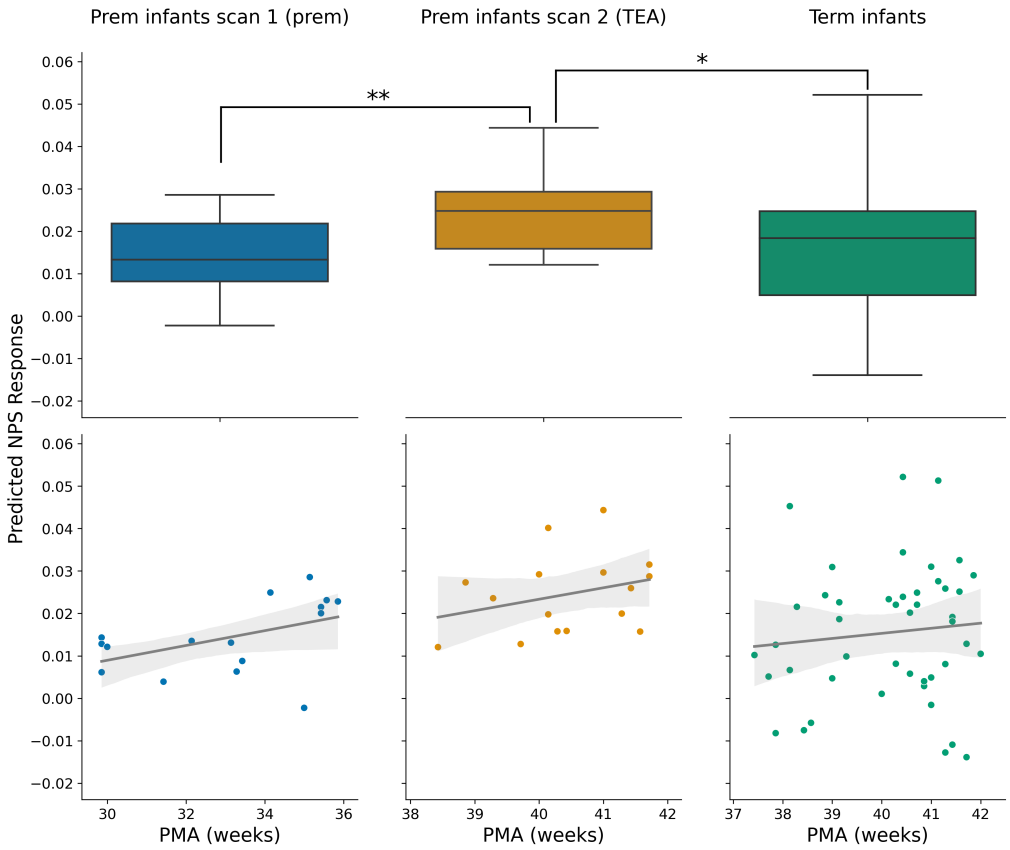


Figure 4.3: Predicted noxious-evoked responses. Top: predicted responses categorised by group, significant difference between groups are indicated by asterisks (* $p < 0.05$; ** $p \leq 0.01$). A paired t-test between Prem infants at TEA and Prem infants at prem scan gives $t = 3.6$, $p = 0.002$. An independent-samples two-sided t-test between Prem infants at TEA and Term infants gives $t = 2.2$, $p = 0.03$. Bottom: how predicted noxious responses change over PMA within each group. For premature infants, predicted noxious responses significantly increase with PMA using repeated measures correlation (R_{rm}) for the combined premature group ($R_{rm} = 0.71$, 95% CI: $0.55 - 0.85$, $p = 0.001$) whilst this trend is weaker and non-significant in the term-born infants (Pearson R $= 0.10$, 95% CI: $-0.25 - 0.48$, $p = 0.5$). Abbreviations: PMA = postmenstrual age, Prem = premature; TEA = term-equivalent age.

$p = 0.001$) whilst this trend is weaker and non-significant in the term-born sample (Pearson R $= 0.10$, 95% CI: $-0.25 - 0.48$, $p = 0.5$), as shown in Figure 4.3. Confidence intervals are calculated using bootstrap resampling, with 1000 resamples of 20 random samples without replacement from each group. Using the same bootstrap resampling parameters to calculate CIs for the difference between correlations of predicted noxious response with age in each group, these correlations are significantly different from each other (95% CI for correlation difference between groups is $0.22 - 0.98$).

4.3.2 Investigating Differences in MRI Metrics of Brain Development Underlying Differences in Predicted Noxious-Evoked Responses

The following results explore differences in measures of brain function and structure between these infant groups, how that relates to observed differences in predicted noxious-evoked responses between the infant groups, and whether these features of functional and structural brain maturation may drive observed differences in predicted noxious-evoked responses between the infant groups.

4.3.2.1 Differences in Resting State Network Amplitudes Between Infant Groups

Figure 4.4 displays comparisons of RSN amplitudes for each of the nine networks included in the ML model, as well as a group-mean RSN amplitude comparison. It is evident that trends in overall RSN activity do not match the trends in predicted noxious-evoked response, since RSN amplitudes increase from premature infants at prem., to premature infants at TEA, to term-born infants. These differences are not significant across all networks, when corrected for multiple comparisons, although this is a strong and significant effect in the group mean RSN amplitude (paired t-test for prems at TEA vs prems at prem: $t = 6.6$, $p = 8 \times 10^{-5}$; independent t-test for prems at TEA vs term-borns: $t = 3.0$, $p = 0.03$, corrected for multiple comparisons with effective number of tests $M_e = 9.2$). Some significant differences in RSN amplitudes are observed between groups in both visual and both somatomotor networks, as well as the executive control network. These trends of increasing RSN amplitudes across infant groups, from prematurity to term, follow expected trends of brain maturation, since increased functional activity is coupled with neurodevelopment which is reflected in greater RSN amplitudes. Differences in functional network activity between groups reflect the trajectory of typical brain maturation, since premature infants' brains are generally less developed than brains of healthy term infants of the same PMA. However, this does not match the pattern of differences between infant groups' noxious-evoked activity predicted from these RSNs. This indicates more complex relationships at play between RSN development and

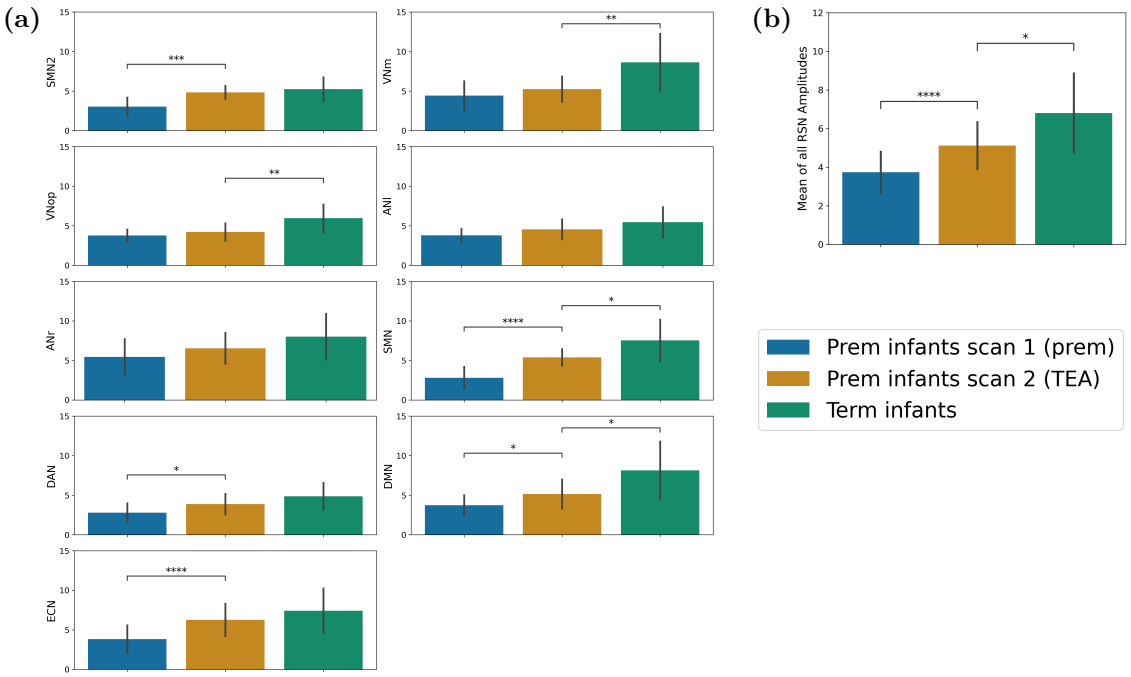


Figure 4.4: (a) RSN amplitudes for each infant group (a) for each RSN; (b) mean across all RSNs. Asterisks indicate significant difference between groups with correction for multiple comparisons using effective number of tests, $M_e = 9.2$ (* $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; **** $p \leq 0.0001$). For the group mean RSN amplitudes, a paired t-test between Prem infants at TEA and Prem infants at prem scan gives $t = 6.6$, $p = 8 \times 10^{-5}$. An independent-samples two-sided t-test between Prem infants at TEA and Term infants gives $t = 3.0$, $p = 0.03$. Abbreviations: prem = premature, RSN = resting-state network; TEA = term-equivalent age.

predicted noxious-evoked activity, which I explore in Section 4.3.2.3, by examining the Principal Components of the RSNs which form the predictors in the ML model.

4.3.2.2 Differences in Structural Brain Development Between Infant Groups

Differences in RSN amplitudes themselves did not match patterns of predicted noxious-evoked activity between infant groups, therefore I proceeded to assess the patterns of microstructural brain development to determine whether these predictions are more clearly related to underlying brain structure. The measures of microstructural development used are FA and MD in the five pain-relevant tracts identified to be significantly correlated with noxious-response in the original work by Baxter et al. [231]. In Figure 4.5 results are presented both for individual WM tracts, as well as group means for MD and FA across these five tracts. These results mirror

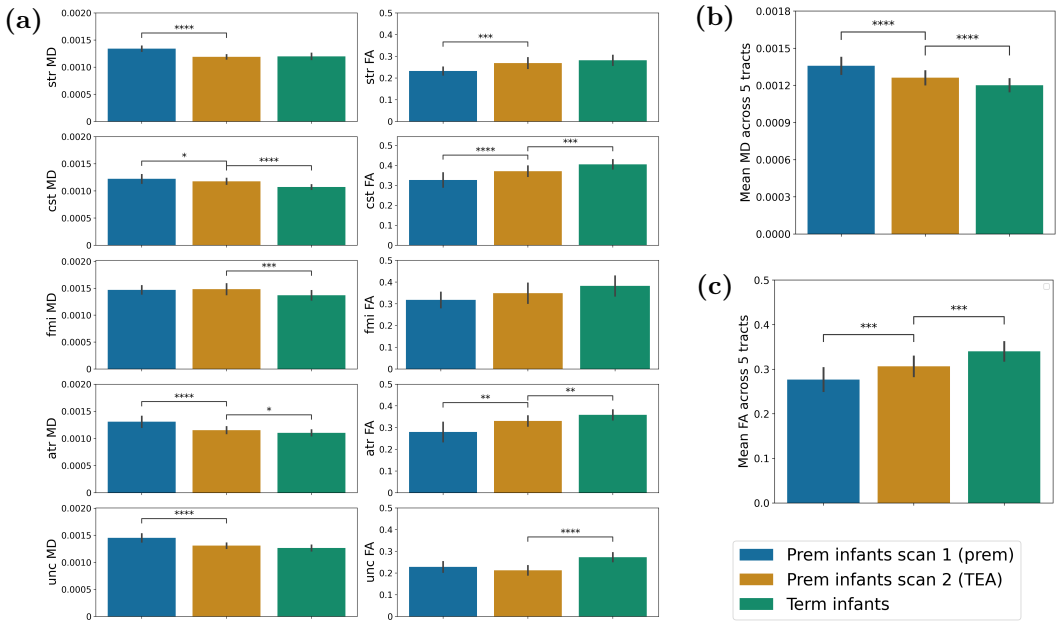


Figure 4.5: (a) Individual tract MD (left) and FA (right) for each infant group; (b) Mean MD across these five tracts for each infant group; (c) Mean FA across these five tracts for each infant group. Asterisks indicate significant difference between groups with correction for multiple comparisons using effective number of tests, $M_e = 8.5$ (* $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; **** $p \leq 0.0001$) For the group mean amplitudes, a paired t-test between Prem infants at TEA and Prem infants at prem scan for mean MD gives $t = 7.5$, $p = 6 \times 10^{-5}$, and for mean FA gives $t = 6.2$, $p = 4 \times 10^{-4}$. An independent-samples two-sided t-test between Prem infants at TEA and Term infants for mean MD gives $t = 3.8$, $p = 0.003$, and for mean FA gives $t = 5.0$, $p = 5 \times 10^{-5}$. Abbreviations: FA = fractional anisotropy, MD = mean diffusivity, prem = premature; TEA = term-equivalent age.

the observation in Section 4.3.2.1 where underlying measures of brain structure reflect expected brain maturation trajectories from prematurity to term, but do not share the pattern of predicted noxious-evoked responses across infant groups. As previously described, MD decreases with age, whilst FA increases with age, in healthy infants. In these data, in general for the individual tracts, and for the group mean, MD decreases with increasing maturity from premature infants at prem, to at TEA, to term-born infants. Similarly, FA increases for individual tracts in general, as well as for the group mean across tracts, from premature infants at prem, to at TEA, to term-born infants.

4.3.2.3 Principal Components of RSN Amplitudes Reflect Group Differences in Predicted Noxious-Evoked Response

The ML model used Principal Components of the RSN amplitudes as predictors. Therefore, I assessed these to better understand the source of predicted differences in noxious-evoked response. Indeed, the pattern of some PCs reflects the pattern of predicted noxious-evoked activity between groups (Figure 4.6). This is the case for the mean PC values for first, third, fourth and seventh PCs. As with predicted noxious-evoked responses, the mean PC values for the first, third, fourth and seventh PCs are lowest in premature infants studied in prematurity, but highest in premature infants studied at TEA, with term-born infants at term intermediate between them. After correction for multiple comparisons, only differences in PC values for the first PC are significant (Figure 4.6). Investigating the underlying causes of the patterns observed in PC differences is complex, since different RSNs contribute to each PC with different magnitude and sign (positive or negative). For example, the first PC comprises majority weights of the ANl, ECN and SMN2 networks, with all RSNs weighted with the same sign. Whilst the third PC comprises majority absolute weights of ANr, SMN2, ECN and VNop. but the ANr and SMN2 are positively weighted whilst the ECN and VNop are negatively weighted. The first PC is the only PC where all RSNs are included with the same sign, whereas all other PCs combine a mixture of positively and negatively weighted RSNs. This suggests that it is not overall RSN activity which drives differences in noxious-evoked response, but differentiation between the relative strengths of different RSNs. The development of premature infant functional networks must diverge from term-born infants such that functional activity in different RSNs does not develop in premature infants with the same trajectory as for term-born infants. The differences in premature and term-born infant PCs may be explained by relative functional acceleration or arrest of select functional networks in premature infants as a result of their increased extrauterine exposure.

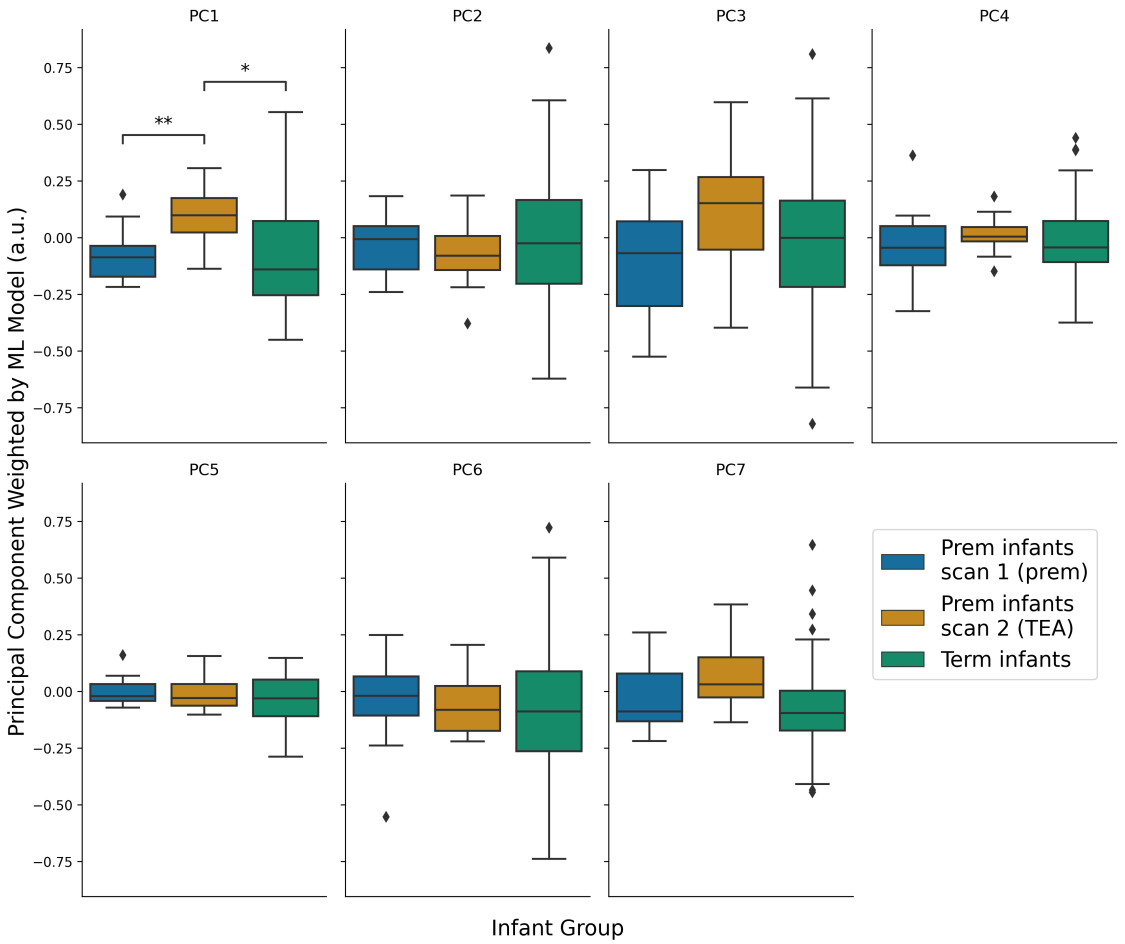


Figure 4.6: ML model-weighted Principal Components of Resting State Network amplitudes which comprise the model predictors for noxious-evoked responses. Asterisks indicate significant difference between groups with correction for multiple comparisons using effective number of tests, $M_e = 9.2$ (* $p < 0.05$, ** $p \leq 0.01$). The only significant differences between groups are values of the first principal component, PC1. A paired t-test between PC1 values for Prem infants at TEA and Prem infants at prem scan gives $t = 5.0$, $p = 0.001$. An independent-samples two-sided t-test not assuming equal variance between PC1 values for Prem infants at TEA and Term infants gives $t = 3.1$, $p = 0.03$. Abbreviations: prem = premature; TEA = term-equivalent age.

4.3.3 Results of Infant Neurodevelopmental Follow-up Assessments

In the following sections I explore how predictions of noxious-evoked response and brain function and structure near birth relate to BSID-III scores at approximately 18-month follow up. For premature infants, the follow-up is intended to be at approximately 18-months corrected age, which is calculated by subtracting the difference between the infant's GA and typical term GA (40 weeks) from the infant's

total PMA. This is in an effort to partially counter the overestimated development of premature infants assessed by BSID-III as reported in the literature [368]. Although dHCP data is stated as being at 18-month follow up, for the premature infant group the true follow-up age range is $17.6 - 20.5$ months corrected age (mean = 18.4 months corrected age), and for the term infant group the follow-up age range is $17.6 - 22.0$ months (mean = 18.8 months). Where I compare to metrics derived from MRI scans, for premature infants I use the scan taken near birth (scan 1, PNA $<$ 21 days) and compare to term infant scans, which were near birth (PNA $<$ 7 days). Therefore, I am comparing follow-up outcomes to brain function and structure measurements taken with minimal ex-utero experience in these infant groups.

4.3.3.1 Infant BSID-III Outcomes and Age at Birth

Cognitive and motor composite BSID-III are positively correlated with GA in premature infants, whilst there is no positive trend between any BSID-III scores and GA in term-born infants (Figure 4.7). This suggests that degree of prematurity affects neurodevelopmental outcomes even at 18-months corrected age, but that term-infant development is more affected by other factors than age at birth. Pearson correlation between cognitive scores and GA in premature infants is R = 0.52 (uncorrected $p = 0.04$), and between motor scores and GA in premature infants is R = 0.58 (uncorrected $p = 0.02$). Since GA is correlated with BSID-III in premature infants, and age-relationships may bias any associations with other measures such as RSN amplitudes which have known relationships with age, I regressed GA out of the BSID-III scores before testing associations between measures of brain function and structure and BSID-III outcomes in the following sections.

4.3.3.2 Infant BSID-III Outcomes and Functional Brain Activity

I explored whether neurodevelopmental outcomes at 18 month follow-up, assessed by BSID-III scores, correlate with functional brain activity near-birth (premature PNA $<$ 21 days; term PNA $<$ 7 days). I first investigated relationships between predicted noxious-evoked activity and BSID-III scores. The motivation for testing

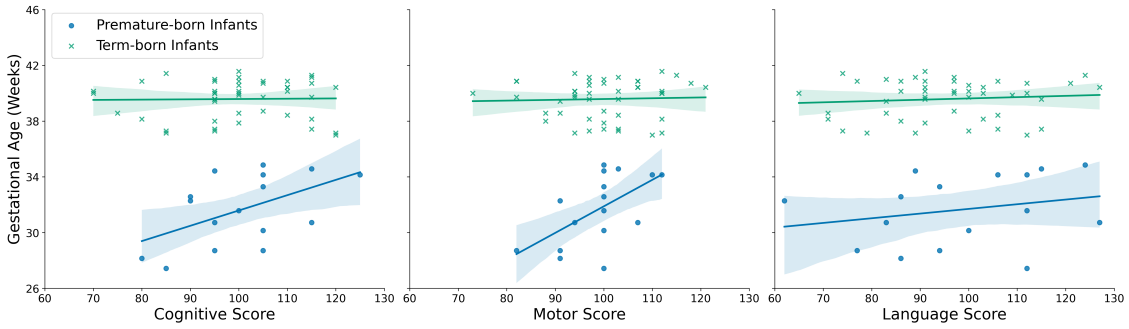


Figure 4.7: Pearson correlation between GA and BSID-III scores for premature-born and term-born infants. Pearson correlation between Cognitive scores and GA in premature infants is R = 0.52 (uncorrected $p = 0.04$), and between Motor scores and GA in premature infants is R = 0.58 (uncorrected $p = 0.02$). However, for Language scores and GA in premature infants, Pearson R = 0.24; uncorrected $p > 0.3$, and for all term infant associations, Pearson R < 0.1; uncorrected $p > 0.5$. Abbreviations: BSID-III = Bayley Scores of Infant Development version III, GA = gestational age; RSN = resting-state network.

correlation with predicted noxious-evoked activity was to see whether increased pain sensitivity (higher predicted noxious-evoked activity) might be reflected in adverse neurodevelopmental outcomes, in a similar vein to higher pain exposure in neonates leading to adverse neurodevelopmental outcomes at 8 and 18 month follow-up [55]. However, I did not find a strong association between predicted noxious responses and BSID-III outcomes in either term or premature infants (Figure 4.8a). I also investigated whether BSID-III outcomes were directly correlated with mean RSN activity near birth in these infant groups (Figure 4.8b). In this case, motor development was strongly correlated with mean RSN activity in premature infants from their premature-age scan (Pearson R = 0.76, uncorrected $p = 0.0006$), but not term infants. There were no significant relationships between mean RSN amplitudes and cognitive or language development in premature or term infants.

Only observing a relationship with motor development does not necessarily rule out interactions with cognitive or language development. Relationships with motor outcomes at 18 months are more likely to be discovered, since impaired motor development is clearer from an earlier age, whilst impairments in language and cognitive development may not be noticeable until assessments at 4-years or more [363]. These results suggest that impaired development is related to RSN

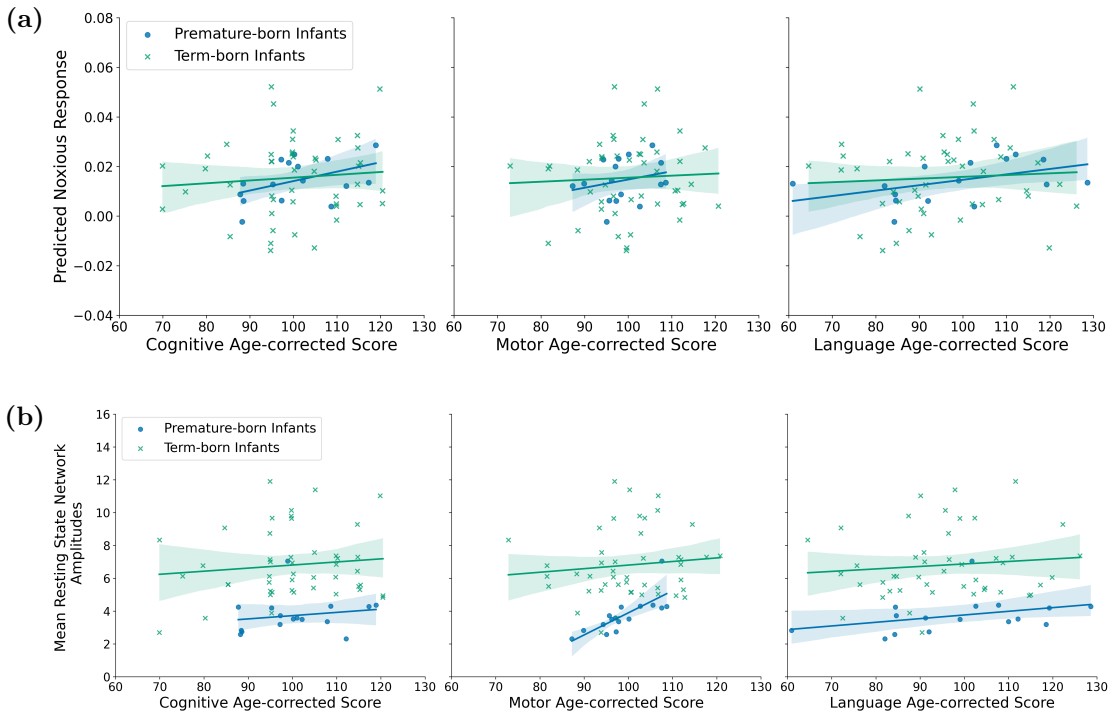


Figure 4.8: (a) Predicted Noxious Response and (b) RSNs and Bayley Scores of Infant Development in Premature and Term dHCP Infants. Pearson correlation between Motor scores and mean RSN amplitude in premature infants is R = 0.76 (uncorrected $p = 0.0006$). All other associations are non-significant between mean RSN amplitude and BSID-III scores (premature infants: Pearson R = 0.2 − 0.4, uncorrected $p = 0.2 - 0.5$; term infants: Pearson R = 0.05 − 0.09, uncorrected $p = 0.5 - 0.7$). All associations are non-significant between predicted noxious response and BSID-III scores (premature infants: Pearson R = 0.2 − 0.4, uncorrected $p = 0.08 - 0.4$; term infants: all Pearson R = 0.1, uncorrected $p = 0.4 - 0.5$). Abbreviations: BSID-III = Bayley Scores of Infant Development version III; RSN = resting-state network.

activity near birth in premature infants, after age-regression of BSID-III scores, and therefore measurements of functional brain activity may be capable of identifying at-risk infants requiring interventions to assist their development from an early age. However, our RSN-derived noxious response predictions are not correlated with development at 18 months, and do not appear to be an appropriate indicator for risk for neurodevelopmental impairment. We are missing potentially useful information which could help to explore and understand this further, such as the interaction between predicted noxious response, actual pain exposure (number of painful procedures) in early life, and neurodevelopmental outcomes. Without quantitative information on pain exposure in infancy for the dHCP cohorts, we

cannot test the effect that higher predicted pain sensitivity along with higher pain exposure, compared to equal pain exposure for a less pain-sensitive individual, may have on adverse development.

4.3.3.3 Infant BSID-III Outcomes and Structural Brain Development

I also explored whether we observe correlations between BSID-III outcomes at 18 month follow-up and the measures of structural brain development in pain-associated tracts from scans near birth (premature PNA < 21 days; term PNA < 7 days). In this case, there are no clear trends for mean MD or mean FA across the five tracts and any BSID-III scores in either premature or term infant cohorts (Figure 4.9).

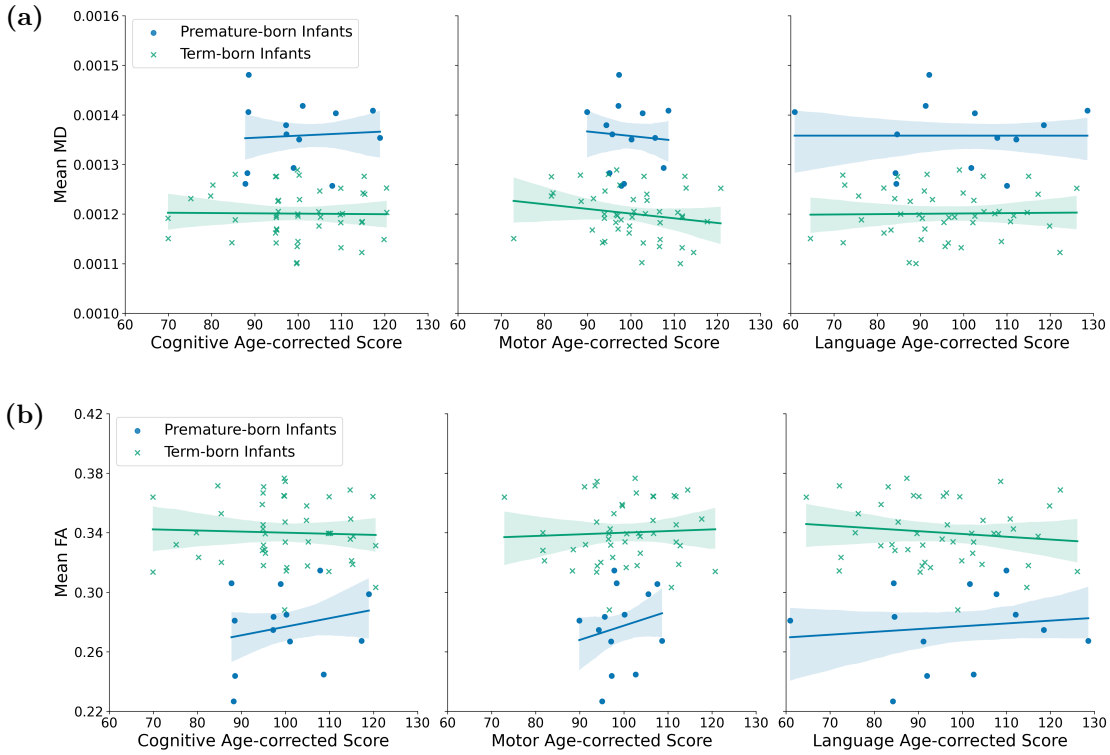


Figure 4.9: (a) Mean MD and (b) Mean FA across 5 white matter tracts and Bayley Scores of Infant Development in Premature and Term dHCP Infants. No Pearson correlations are significant for any BSID-III score associations with MD or FA (all Pearson R ≤ 0.2, uncorrected $p = 0.2-1.0$). Abbreviations: BSID-III = Bayley Scores of Infant Development version III, FA = fractional anisotropy, GA = gestational age; MD = mean diffusivity.

4.4 Discussion

In this Chapter, I found that predicted noxious-evoked response of premature infants studied in prematurity is lower than those of term-born infants at term, but when premature infants are studied at TEA their predicted noxious-evoked responses are greater than those of term-born infants at term. I also showed that predicted noxious-evoked response was positively-correlated with PMA in the premature-born infant cohort. According to the brain maturation hypothesis, these predictions would suggest that the premature infants at prem are the least functionally mature, the term-born infants are intermediate in functional maturity, and the premature infants at TEA are the most functionally mature. The analysis of MRI metrics of functional and structural brain maturity showed that the term-born infants had the greatest overall brain maturity, followed by premature-born infants at TEA with intermediate overall brain maturity, and premature-born infants during the premature period had the least functionally and structurally mature brains. Investigation of relationships between brain structure, function, and predicted noxious-evoked responses assessed near-birth and results of neurodevelopmental follow-up assessments at 18-months in these infants revealed limited associations. The only significant association was a positive correlation between RSN activity measured in premature infants during the premature period and motor development assessed at 18 months. Gestational age was found to be an important factor for neurodevelopmental outcomes in premature infants, as both motor and cognitive development at 18 months were positively associated with gestational age in the premature infant cohort. Gestational age was not significantly correlated with any neurodevelopmental outcomes in term-born infants, suggesting that it is the degree of prematurity which is important for long-term outcomes.

Premature infants' brains are generally less mature than brains of healthy term infants of the same postmenstrual age, when considering metrics of overall brain development, as reproduced in these data. However, there is evidence for functional acceleration of specific brain regions in premature infants, which could

explain the fact that premature infants at TEA have greater predicted noxious-evoked responses than age-matched term-borns despite mismatch with the opposite differences in overall brain maturity, without refuting the brain maturation hypothesis for nociceptive sensitivity. Increased noxious-evoked responses could relate to accelerated development of pain-related brain networks in premature infants by term-equivalent age, rather than depending on overall brain maturation. For example, these results could link with literature findings that greater pain exposure leads to increased pain sensitivity and functional-acceleration of brain networks with greater exposure in the extrauterine environment, such as visual and auditory stimuli, in premature infants. Greater pain exposure could also produce environmentally-induced acceleration of the functional maturation of the nociceptive system in infants. For any particular infant born prematurely, necessitating routine hospital stays and examinations, the longer the extrauterine period the more pain exposure they will have had, due to routine blood tests or other painful clinical procedures during their time in the neonatal care unit. The predictions suggest that whilst a general brain maturation hypothesis may be supported within healthy term-born samples, as demonstrated in Baxter et al. [231], this hypothesis may need to incorporate nuance for accelerated functional development in the case of premature infants. Accelerated development of the nociceptive system could be due to greater pain exposure of the premature infants, and differences in how premature infant brains develop ex-utero compared to the in-utero term-born infant brain development.

4.4.1 The Brain Maturation Hypothesis in the Context of Prematurity

Three expectations were outlined for the preterm and term infant dHCP data (Section 4.1.3), which led on from the brain maturation hypothesis proposed from results seen in Baxter et al. [231]. These expectations were (1) that premature infants' predicted noxious-evoked responses would be greater at TEA than during the preterm period, since the premature infants' brains will have matured significantly with increasing PMA from the early premature period to TEA; (2) that there

will be a difference in predicted noxious-evoked responses between term-born and premature-born infants at TEA, owing to differences in brain maturation, since pre-term birth is associated with altered brain development driven by differences in the uterine and extrauterine environments, pre-term care and other biological and environmental factors [350, 364, 365]. (3) Following results in Baxter et al., we also expected that differences observed in predicted noxious-evoked response between groups would be positively correlated with metrics of greater brain maturation, in particular the group RSN activity and the structural development of five WM tracts, identified as significant in Baxter et al., measured with MD and FA.

The third hypothesis was based on the original results which only comprised a peri-term and term cohort, and therefore the identified metrics in that cohort may be more descriptive of overall development in brain maturation which are most relevant to term born infants. However, premature infants can display specific accelerated functional maturation compared to that of age-matched term-born infants, and the nuance of functionally-accelerated differences may not have factored into the variation within the original healthy term cohort. Therefore, whilst we found that (1) and (2) held true, and (3) was not supported, this does not discount the significance of an interaction between brain maturation and infant pain sensitivity. Rather, this could suggest that the acceleration of development of specific functional networks are more important to the infant noxious-evoked response than the assessments of general brain maturation and the WM tract metrics considered in this study. Therefore, a brain maturation hypothesis should not be discounted, but should be adapted to include nuance towards potential functionally-accelerated development of specific networks involved in noxious response, rather than overall brain development. Despite the fact that the five WM tract relationships were identified in a specific noxious-evoked response study in term neonates, an additional study of more specific metrics of development of the pain network which can assess functionally-specific development in premature infants needs to be planned and conducted.

Accelerated functional brain development has been observed during prematurity with regards to stimuli which are provided with greater exposure to the infant in the extrauterine environment, such as the accelerated development of the brain's visual and auditory or language processing networks in premature infants [348, 349, 382, 383]. In particular, the development of the visual network and effects of prematurity have been studied extensively in humans and animal models. In infant EEG studies, prematurity has been shown to accelerate the development of binocular vision which develops at approximately 4 months after birth in both premature and full-term infants [349], demonstrating that onset of binocular function is dependent on PNA, rather than GA. Thus, aspects of visual network development, and binocular vision, are experience-dependent processes influenced by extrauterine sensory stimulation, rather than pre-programmed processes independent of external stimuli. Complementary EEG research studying visual evoked potentials (VEPs) [348], and Jando et al.'s own binocular vision study [349], have nevertheless shown that the myelination of the visual pathway is pre-programmed, and is dependent on total PMA rather than PNA [348, 349]. Yet, morphological changes of the VEPs in premature infants also represent accelerated development compared to age-matched infants with less extrauterine experience [348]. Therefore, functional visual development is accelerated in premature infants despite the fact that structural development to improve efficiency of signal transmission, evidenced through myelination and cell maturation of the visual pathways, is not similarly accelerated by prematurity [349]. Additionally, studies of auditory function have demonstrated accelerated development of premature infants' language processing using NIRS [383] and cortical auditory evoked potentials using EEG [382]. Nishida et al. used NIRS to compare brain activity in response to an audio recording of a fairy tale in eight extremely premature infants at TEA to ten newborn full-term infants to assess language recognition and processing [383]. They found that PNA was more important for explaining development of neurovascular coupling and auditory system development than PMA, as latency of response to the verbal stimulation was significantly shorter in preterm infants compared to the term infant cohort [383]. Similarly, Cavalcanti et al. studied peak latencies of

EEG auditory evoked potentials to speech stimuli in late preterms, at 1-3 months corrected age, compared to age-matched terms [382]. They found that peak latency was shorter in preterms at 3 months corrected age compared to terms, suggesting accelerated development of auditory cortical pathways in preterms, likely due to their increased exposure to speech stimuli in the extrauterine environment.

By comparison to increased exposure to auditory and visual stimuli in the extrauterine environment, premature infants also undergo a large number of presumed painful procedures during hospital stays over the premature period, including regular routine blood tests, intubation, and retinopathy of prematurity examinations [42]. EEG studies of noxious-evoked response also suggest that premature infants may display accelerated development of cortical noxious response [228, 297]. Research has explored the interaction between premature infants' greater exposure to pain and stress in early life, and infants' altered responses to gentle non-noxious tactile stimuli [214], and abnormal brain development [384]. A thorough and interesting study from Maitre et al. in 2017 explored how early-life exposure to pain and stress impacts infants' processing of non-painful gentle touch [214]. This EEG study has the benefit of being a large cohort study, including over 100 infants, which is rare in neonatal and infant research due to aforementioned challenges with recruitment and ethics. They measured the tactile-evoked responses of preterm and full-term infants to a gentle air-puff stimulus, and a sham control, prior to discharge from hospital. Not only did they investigate the effects of prematurity, but also the interaction of analgesics, positive supportive experiences including breastfeeding and skin-to-skin care, and painful experiences such as skin punctures and intubations, on tactile processing. They presented several key findings. Firstly, that tactile-evoked response amplitudes were significantly lower amplitude in ex-premature infants compared to full-term infants, and there were topographic differences in the tactile-evoked activity. Secondly, that supportive tactile experiences were significantly associated with greater tactile-evoked response amplitudes, when controlling for prematurity (GA and PNA) and analgesics. Thirdly, that prior nociceptive experiences demonstrated the opposite interaction, as cumulative nociceptive exposure was significantly associated

with decreased tactile-evoked response amplitudes, when controlling for prematurity (GA and PNA) and analgesics. This study shows the importance of early-life exposure to both supportive tactile and negative noxious sensory inputs, which can impact development of tactile processing. Premature infants are particularly vulnerable to the consequences of early-life painful experience, and infants in this study had a median of 32 (10 - 103) painful procedures between birth and time of EEG recording, which was measured shortly before their discharge from the NICU. The opposing effects of supportive touch and procedural pain demonstrate the complex multisensory extrauterine environment, and the necessity of recording all these variables to be able to accurately ascertain the consequences of prematurity and the impact of environmental factors on somatosensory development.

A recent holistic multisensory study by Schmidt Mellado and colleagues demonstrated accelerated development in premature infants' responses to tactile, noxious and visual stimuli, using EEG [228]. They conducted a longitudinal study of very preterm infants through to term-equivalent age, and found that EEG measures of visual and noxious-evoked response magnitudes increased whilst tactile-evoked response magnitudes decreased with PMA, reflecting developmental maturation. Tactile response magnitudes decreasing with age aligns with typical maturation in infants [228]. Additionally, greater visual and reduced tactile response magnitudes were observed in these very preterm infants compared to age-matched late preterm infants and compared to average term infants, suggesting that greater extrauterine exposure has a significant impact on accelerating development of the visual and tactile functional networks. Whilst Schmidt Mellado et al. did not find that noxious-evoked response magnitudes were significantly increased in very preterm infants compared to age-matched late preterm infants, they explain that their very preterm infant cohort were unusually healthy with a low burden of pain exposure, which perhaps explains the lack of difference observed in their study compared to other studies observing an effect of prematurity on noxious sensitivity [297]. We need to conduct additional studies of noxious-evoked brain activity which explore the effects of prior pain exposure, stress, and other factors which may influence the

maturation and development of pain networks as well as prematurity, gestational and postmenstrual age. A combination of EEG and MRI studies would be most informative. EEG is straightforward to perform cotside during clinically-required procedures such as heel lances, as well as at rest and under experimental conditions. Whilst MRI studies cannot be performed during clinical procedures and necessitates careful transport of infants off the NICU to the MRI scanner, MRI data is able to provide higher-dimensional spatial information of resting-state brain activity and experimental noxious-evoked brain activity.

There are also noteworthy MRI studies demonstrating accelerated functional development in preterm infants. One large cohort study of almost 300 term and preterm infants with MRI scans recorded at TEA investigated the effects of preterm birth on structural brain development [385]. They found that lower GA was associated with smaller brain volumes in several regions at TEA, but lower GA was also associated with several brain regions being larger at TEA [385]. In particular, lower GA was associated with larger primary motor, visual and somatosensory cortices, which, although not studied and quantified directly in the study, is likely to reflect increased cortical development due to increased visual and somatosensory input and increased physical movement in the extrauterine environment. Eyre et al. also investigated typical and abnormal development in over 300 term and preterm infants from the dHCP dataset with MRI scans at TEA [158]. They measured functional connectivity within RSNs, as a reflection of functional network maturation. Complementing the previous paper, Eyre et al. found that infants born prematurely had decreased FC across almost all RSNs, except for finding increased connectivity within the lateral motor network compared to age-matched term-borns, an imbalance in motor connectivity which they suggest could be an explanatory factor for developmental coordination disorders which often occur in preterm children [158]. The lateral motor network concerns movement of the upper limb, hand and face [158]; regions of the body which naturally move more in the less restrictive extrauterine environment and in response to external input. Other studies have also demonstrated increases in FC in premature infants in select networks associated with increased

sensory extrauterine exposure including auditory [386], visual [386, 387], stress/pain networks [386], and primary sensorimotor brain regions [387, 388], each of which may lead to disorders and poor outcomes by causing imbalances and developing at the expense of other key brain regions such as higher-order cognitive networks [158, 388].

Just because stimulus exposure can lead to accelerated functional development in preterm infants, this does not imply a positive association between the accelerated development and longer-term health outcomes and quality of life. As explored earlier in the text, longitudinal studies have shown that overall neurodevelopmental outcomes in later life are worse throughout childhood, adolescence and adulthood for those born premature compared to at term, including poorer educational achievements, increased neurodevelopmental disabilities and psychiatric disorders [350, 370]. More specifically, accelerated functional development in premature infants may also be linked to worse outcomes in later life. Scherjon et al. showed a discrepancy between functionally-accelerated maturation of VEPs in premature infants and cognitive outcomes at 5 year follow-up [389]. They found that early-life acceleration in visual maturation, represented by lower VEP latencies at 6 months, was associated with lower IQ and poorer performance in a test of cognitive function, visual recognition, and fine motor co-ordination at 5 years old [389]. Furthermore, altered brain development in response to sensory input from cumulative stressors in intensive care units may have negative consequences [384]. Smith et al. quantified stressful exposures in hospital settings to preterm infants using the Neonatal Infant Stressor Scale which counts painful procedures and non-painful stressors such as nappy change [390], and conducted MRI scans and neurobehavioural assessments at TEA. They found that greater stress exposure is linked to decreased frontal and parietal brain width, reduced brain volume, and altered functional connectivity and diffusivity in the temporal lobes as well as abnormal motor behaviour at TEA. More recently, Duerden et al. found that greater procedural pain exposure in preterms in early life is associated with smaller somatosensory thalamus volumes in infancy and worse functional outcomes at 3 years [80]. A study by Maitre et

al. suggests that adverse experiences of noxious stimuli could lead to a functional adaptation of hyposensitivity to reduce the stress of further noxious stimuli, but which secondarily damages the sensory experience of gentle touch [214]. Although certain developmental adaptations to the extrauterine environment in infancy may be initially beneficial to the premature infant for function and survival, disruption of the typical stages and pathways of brain development may be harmful to later cognitive function.

Another example of accelerated development to the extent of pathology in premature infants is increased exposure to oxygen in the extrauterine environment, which increases the risk of the disorder of retinopathy of prematurity (ROP). ROP can be caused by this relative excess of oxygen postnatally, since the hyperoxia promotes increased, atypical, vascularisation of the retina [391], and a degree of hypoxia is important to balanced and healthy development of vascular network [392]. This disruption to balanced development of the vascular network can lead to retinal detachment and irreversible vision loss. Fortunately, ROP is straightforward to screen for and treat in its early stages, but ROP remains a leading cause of preventable blindness globally, and particularly in low and middle-income countries due to increased preterm survival rates coupled with inadequate resources and training to screen and treat ROP [393, 394]. Thus, whilst factors associated with prematurity, or increased extrauterine exposure, can drive increased vascularisation this is not fundamentally beneficial, and can be pathological.

Differences were observed between the premature and term infants' predicted noxious-evoked responses. One biological explanation for a difference between premature and term infant noxious-evoked responses stems from the co-development of neural activity and network vascularisation, and that sensory-related changes in neural activity can affect the degree of the related networks' vascularisation. Neural and vascular networks develop in parallel in the postnatal brain, and research has observed this phenomenon of mutual development whereby regional neural activity drives regional vascular growth [392]. Fluctuations in neural activity and oxygen

supplied by blood flow to areas of the brain influences angiogenesis (the formation of blood vessels) [392, 395, 396]. Studies in neonatal rodents have investigated how modulating neural activity, and therefore oxygen supply, through moderate and high-intensity stimulation affects angiogenesis postnatally [395, 396]. Lacoste et al.'s study in neonatal mice found that enhancing sensory inputs led to increased vascular density and branching, whilst multiple methods of decreasing sensory input led to reduced vascular density and branching of the primary somatosensory cortex [395]. These results suggest that enhanced sensory stimulation leads to enhanced angiogenesis in relevant brain regions. This finding is not unique, with multiple complementary studies demonstrating similar effects. For example, early studies found that rodents raised in dark environments presented with decreased vascularisation in the visual cortex [397], and others have demonstrated that spontaneous neural activity is a driver of functional maturation in both the foetal [398] and postnatal [399] period. The phenomenon of spontaneous brain activity driving neural development is well-established and includes studies of multiple animals and humans. For example, Marla Feller provided a review of correlated spontaneous activity driving development of the retina, spinal cord, and hippocampus [400]. However, Whiteus et al. found that extreme stimulation in the early postnatal period in mice impaired vascularisation of the auditory, motor and sensorimotor networks [396]. Whiteus et al. also found that the same intensity of stimulation did not affect vascularisation in adult mice, and thus the infant is particularly sensitive to excessive external stimuli and abnormal development during the postnatal period due to their greater vascular plasticity [396, 401]. These seemingly contradictory results, that enhanced stimulation can enhance angiogenesis [395] or arrest it [396] is explained through concentration-dependent effects of nitric oxide on angiogenesis [287, 396, 402]. Nitric oxide is produced in response to neural activity, and either acts as pro- or anti-angiogenic depending on the concentration of nitric oxide and other mediating factors [287]. Therefore, sensory-evoked neural activity in early life either has the potential to promote network vascularisation or to prevent a network achieving normal levels of vascularisation. This could provide one explanation for why premature infants have been observed

to display both increased and decreased pain responses compared to term-born counterparts. Since premature infants receive an average of ∼10 painful routine procedures per day during their hospital stay [42], they will receive far more early-life noxious stimulation than typical healthy term infants. This noxious stimulation could therefore promote angiogenesis of the vascular networks involved in noxious input and pain signalling, leading to relative over-development and increased sensitivity to noxious input, or, if the stimulation exceeds an unknown threshold, it could arrest normal development of these networks and lead to under-development of the networks and decreased sensitivity to noxious input. Unfortunately, the dHCP dataset does not include data on the number of painful procedures infants receive, or number of blood tests or other clinical examinations which could be used as a proxy for prior pain exposure. Therefore, an informative future study would be to collect MRI data for premature infants as well as data of their clinical examinations including blood tests, intubation, surgery and other potentially painful procedures. We could compare MRI data on structural development of networks as well as resting-state functional brain activity, and predicted noxious-evoked responses, to their prior pain exposure to determine whether we see an intensity-dependent effect between pain exposure, brain development, and predicted noxious-evoked responses. Whilst we have postulated that the differences we see between premature and term infant predicted noxious-evoked responses may be due to functionally-accelerated development of the pain network in premature infants, given their greater exposure to painful events, having the data to link prior pain exposure to noxious-evoked responses would allow us to test this interpretation. Ideally, whilst we would also have similar data for term-born infants and their prior pain exposure, term infants who have a greater number of painful procedures will also have confounding effects of ill-health to consider, and/or may not be healthy and stable enough for an MRI scan during the neonatal period.

4.4.2 Developmental Outcome Measures

With respect to links between developmental outcomes and early-life brain structure and function, we only found a positive correlation between motor development and increased RSN activity in premature infants near birth. Research suggests that it is easier to measure and predict motor outcomes from an early age, where cognitive and language deficits often become apparent later on [363]. Therefore, it is possible that we would find interactions between MRI measures from near birth and cognitive and language assessments completed at 4 or more years of age. However, cognitive and language functions involve more complex brain networks than motor functions which rely on few known regions [363]. Cognitive outcomes are also influenced more by environmental factors such as socioeconomic status, which can hinder or help early-life development [55, 363].

Lowe et al. did find a significant interaction between cerebral WM volume and neurodevelopment, assessed by results of an executive function test and BSID-III composite language scores at $18-22$ months, for very low birthweight preterms compared to term controls, however this comparison was made using MRI data also collected at $18-22$ months [403]. As Lowe et al. reported, this is a rare study in that it collects and compares MRI data in preterms and term-born infants at $18-22$ months old, where other studies have often compared outcomes to MRI data collected at TEA or otherwise during the neonatal period, as in this research.

4.4.3 Limitations and Future Work

The training and validation of our ML model only included infants of $35.3-41.4$ weeks GA and $35.9-41.7$ weeks PMA, and this may limit the generalisability of our interpretations from these results to other age groups, such as the premature cohorts included in this study. The premature sample I studied here includes infants at less than 30 weeks GA and PMA. Therefore, the extension of the ML model and its interpretations with respect to noxious-evoked activity may be a flawed over-generalisation in this younger sample of premature infants. Arichi et al. demonstrated

how the HRF, used in fMRI BOLD analyses, evolves postnatally [325]. The HRF used for these data was based on the typical term infant BOLD response [325], but as stated this may be inappropriate for the premature infants, and consequently present an issue for the interpretation of these results. Further work should tailor the analysis to use the appropriate HRF for different infant postnatal and postmenstrual ages, and incorporate premature infant measurements into the training dataset of the ML model.

Comparisons with results from the study presented in Baxter et al. which identified five nociception-relevant WM tracts may be affected by using different noxious-evoked response measures. In Baxter et al., results were derived from the within-Oxford-dataset SRT measure, whereas I chose to use the dataset-independent NPS template measure in this ML model to quantify noxious-evoked responses. However, before making this decision, I did check the correlation between noxious-evoked responses derived from both the SRT and NPS measures in both true and predicted data, and found strong agreement (Pearson correlations > 0.8). Therefore, I did not expect large differences in results arising from this decision. I chose to only use NPS predictions for exploratory comparisons in the dHCP dataset, rather than both measures, to minimise issues with multiple comparisons.

This study relied on predicted noxious-evoked responses, as a useful resource-light and ethically-straightforward way to explore and generate new hypotheses. Different patterns of functional and structural maturation between premature and term infant groups compared to predicted noxious-evoked responses suggested that relative strengths of infants' functional networks, i.e., functional acceleration, may be more important to nociceptive sensitivity than overall brain maturation. However, testing whether there is evidence for functional acceleration of networks implicated in nociceptive processing in premature infants, and structural change associated with functional acceleration, requires further direct investigation using measured noxious-evoked activity and measures of brain structure and resting-state function. Additionally, whilst MRI studies provide high spatial resolution for scientific research

to improve general understanding of infant neurodevelopment, MRI is not practical for clinical application. Therefore, translating this research into developing accurate prediction models of nociceptive sensitivity, or measured nociceptive response, using continuous EEG recording would be more clinically beneficial. EEG is already used in neonatal units for monitoring purposes, and could provide additional support for pain management.

4.5 Chapter Summary

Differences in infant inter-individual pain sensitivity are poorly understood. It is a difficult balance between under and over-treating pain since both approaches can cause harm. Mapping traits which underlie differences in pain sensitivity would be useful for appropriately tailoring pain management strategies to individual infants, since they are unable to communicate their needs directly. The ML model to predict noxious-evoked response magnitudes from rs-fMRI, from Chapter 3, was applied to infant rs-fMRI datasets from the independent large dHCP database in this Chapter. Using the dHCP data, I explored relationships between predicted noxious-evoked response and age, brain structure and function, and neurodevelopmental outcomes. Predicted noxious-evoked responses were significantly higher in premature infants at TEA than during the premature period, with predicted noxious-evoked responses also significantly positively correlated with PMA in this group. Predicted noxious-evoked responses in term-born infants at term were between those of premature infants at premature and TEA. The trend of predicted noxious-evoked responses between age groups did not match patterns of RSN amplitude nor WM maturation across groups. Principal Component Analysis of RSN amplitudes suggests that rather than general maturation of functional and structural networks, there is a more complex relationship where relative maturation of functional networks in the brain influence noxious response. I identified MR evidence for complex developmental trajectories of pain responses associated with prematurity, revealing potential neurophysiological correlates. This study relied on predicted noxious-evoked responses, as a useful resource-light and ethically-straightforward way to explore and generate hypotheses,

however these findings require further investigation using measured noxious-evoked activity for direct testing.

5

Concluding Remarks

Neuroimaging techniques are valuable in the pursuit of understanding neonatal pain and brain development. These studies require bespoke methodology for neonatal acquisition and analysis. In this thesis, I have used bespoke approaches, including age-specific templates of brain structure and functional activity. There are many challenges with neonatal neuroimaging research, including validation, reproducibility, required technical knowledge, and improving uptake of neuroimaging methods. I have addressed some of these issues. In Chapter 2 I demonstrated the reproducibility and generalisability of an EEG template of infant noxious-evoked activity, which worked towards establishing this tool as a validated pain measure. I have used modern best-practice throughout, including pre-registration to avoid bias and increase transparency. Pre-registration involves open and detailed sharing of methods and analysis which improves the reproducibility of the research. Whilst the following chapters appear later in the thesis, chronologically they pre-dated any practical exploration into pre-registered reports, hence my MRI research did not follow this same framework. In Chapter 3 I developed a machine-learning pipeline incorporating best practices in machine-learning methods, to reduce bias in the model and maximise generalisability. This model used data from resting-state and noxious stimulus-evoked functional MRI, representing cutting-edge methods in infant pain research.

In Chapter 4 I applied this validated machine-learning model to predict noxious-evoked activity from resting-state brain activity in infants from the open dHCP dataset, and to explore relationships between infants' predicted noxious-evoked responses and their demographics and brain maturation.

In this thesis I discussed issues relating to reproducibility in research, with a particular lack of reproducibility studies in infant neuroimaging research. Infant studies are usually small sample, due to ethical considerations and difficulties recruiting and studying the infant population. These shortfalls are intrinsically linked, since results based on small sample sizes would benefit from reproducibility research, and yet reproducibility studies are not prioritised in resource-poor settings where recruitment is difficult. The institutes which enable academic careers and success do not currently value and support reproducibility studies enough, nor open science practices which would ultimately lead to greater understanding and confidence in our results. Statistical confidence in the accuracy of our results, and the contexts in which they do and do not apply, is particularly important when decisions based on research evidence can impact the care of vulnerable populations, such as infants. Fortunately, there is a growing open science movement and the research landscape is beginning to change, with credit owed to grass-roots initiatives from passionate researchers to improve the culture of open research and publishing. Frameworks and systems are now in place to support open science including peer-reviewed and un-reviewed pre-registration, and data sharing platforms (e.g., OSF `https://osf.io/`), and journals are increasingly offering Registered Report options (e.g., `thejointone`). However, further work is needed to ensure these frameworks are well-implemented, understood and valued, by potential authors, peer-reviewers, and readers, as well as the systems which support academic careers, including conference organisers and funders. A set of recommendations to the reader, whether they are a new researcer in infant pain or neuroimaging fields, or have power to influence publishing and funding institutions, are outlined in Table 5.1.

To	Recommendation	Notes
The Researcher	**Open Science, Transparency and Reproducibility:**	Following open, transparent and reproducible research practices, including the below
	Share Data	Sharing data is important, particularly in fields where we need to be particularly mindful of ethical research, and where data collection is difficult and sample sizes are limited. Sharing data ensures we maximise the appropriate use of existing data to further knowledge, without wasting our limited resources or causing undue harm. It also allows us to direct our resources towards worthwhile research. For example, combining and utilising larger datasets allows exploratory analysis which can help to identify hypotheses to be tested in prospective datasets. Projects to produce open neuroimaging dataset such as dHCP have enabled a wealth and diversity of research which would never have been possible through individual researchers collecting smaller isolated datasets independently.
	Share Methods	To truly advance research and understanding, others must be able to follow and repeat your methods, whether they intend to reuse them, or to test their generalisability or reproducibility in new contexts.

Table continued on next page

To	Recommendation	Notes
	Share Research (Early and Agnostically)	Peer-reviewed pre-registration and preprints to support well-designed research, promote and allow publication of valuable research regardless of outcomes. Pre-registration may not be appropriate in every case, but recording carefully designed analysis plans which meet statistical requirements will prevent potential bias, or other methodological mistakes, and can encourage sharing of outcomes whether "positive" or "negative". Research which yields null or negative results may still be valuable; it can prevent others spending valuable time on known dead-ends, and highlight cases where prior hypotheses do not generalise. Therefore, this research should also be shared.
	Open Access	Research should be widely accessible. Open access promotes transparency and sharing of knowledge between researchers, fosters wider collaboration and impact.
	Ethics and Integrity	In the pursuit of research, it is important not to forget ethical research practices and research integrity. This category includes considering and avoiding potential biases, such as statistical errors in p-hacking, or presenting exploratory research as a-priori hypotheses.
Institutes, Funders, or Journals	**Value and Support Open Science Practices**	Institutions need to value and reward open science to make the movement sustainable, and to improve the landscape and culture of research and academia long term.

Table continued on next page

To	Recommendation	Notes
	Support True Open Access Research	Research should be widely accessible to promote transparency, foster collaboration, impact and innovation. However, Open Access publishing needs to be supported by institutions and journals, and valued by institutions and funders. Open Access itself should be accessible and not come at an unachievable cost to those from less wealthy institutions.
	Publish and Value Negative/Null Findings	Journals should publish research which fails to replicate results, shows null results, or shows negative findings. Current publishing and funding-value systems which reward positive findings too much lead to publication bias including the file-drawer problem, ultimately encourage resource-wastage as people repeat methods others have failed before, and encourages methodological biases which skew the research landscape such as HARKing and p-hacking. Many of these issues would be reduced by fairer publishing and value of high quality research regardless of the direction of outcomes.
	Fund and Platform Reproducibility Research	Similar to the above, it is one thing to encourage researchers to conduct reproducibility research, but institutions, journals, and funders need to support this research in order to sustain the reproducibility movement. The benefits of reproducibility research should be understood and valued by all, and these researchers should be able to be platformed in journals and conferences, for appropriate knowledge-transfer.

Table continued on next page

To	Recommendation	Notes
	Provide and Appreciate Time and Resources to Follow Best Research Practices	Researchers should prioritise methodological and statistical best practices, and such rigorous and robust research should receive credit. Research content and quality should be valued together. This should include value afforded to sharing methods, data, and materials that can be reused by others.
Everyone	**External Resources**	There are many researchers and organisations now promoting and enabling open science, with their own guidelines and initiatives. I would encourage those interested to engage with these organisations and utilise their well-developed resources. These include the Centre for Open Science (COS, https://cos.io/) which produced the OSF (https://osf.io/) for transparently hosting research and associated materials, and TOP Factor (https://topfactor.org/) to review journals' open science practices.

Table 5.1: Recommendations to researchers and to Institutes which shape the research landscape, which relate to open science principles and upholding research integrity, transparency and reproducibility.

Developing methods is fruitless unless they can be used to answer research questions and advance scientific knowledge. One of my interests is the clinical translation of research, including translation to clinical trials. The EEG noxious-evoked template is an objective measure of infant noxious response, which provides an excellent opportunity to use this tool as a primary outcome measure to improve the quality of clinical trials for pain relief. By developing a machine-learning prediction model, I was able to integrate a small locally-acquired dataset with a large open neuroimaging dataset; the dHCP. This feeds into an exciting and thriving new area of science, seen in many fields, e.g., utilisation of the UK Biobank data. In this analysis, I was able to demonstrate that pain sensitivity is associated with resting-state brain activity in infants. I was therefore able to explore factors of brain

development underlying pain sensitivity and identify potential future avenues of research. MRI data is very difficult to acquire in neonates due to challenges around recruitment, acquisition, and resources. By utilising the large dHCP dataset for exploratory analysis, this means that precise research questions arising from this analysis can be tested in smaller-scale neonatal studies without resource wastage.

5.1 Limitations and Future Work

The EEG noxious-evoked template affords the opportunity to objectively assess pain response in infants, and provide objective outcome measures for clinical trials. There are two important directions to develop these measurement tools. Firstly, the template reproduced in Chapter 2 is only validated in infants aged 34 weeks or more. A similar reproducibility study for the recent preterm-specific ($30-33$ week) noxious-evoked EEG activity template would be beneficial. Premature infants are often the most vulnerable, where the most premature infants are exposed to the greatest number of painful procedures, and therefore validating appropriate objective measures of pain in the most preterm infants is of utmost importance. Validation of this preterm-specific EEG noxious-evoked activity template could be achieved by testing for reproducibility in an independent prospective cohort study, and establishing responsiveness to pain-relieving interventions in independent samples. Secondly, recent evidence suggests that multimodal measurements of infant pain including cortical activity, behaviour, and physiological signs, are the most sensitive to discriminating noxious from non-noxious responses, particularly in the most preterm infants. EEG templates of noxious-evoked activity in each age group should therefore be combined with physiological and behavioural measures of noxious response to provide the best holistic assessment tool for clinical trial use.

Another future area of interest is developing these methods to provide an ongoing assessment of pain in NICU patients. This is particularly relevant to the management of prolonged pain in infants following painful procedures, or due to invasive equipment such as intubation and IV. These current methods have been designed for the

assessment and management of acute pain, but it would be very interesting and clinically valuable to build on these methods to detect pain, through continuous monitoring outside of an acute procedure. This is a complex task which will require the collection of continuous EEG and vital signs data in infants, and developing accurate automated detection of pain events from these data. However, achieving this task could revolutionise infant pain treatment and immeasurably improve quality of life.

Machine-learning models could be used to derive clinical biomarkers of nociceptive sensitivity to assist infant pain management. MRI can provide useful scientific insights and uncover underlying neurobiological relationships to infant nociceptive sensitivity. However, this knowledge needs to be converted to more practically implementable systems for clinical applications, including EEG.

Appendices

A

Chapter 2 Supplementary Material

A.1 Documentation for EEG Studies

Consent Form

Study ID: | | | | | | | | | |

Infant's name:

Study Title: Investigating pain in the developing human brain

Chief Investigator: Prof Rebeccah Slater [redacted]

Please initial each box

Please complete in black ballpoint pen.

1	I confirm that I have read and understood the information sheet (clinical procedures) (**v. , dated / /**), for the above study. I have had the opportunity to ask questions and have had these answered satisfactorily.	☐
2	I understand that my child's participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my child's medical care or legal rights being affected.	☐
3	I understand that relevant sections of my child's medical notes and data collected during the study may be looked at by individuals from the University of Oxford or Oxford University Hospitals NHS Trust, where it is relevant to my child's taking part in this research. I give permission for these individuals to access to my child's records.	☐
4	I agree to my child being videoed during the study. I understand that recorded images will not be used for public use, only analysis. No identifiable information, including video recordings or imaging, will be used in any publications/presentations. Only anonymised data will be published or presented at meetings.	☐
5	I agree for the collected data to be used for teaching or academic research presentations.	☐
6	I agree to my child taking part in the above study.	☐
	OPTIONAL	
7	I agree to my child being studied on more than one occasion, up to a maximum of 5 occasions.	☐
8	I agree to complete a parental questionnaire related to my child's study.	☐
9	I consent to being approached in the future about other research studies that my child may be eligible for.	☐
10	I agree to the images/videos of my child recorded during this study being used for publications and presentations.	☐

Name of parent: _____

Relationship to baby: _____

Signature: _____

Date: _____

Name of investigator taking consent: _____

Signature: _____

Date: _____

1 to be kept as part of the study documentation (original)
1 form for parent
1 to be kept with hospital notes

NIPI Consent Form (Oxford): clinical procedures v13.0 08/11/2021
REC Ref: 12/SC/0447 Tel: [redacted] ***Newborn Care Unit, John Radcliffe,*** *Headley Way, Headington, Oxford, OX3 9DU*

Figure A.1: Sample consent form for recruitment to infant EEG studies. Personal contact information has been redacted.

Investigating pain in the developing human brain

Parent Information Leaflet

John Radcliffe Hospital

DEPARTMENT OF PAEDIATRICS

UNIVERSITY OF OXFORD

Oxford University Hospitals NHS
NHS Foundation Trust

Your child is, or may be, eligible to take part in a research study. Before you decide, it is important that you understand why the research is being done and what it involves. Please read the following information carefully and ask us if anything is unclear or if you would like more information.

1. **Study title:** Investigating pain in the developing human brain

2. What is the purpose of the study?
Infants in hospital often need to have many procedures as part of their routine medical treatment, which may cause discomfort. As they cannot tell us how much these procedures hurt, it is difficult to know how much pain they are feeling and to make sure that they receive the right medicines. We know that infants can process discomfort and pain in their brains and we have developed a method of assessing this pain-related brain activity. We also know that infants show that they are in pain using different behaviours. These may be indicated by changes in heart rate and breathing rate in response to pain.

The aim of this research is to understand more about how infants experience pain, so that better ways of treating pain can be developed. We are also interested in how infants respond to different stimuli from their environment, such as light and sound, and how this might change across development.

3. Does my child have to take part?
No, it is your decision whether or not your child takes part. If you decide to allow your child to take part, you will be asked to sign a consent form. If you decide you do not want your child to take part, this will not affect your child's care.

If you decide you would like your child to take part, you can change your mind at any time and withdraw your child from the study by telling the research team. You do not have to give a reason. You will be asked if we can use the data/images that have already been collected for analysis (all data) and for publication of results (anonymised data only).

4. What is involved in the study?
In this study we would like to understand how infants respond to clinically-required procedures, such as blood tests. **No clinical procedures will be carried out solely for research purposes.** We will only study your child during a procedure that is needed for clinical purposes.

As we are interested in how infants respond to different stimuli in their environment, we may also ask to study your baby during a control procedure and in response to sharp touch. These do not pierce the skin, but they stimulate the receptors that we are interested in generally without waking or upsetting the infant.

We will assess your child's response by measuring their activity. We may also video your child's face, and measure other responses such as muscle activity, heart rate and oxygen saturations. We will monitor your child before, during and after the clinical procedure. On rare occasions we may ask to monitor your child for up to 36 hours before and/or after the clinical procedure.

Clinical procedures will be completed in the routine way. The study will not interfere with your child's clinical care, nor will there be any delay if an emergency procedure is required. We may also explore the impact of pain relief and comfort measures on your child and may ask you to complete a questionnaire following the study.

As we are interested in how your child's response to pain changes as they grow, we may ask if we can study your child more than once during their stay in hospital. We will also ask you if we can contact you in the future, to ask if you would be happy for your child to take part in other research studies. If you agree that we can contact you in the future about other research studies, we will also record your contact details. Your contact details will not be passed onto anyone outside of the research team. You can opt-out of this at any point by contacting Prof Rebeccah Slater (details below). Your agreement for us to contact you does not form any obligation to participate in future research.

We may use the following recording measures for your child:

Measuring brain activity

Electroencephalography (EEG)**:** EEG is a portable imaging system to measure brain activity. It involves gently placing electrodes (small metal discs) on the head using a paste that can be washed off with soap and water. EEG is routinely used on the neonatal unit, children's wards and clinics.

Near Infrared Spectroscopy (NIRS): NIRS is a non-invasive technique to measure brain activity. It involves placing lights and detectors on the head to record changes in blood and tissue oxygen levels.

Ultrasound: an ultrasound machine uses sound waves to create images of the brain. Ultrasound is routinely used to monitor babies' development during pregnancy and to assess brain development on the neonatal unit. In our research we also use a special type of ultrasound called functional ultrasound. This can measure which areas of the brain are active. An ultrasound scan involves placing an ultrasound probe on your child's head. To make contact, some gel will be applied between the head and the probe.

Measuring other responses

Electromyography (EMG): EMG is a safe non-invasive technique to record muscle activity. Small electrodes will be placed on the skin over the muscle to see if your child pulls away during the stimulation (and clinical procedure if relevant).

Vital sign monitoring: Small adhesive electrodes may be placed on your child's chest to measure changes in heart rate (this is called an ECG) and breathing rate. A small probe may also be wrapped around your child's foot to measure changes in blood oxygen levels.

Videoing your child: We may also video your child during the study. This is so that we can assess changes in facial expression and body movements, and to record the exact timing of the stimulation or clinical procedure.

We may also approach you to ask if you are happy for us to use these images for teaching, publicity and/or scientific journals. If you agree, we will take separate consent for this as your child's face would be visible in the video footage. This is not a mandatory part of the study. If you choose not to allow us to use the images in this way, this will not affect your child's care or prevent your child from participating in this research.

5. Are there any additional risks or benefits for my child?

Obtaining video footage of your child is non-invasive and does not present any risk to your child. EEG, EMG and ECG have been used clinically for over 20 years without any adverse effects. Ultrasound is a tool that is routinely used in clinical practice. All studies have a dedicated team of healthcare professionals and researchers that will ensure the safety of your child at all times. We are not aware of any risks for your child taking part in this study.

The data collected are for research, so will not be reviewed by a doctor routinely. If any clinically significant findings are identified at the time of the study then the research team will report these to the clinical care team to handle as appropriate.

There are no direct benefits of participating in this research. This study is designed to gather information, to help guide improvements in care for infants in the future. If your child becomes distressed, the research study will be paused or stopped. Any clinically required procedures will still go ahead if the treating clinician feels that this is appropriate.

6. What information will be collected about my child?
We will collect information about your child from the medical notes, including demographic (e.g. ethnicity), clinical (e.g. number of blood tests in hospital), environmental (e.g. ward transfers) and social factors (e.g. postcode). This information helps us to determine which factors may influence the way an infant copes with pain. We will also collect information about your child's brain and its activity, and may collect information about your child's muscle activity, vital signs (such as heart rate and breathing rate), and recordings of their facial expressions and body movements.

All information and videos that are collected during this research study will be kept strictly confidential. Each infant will be allocated a study number which will be used to label all data. This study has been registered with the data protection registration office and forms part of an educational programme.

7. What will happen to the results?
Results will be analysed and published in a journal. All publications will be made available on our website **https://neuroimaging.paediatrics.ox.ac.uk**. The findings may also be used for teaching or academic research presentations. No identifying information will be presented about you or your child, unless you have provided specific consent for us to use videos/images of your child in this way.

8. What will happen to my child's data?
We will be using information collected from your child and their medical records in order to conduct this study. Research is a task that we perform in the public interest. The University of Oxford, based in the United Kingdom as Sponsor, is the data controller. This means that we, as University of Oxford researchers, are responsible for looking after the information collected and using it properly. We will use the minimum personally-identifiable information possible. We will keep identifiable information about your child for up to 5 years after the study has finished. This excludes any research documents with personal information, such as consent forms, which will be held securely at the University of Oxford for 25 years after the end of the study.

UK Data protection regulation provides you with control over your personal data and how it is used. When you agree to your information being used in research, however, some of those rights may be limited in order for the research to be reliable and accurate. Further information about your rights with respect to your personal data is available at
https://compliance.web.ox.ac.uk/individual-rights
You can find out more about how we use your information from the contacts in section 12.

Research data may be shared with other researchers doing similar work, both here and abroad. Responsible members of the University of Oxford or the Oxford University Hospitals NHS Trust may be given access to data for monitoring and/or audit of the study to ensure we are complying with regulations.

9. Who is organising and funding this research?
This study is sponsored by University of Oxford and has been funded by The Wellcome Trust. Your doctor will not be paid for including you in this study.

10. Who has reviewed the study?

All research that involves NHS patients has to be approved by a Research Ethics Committee. Approval means that the Committee is satisfied that yours and your child's rights will be respected, that any risks have been reduced to a minimum and balanced against possible benefits, and that you have been given sufficient information on which to make an informed decision about whether to take part. The South Central Oxford C Research Ethics Committee has reviewed and approved this study.

11. Comments or concerns during the study

The University has arrangements in place to provide for harm arising from participation in the study for which the University is the Research Sponsor. NHS indemnity operates in respect of the clinical treatment with which your child is provided. If you wish to complain about any aspect of the way in which you have been approached or treated during the course of this study, you should contact Prof Rebeccah Slater (details below) or the University of Research Governance, Ethics & Assurance (RGEA) office [redacted]

12. Contact for further information

[redacted]

Picture shows example of an EEG study.

Thank you for reading this information leaflet.

Figure A.2: Sample patient information leaflet for infant EEG research. Personal contact information has been redacted.

SCOPE

This SOP describes how to withdraw a participant from a research study. A participant can be withdrawn at any time (before, during or after data collection) by either a parent or a member of the research/clinical team. Parents do not need to give a reason for the withdrawal.

PROCEDURE

NIPI studies

1. Fill in the paper NIPI Withdrawal form (any staff member can use this form). Blank forms are stored in the NIPI document box in Office 4, or electronic copies are on the shared drive.
2. The form will ask if a reason for withdrawal has been provided, and if the parents have agreed that data already collected can still be used.
3. It is important to reiterate to parents that clinical care will not be affected by the decision to withdraw. All clinical care should continue as planned.
4. The completed paper form should be kept in the NIPI filing cabinet with the participant's Consent Form etc.
5. **The time and date of the withdrawal should be documented in the medical notes, together with any other necessary information.**
6. The information from the Withdrawal Form should be entered onto the NIPI database.

WITHDRAWAL FORM

nipi

Section A: Baby Details

A1. Study number:

A2: Infant's date of birth:

Section B: Withdrawal Details

B1: When was the infant withdrawn? Date / Time

B2: Status at time of withdrawal: Before study conducted / During study / After study complete

B3: Decision to withdraw made by: Parent/Guardian / Research team / Clinical team / Other

B4: Was a reason for withdrawal given? Yes / No

If YES to B4:

Reason for withdrawal:

B5: If withdrawn during study or after the study was complete, have the parents agreed that any data already collected can be used?

Yes / No

Section C: Form Completion

C1: Name: (Please print)

C2: Signature:

C3: Role:

C4: Date form completed:

Non-NIPI studies

If a Withdrawal Form for your non-NIPI study exists, use this form to record any withdrawals. If there is no Withdrawal Form for your non-NIPI study, create a file note to store with the participant's Consent Form and any other data collection forms, and record the following information:

- Participant ID
- Study withdrawn from
- Date and time of withdrawal
- Participant status at time of withdrawal (prior to study, during study, after study)
- Who made the decision to withdraw (and a reason, if given)
- If the withdrawal took place during or after study, have the parents given permission for the data already collected to still be used.

The withdrawal should also be documented in the medical notes as for NIPI studies.

CHANGE HISTORY

SOP and Version No	Effective Date	Significant Changes	Previous SOP and Version No
SOP 16	23.01.2019	First version	N/A

Figure A.3: Standard Operating Procedure for study withdrawal. Title page omitted due presence of personal details.

SCOPE

This SOP describes the process for safety reporting in any research studies not involving an investigational medicinal product (i.e. any research studies which are not drug trials).

PROCEDURE

A **SERIOUS ADVERSE EVENT (SAE)** is any untoward medical occurrence that:

- Results in death
- Is life-threatening
- Requires inpatient hospitalisation or prolongation of existing hospitalisation
- Results in persistent or significant disability/incapacity
- Consists of a congenital anomaly or birth defect
- Other 'important medical events' may also be considered serious if they jeopardise the participant or require an intervention to prevent one of the above consequences.

The time period for reviewing SAEs should be considered for each individual research projects, but within the NIPI study any SAEs occurring within 12 hours of studying the baby should be recorded (at a minimum). SAEs should be recorded on an SAE Report Form as soon as the research team become aware of them. This includes SAEs which are unlikely to have any association with the research study. Any member of staff can report an SAE. Paper SAE forms are stored in the Document Box in Office 4.

Any new, updated, or corrected information on previously reported SAEs must be reported on a separate SAE form when the information becomes available, with 'Follow-up information' selected instead of 'Initial report'. The outcome of events 'Resolving' or 'Not Resolved' must be followed up until 'Resolved'. All information on any SAE Report Form (i.e. both 'initial report' and 'follow-up information') should be entered into the corresponding participant's record in the database. This is important as cumulative numbers of SAEs need to be included in the annual safety reports to the REC.

All SAEs should also be reported to the CI and Clinical Research Director. Where necessary, the REC will be informed by the Clinical Research Director. Where there is an immediate hazard to patient health and safety requiring urgent safety measures, these measures should be taken immediately and the REC and other investigators must be notified within 3 days.

Causality

The casual relationship of each SAE must be assessed on the paper SAE Report Form by a medically qualified individual who has been delegated this duty on the Delegation Log (stored in the yellow NIPI folder). The definitions for assessing causality are:

- **Unrelated** – where an event is not considered to be related to the research study
- **Possibly** – although a relationship to the research study cannot be completely ruled out, the nature of the event, the underlying disease, concomitant medication or temporal relationship make other explanations possible
- **Probably** – the temporal relationship and absence of a more likely explanation suggest the event could be related to the research study
- **Definitely** – the research study is the most likely cause.

CHANGE HISTORY

SOP and Version No	Effective Date	Significant Changes	Previous SOP and Version No

Figure A.4: Standard Operating Procedure for safety reporting during infant neuromimaging research. Title page omitted due presence of personal details.

A.2 Flow Diagrams for Sample Selection used in Pilot Datasets for Registered Report Pilot Analysis

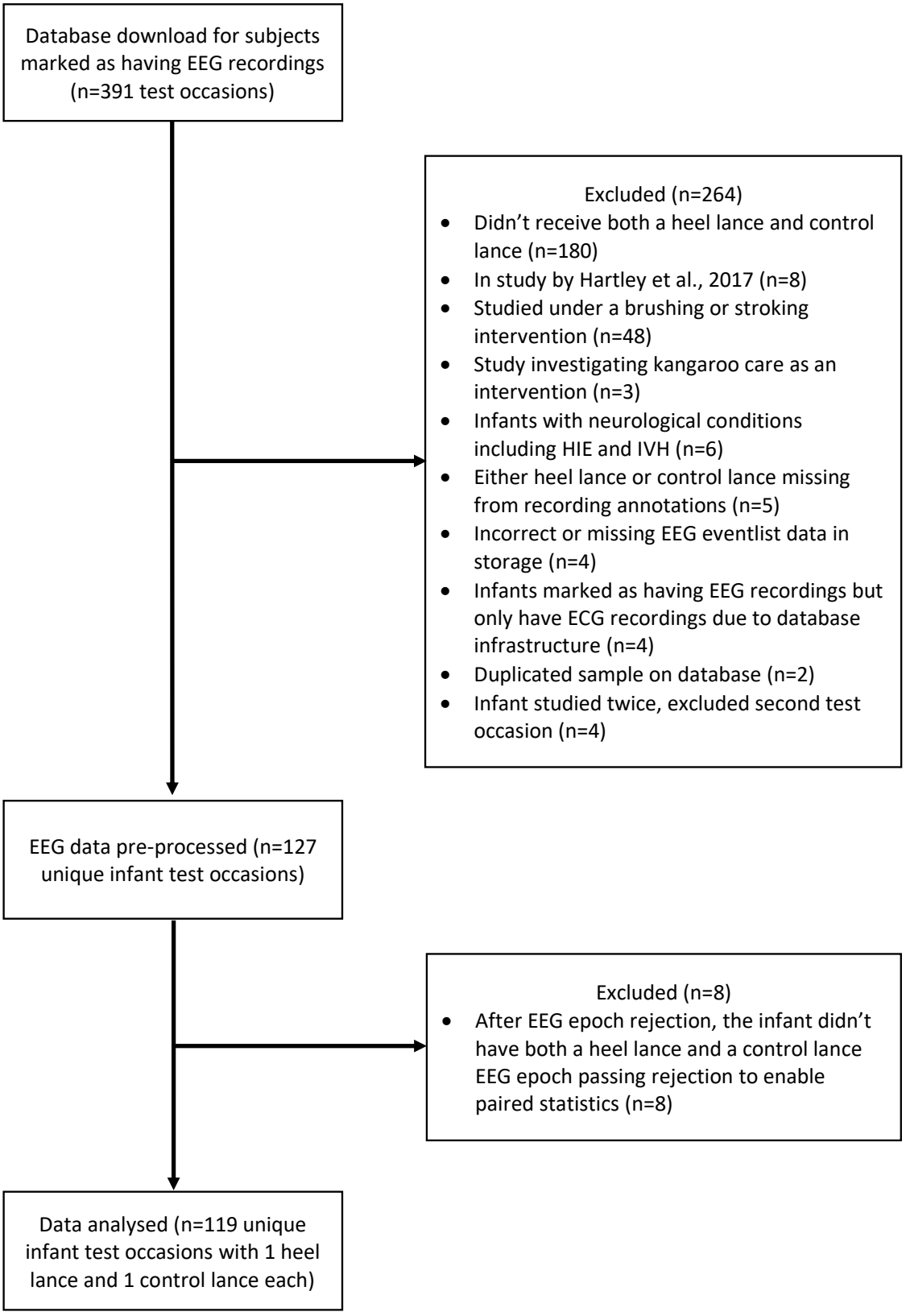


Figure A.5: Flow diagram illustrating the exclusion decisions for Oxford Pilot Dataset A

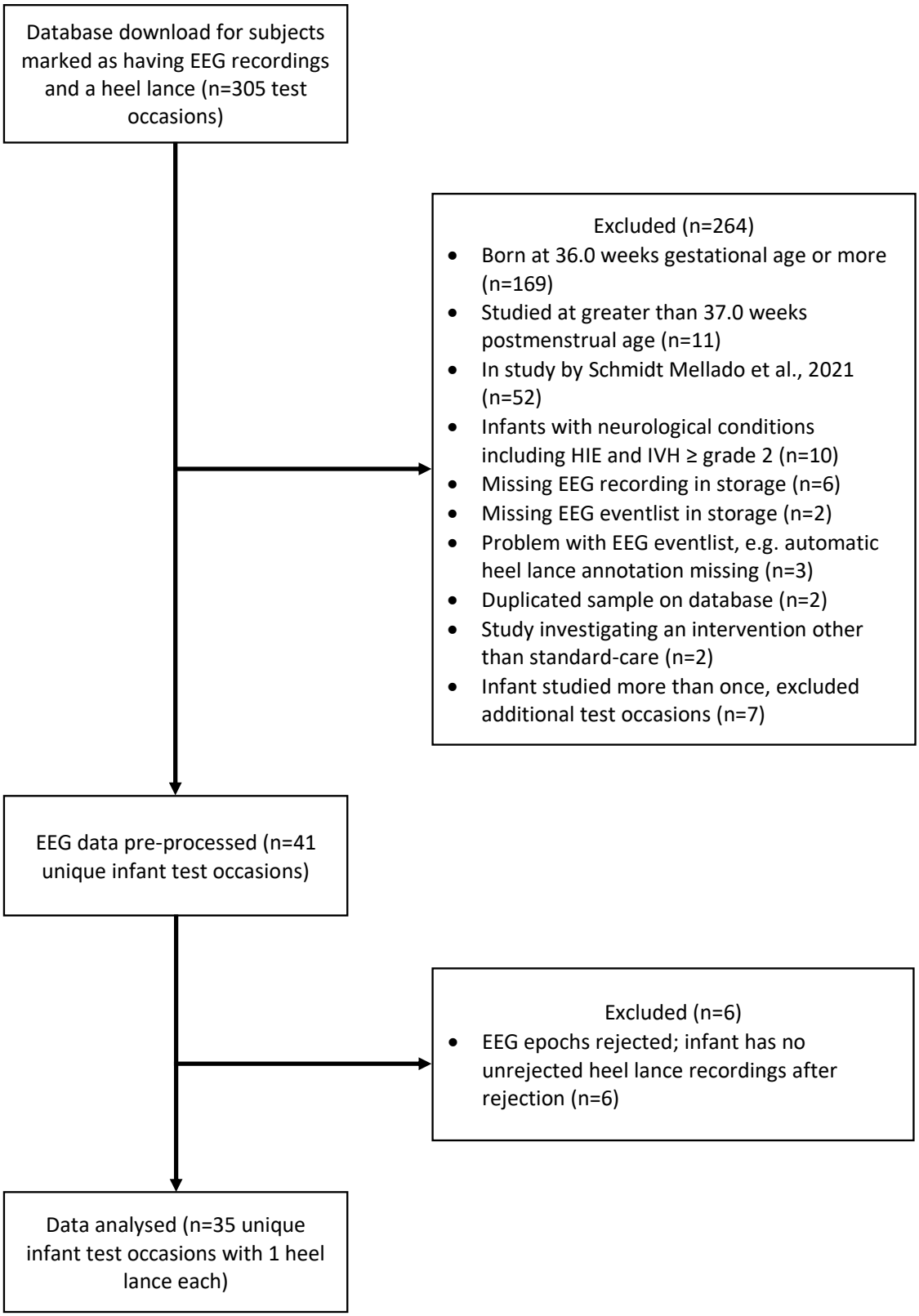


Figure A.6: Flow diagram illustrating the exclusion decisions for Oxford Pilot Dataset B

B

Chapter 3 Supplementary Material

B.1 Documentation for MRI Studies

In addition to the MRI-specific study procedures in the figures below, general SOPs for participant withdrawal and safety reporting are shown in the earlier appendices in figures A.3 and A.4 respectively.

Consent Form

Study ID: | | | | | | | | | |

Infant's name:

Study Title: Investigating pain in the developing human brain
Chief Investigator: Prof Rebeccah Slater [redacted]
[redacted]

Please initial each box

Please complete in black ballpoint pen.

1	I confirm that I have read and understood the information sheet (MRI) (**v. , dated / /**), for the above study. I have had the opportunity to ask questions and have had these answered satisfactorily.	☐
2	I understand that my child's participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my child's medical care or legal rights being affected.	☐
3	I understand that relevant sections of my child's medical notes and data collected during the study may be looked at by individuals from the University of Oxford or Oxford University Hospitals NHS Trust, where it is relevant to my child's taking part in this research. I give permission for these individuals to access to my infant's records.	☐
4	I understand that the MRI scan is for research and is not useful for medical diagnosis, and that scans are not routinely looked at by a doctor. If a concern is raised about a possible abnormality on my child's scan, I will only be informed if a doctor thinks it is medically important such that the finding has clear implications for my child's current or future health.	☐
5	I agree to my child being videoed during the study. I understand that recorded images will not be used for public use, only analysis. No identifiable information, including video recordings or imaging, will be used in any publications/presentations. Only anonymised data will be published or presented at meetings.	☐
6	I agree for the collected data to be used for teaching or academic research presentations.	☐
7	I agree to my child taking part in the above study.	☐
	OPTIONAL	
8	I agree to my child being studied on more than one occasion, up to a maximum of 5 occasions.	☐
9	I agree to complete a parental questionnaire related to my child's study.	☐
10	I consent to being approached in the future about other research studies for which my child may be eligible.	☐
11	I agree to the images/videos of my child recorded during this study being used for publications and presentations.	☐

Name of parent: _____

Relationship to baby: _____

Signature: _____

Date: _____

Name of investigator taking consent: _____

Signature: _____

Date: _____

1 to be kept as part of the study documentation (original)
1 form for parent
1 to be kept with hospital notes

Figure B.1: Sample consent form for recruitment to infant MRI research studies. Personal contact information has been redacted.

Investigating pain in the developing human brain

Parent Information Leaflet

John Radcliffe Hospital

DEPARTMENT OF PAEDIATRICS

UNIVERSITY OF OXFORD

Oxford University Hospitals NHS
NHS Foundation Trust

Your child is, or may be, eligible to take part in a research study. Before you decide, it is important that you understand why the research is being done and what it involves. Please read the following information carefully and ask us if anything is unclear or if you would like more information.

1. **Study title:** Investigating pain in the developing human brain

2. **What is the purpose of the study?**

Infants in hospital often need to have many procedures like blood tests as part of their routine medical treatment, which may cause discomfort. As they cannot tell us how much these procedures hurt, it is difficult to know how much pain they are feeling and to make sure that they receive the right medicines. We know that infants can process discomfort and pain in their brain. By using a special scanner called an MRI (Magnetic Resonance Imaging) scanner, we are able to take detailed 3D pictures of an infant's brain while resting, as well as see how their brain activity changes in response to light, touch, sound or pain. We also know that infants show that they are in pain using different behaviours. These may be indicated by changes in heart rate and breathing in response to pain.

The aim of this research is to understand more about how infants process the outside world and in particular how they feel touch and pain, so that better ways of treating pain can be developed. We are also interested in what infants' brains looks like at rest and how infants respond to different stimuli from their environment, such as light and sound, and how this might change across development.

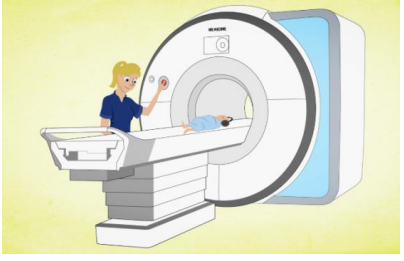
3. Does my child have to take part?

No, it is your decision whether or not your child takes part. If you decide to allow your child to take part, you will be asked to sign a consent form. If you decide you do not want your child to take part, this will not affect your child's care.

If you decide you would like your child to take part, you can change your mind at any time and withdraw your child from the study by telling the research team. You do not have to give a reason. You will be asked if we can use the data/images that have already been collected for analysis (all data) and for publication of results (anonymised data only).

4. What is involved in the study?

If you wish to take part in our study, we will take you and/or your partner and your child to the FMRIB Centre, which is where the research MRI scanner is based at the John Radcliffe Hospital. We will use the scanner to record images of your child's brain and its activity. MRI is a safe magnetic imaging technique, which does not involve radiation, and is used to scan infants in hospital. It produces images by measuring changes in blood oxygen levels.



Before the scan, we will check that your child does not have anything metallic on them, and we will give them ear plugs and ear muffs to protect their hearing and to help them sleep soundly through the scan. If possible, giving your child a feed before the scan will make them sleepy and helps to settle them. We will make sure your child is warm and comfortable before we start scanning.

We may use a soft touch and a sharp touch device, neither of which pierces the skin but stimulates the receptors we are interested in generally without waking or upsetting the infant. We may also stroke your child with a soft brush, turn a light on and off, or play a few sounds to them to look at how your child's brain activity changes in response to these stimuli.

During the scan, we may also record your child's heart rate and oxygen saturations using a small probe wrapped around their foot. We may also video your child's face and/or body movements. We may also approach you to ask you if you are happy for us to use these images for teaching, publicity and/or scientific journals. If you agree, we will take separate consent for this as your child's face would be visible in the video footage. This is not a mandatory part of the study. If you choose not to allow us to use the images in this way, this will not affect your child's care or prevent your child from participating in this research.

A clinically trained researcher will be present throughout the study. If your child wakes or needs feeding or changing during the study, we will stop the scanner and ask you whether you wish to continue after their needs have been met. We may ask you to complete a questionnaire following the study.

As we are interested in how your child's response to pain changes as they grow, we may ask if we can study your child more than once during their stay in hospital. We may also ask you if we can contact you in the future, to ask if you would be happy for your child to take part in other research studies. If you agree that we can contact you in the future about other research studies, we will also record your contact details. Your contact details will not be passed onto anyone outside of the research team. You can opt-out of this at any point by contacting Prof Rebeccah Slater (details below). Your agreement for us to contact you does not form any obligation to participate in future research.

5. Are there any additional risks or benefits for my child?

A clinically-trained member of the research team will ensure your child's safety at all times. As the MRI scanner is not portable, we need to transport your child to and from the MRI scanning suite in a pram or transport pod. MRI is a safe technique used to image both premature and term infants in hospital. The scanner produces loud repetitive sounds, and we have measured the noise level to ensure that we provide all infants with appropriate ear protection. As the MRI scanner is magnetic, before entering the scanning suite we make sure that neither your child nor any member of the research team is carrying any metal.

Obtaining video footage of your child is non-invasive and does not present any risk to your child. Vital sign monitoring (such as heart rate and breathing rate) has been used clinically for over 20 years without any adverse effects. The sensory stimuli which may be applied to your child have similarly been used in many other patient groups. All studies have a dedicated team of healthcare professionals and researchers that will ensure the safety of your child at all times.

The MRI scan is for research rather than medical diagnosis, so it will not be reviewed by a doctor routinely. If any clinically significant findings are identified at the time of the study then the research team will report these to the clinical care team to handle as appropriate.

There are no direct benefits of participating in this research. This study is designed to gather information to help guide improvements in care for infants in the future. If your child becomes distressed, the research study will be paused or stopped.

6. What information will be collected about my child?
We will collect information about your child from the medical notes, including demographic (e.g. ethnicity), clinical (e.g. number of blood tests in hospital), environmental (e.g. ward transfers) and social factors (e.g. postcode). This information helps us to determine which factors may influence the way an infant copes with pain. In the MRI scanner, we will collect images of your child's brain and its activity, and we may collect information about your child's muscle activity, vital signs, and recordings of their facial expressions. All information and videos that are collected during this research study will be kept strictly confidential. Each infant will be allocated a study number which will be used to label all data.

This study has been registered with the data protection registration office and forms part of an educational programme.

7. What will happen to the results?
Results will be analysed and published in a journal. All publications will be made available on our website **https://neuroimaging.paediatrics.ox.ac.uk**. The findings may also be used for teaching or academic research presentations. No identifying information will be presented about you or your child, unless you have provided specific consent for us to use videos/images of your child in this way.

8. What will happen to my child's data?
We will be using information collected from your child and their medical records in order to conduct this study. Research is a task that we perform in the public interest. The University of Oxford, based in the United Kingdom as Sponsor, is the data controller. This means that we, as University of Oxford researchers, are responsible for looking after the information collected and using it properly. We will use the minimum personally-identifiable information possible. We will keep identifiable information about your child for up to 5 years after the study has finished. This excludes any research documents with personal information, such as consent forms, which will be held securely at the University of Oxford for 25 years after the end of the study.

UK Data protection regulation provides you with control over your personal data and how it is used. When you agree to your information being used in research, however, some of those rights may be limited in order for the research to be reliable and accurate. Further information about your rights with respect to your personal data is available at
https://compliance.web.ox.ac.uk/individual-rights
You can find out more about how we use your information from the contacts in section 12.

Research data may be shared with other researchers doing similar work, both here and abroad. Responsible members of the University of Oxford or the Oxford University Hospitals NHS Trust may be given access to data for monitoring and/or audit of the study to ensure we are complying with regulations.

9. Who is organising and funding this research?
This study is sponsored by University of Oxford and has been funded by The Wellcome Trust. Your doctor will not be paid for including you in this study.

10. Who has reviewed the study?
All research that involves NHS patients has to be approved by a Research Ethics Committee. Approval means that the Committee is satisfied that yours and your child's rights will be respected, that any risks have been reduced to a minimum and balanced against possible benefits, and that

you have been given sufficient information on which to make an informed decision about whether to take part. The South Central Oxford C Research Ethics Committee has reviewed and approved this study

11. Comments or concerns during the study

The University has arrangements in place to provide for harm arising from participation in the study for which the University is the Research Sponsor. NHS indemnity operates in respect of the clinical treatment with which your child is provided. If you wish to complain about any aspect of the way in which you have been approached or treated during the course of this study, you should contact Dr Rebeccah Slater (details below) or the University of Oxford Research Governance, Ethics & Assurance (RGEA) office ([redacted]).

12. Contact for further information

[redacted]

Picture shows example of an MRI study.

Thank you for reading this information leaflet.

Figure B.2: Sample patient information leaflet for infant MRI research studies. Personal contact information has been redacted.

SCOPE

This SOP describes the role of the clinical member of staff in an MRI study. It covers the transport to and from the ward to the MRI scanner in FMRIB, the pre-scanning safety checks and the monitoring of the infant during the study.

PROCEDURE

Magnet safety training and MRI safety screening

Everyone attending MRI scans must have:

1. Completed magnet safety training in the past year. Access to FMRIB is automatically revoked if you do not renew your magnet safety training yearly.
2. Completed a volunteer MRI safety screening form in the current calendar year. All forms are located at the scanners and will be checked by a Radiographer. If your health status changes (recent surgeries/implants, pregnancy), fill out a new form and speak to a radiographer before attending any future MRI studies.

A record of training will be held in the Paediatric Neuroimaging Group research office.

On the day of the study

Preparation

- The clinical member of the MRI team should ensure that the grab bag, a full oxygen cylinder, resuscitation kit and portable suction unit are prepared, stocked, and in working order. **Note: These items are not MRI-safe and should NEVER be taken inside the scanning room.**
- The clinical member should consult the midwife caring for the infant and ensure that the infant (and the mother if she wishes to attend the scan) is safe to transport and there are no new contraindications.
- The clinical member should examine the baby to ensure that they are satisfied that the baby is clinically stable for transfer and scanning.
- The baby's most recent weight (within 5 days of the scan) should be documented on the study sheet.

The infant will be transported to FMRIB by a trained clinical member of staff, parent, and an MRI-trained researcher.

In the FMRIB subject room

- Fully undress the baby and remove ECG dots, medicated skin patches, jewellery, any metallic objects in linen (e.g. safety pins) and metal fasteners on clothing that are either touching bare skin or near the head. Don't forget to look at the baby's back and in the nappy!
- Dress in plain vest (no logos or decorative stitching) and ensuring no metal poppers anywhere other than between the legs. These should be over the nappy (not in contact with the skin) and a blanket will be placed between the legs once in the scanner to ensure that there is no contact between the legs.
- **If a cannula is in situ:** the cannula splint is NOT 3T MRI safe and must be removed.
- Measure and record the baby's temperature on the data collection sheet. The baby should not be scanned if their temperature is not within the normal range (36.5-37.5^{o}C)
- Apply the first two levels of ear protection:
    - Take approximately half a ball of ear putty and apply to seal the outer ear (not in ear canal).
    - Apply mini muffs on top.
- Swaddle the baby in MRI safe hospital blankets. Make sure there is no skin-to-skin contact, i.e. arms cannot touch, so make sure to apply blanket between the chest (if not wearing a vest) and arms and between arms and legs. Hand should be tucked away in the blankets.
- The researcher will ask the parent(s) to fill out MRI safety forms for their baby and themselves (if they wish to come into the scanner room) and the radiographer will formally screen the parents prior to entering the scanner control room. The radiographer will confirm that the baby has been screened by the clinical member of the research team.
- The MRI team must be in scrubs and check they have emptied their pockets, removed jewellery and are not carrying any contraindicated items. The only items taken into the control should be the MRI-safe foam cradle, MRI-safe step, MRI-compatible pinpricks, disposable dummy (in case of requirement), MRI-safe oxygen saturation probe wrap. All other items should be left in subject room or locker.

CHECKPOINT

Before entering the scanner room, radiographer confirms with the lead researcher and clinical member of the MRI team that the necessary checks have been carried out.

In FMRIB scanner room

- Only take the infant into the scanner room after they have been checked fully screened by the clinical member of staff and approval has been granted to enter by the radiographer.
- Place the foam cradle in the head coil and lay the bean bag cushion on top.
- Place baby on the beanbag/cradle, ensuring the head is positioned centrally in the head coil (see markings and consult radiographer if necessary).
- Apply the third level of ear protection: place the large Ems4Bubs earmuffs over the mini muffs so that they are completely covered and comfortable for the baby. There should be a good seal between the skin and the muffs. Insert padding (foam triangles covered in paper towel provided by radiographer) on either side of the head to ensure the earmuffs are securely held against the head. The muffs should not be able to move but should not cause discomfort.
- Swaddle the infant in blankets, ensuring that neither the baby's arms nor legs are in direct contact with each other or against the metal poppers of the nappy (separate with blanket). Leave the legs free.
- Remove air from beanbag cushion to fix position of the infant once settled and aligned.
- Apply the two saturation probes: On the right foot - place the probe used for real-time clinical monitoring of oxygen saturations in the control room during scanning. On the right leg - place the probe used for recording physiological data. Use MRI-safe wraps to cover the probes and tape if necessary to secure them.
- Cover the legs loosely with another MRI-safe standard hospital blanket.
- Ensure that HR and oxygen saturations are reading well (good trace) and are within normal limits. Record initial vital sign readings on the data collection sheet.
- During the study, the clinical member should continuously monitor the vital signs. They should alert the research team and radiographer if they have any concerns.
- During the scan, a member of the research team must regularly update the parents and ensure they are happy to continue.
- At the end of the scan, the clinical member should record a final reading of the infant's heart rate and oxygen saturations on the data collection sheet prior to removing them from the scanner.

CHECKPOINT

Before starting the scan, ensure that the three levels of ear protection are well fitted, the infant is settled, the probe is reading correctly, and that you have filled out the clinical assessment on the NIPI MRI data collection form.

If during scanning the infant becomes distressed or HR or oxygen saturation readings are outside normal limits, the clinician should enter the scanner to check the infant, ensure the probe is reading correctly, and settle the infant if appropriate. If the clinician deems that further assessment of the infant is required, the clinician should remove the infant from the scanner. Any resuscitation or stabilisation **MUST** take place in the subject room, outside of the scanning environment. **NO** equipment should be brought into the control room or scanner.

In FMRIB subject room

After the scan,

- Remove monitoring and take the infant from the scanner to the subject room.
- Remove the ear protection.
- Take the infant's temperature and record this on the data collection sheet.
- Dress the infant in appropriate clothes for transport including a hat and blankets.

CHECKPOINT

After scan completion, ensure that the clinical assessment NIPI MRI data collection form has been filled out.

Documentation

Use the MRI data collection sheet to record the infant's temperature, heart rate, and oxygen saturations pre- and post-scan.

Figure B.3: SOP for Infant MRI Transport and Safety. Title page omitted due presence of personal details.

CHECKLIST FOR RESEARCH MRI STUDIES

Checklist during consent process

☐ Consent form signed, dated and baby's name filled out
☐ Sticker on consent form
☐ Written in notes
☐ Clinical member available for transport and monitoring during the study
☐ Confirmed with clinical team that baby is appropriate to study
☐ Parents have seen the stimulus
☐ Parents have been informed about the noise and the ear protection
☐ Discussed pick up time with parents
☐ Asked parents to feed before this time (or can feed in subject room if timing difficult)
☐ Asked parents to dress baby in a plain vest without logos or stitching or metal poppers on the shoulders. Metal poppers between the legs are not a problem.
☐ If baby has a cannula in situ; asked parents to cover the cannula with a sock. The splint will need to be removed just prior to scanning.
☐ Discussed with parents who is attending the scan and whether they will want to stay in the scanner room. Explained that they will be required to complete a volunteer form and be screened for contraindication by radiographer on the day of the scan
☐ Asked mum if she needs a wheelchair

Checklist required kit

☐ MRI team in scrubs and jewellery removed.

Pram

☐ Pram set up with sheets and blankets
☐ Hat for baby
☐ Socks for baby
☐ Full oxygen cylinder and flow regulator that is compatible and fitted
☐ MRI-safe step (to help researcher get onto MRI scanner table)

Grab bag – non-medical (blue sports bag)

☐ Note baby's weight
☐ NIPI MRI data collection form
☐ Earmuffs (Ems' 4 bubs)
☐ Mini Muffs (Natus)
☐ Disposable covers for earmuffs
☐ Ear putty (Mack's)
☐ Thermometer or tempadots for baby
☐ Stimulator: MRI compatible PinPricks

WARNING

GRAB BAG, PRAM AND RESUSCITATION KIT CONTAIN ITEMS THAT ARE NOT MRI SAFE!
THEY SHOULD BE KEPT IN THE SUBJECT ROOM AND NEVER TAKEN INTO THE SCANNER ROOM

☐ Disposable dummies
☐ Spare nappies, wipes, cotton wool and bowl
☐ Orange foam cradle
☐ Single-use paper measuring tapes
☐ Laminated copies of SOPs

Grab bag – medical (blue sports bag)

☐ Bag-valve-mask
☐ Oxygen tubing
☐ Tape
☐ Scissors
☐ Stethoscope
☐ Two sets of Level 1 PPE with eye protection (gloves, aprons, surgical masks, goggles or any other eye protection)
☐ Pulse oximeter wrap

Full neonatal resuscitation Kit (green case)

☐ Checked listed contents and expiry dates
☐ Portable suction unit

CHECK Crash Team is aware of scan

Contact Paediatrics crash bleep holders and Neonatal crash bleep holders to inform them that MRI research study is taking place and ensure they know route to FMRIB in case of emergency.

After the study

Any unit linen should be put in the laundry baskets in LDU. Any used equipment should be cleaned.

Figure B.4: Infant MRI Preparation and Conduct Checklist.

References

[1]J. Villar, L. Cheikh Ismail, E. Staines Urias, F. Giuliani, E. O. Ohuma, C. G. Victora, A. T. Papageorghiou, D. G. Altman, C. Garza, F. C. Barros, F. Puglia, R. Ochieng, Y. A. Jaffer, J. A. Noble, E. Bertino, M. Purwar, R. Pang, A. Lambert, C. Chumlea, A. Stein, M. Fernandes, Z. A. Bhutta, and S. H. Kennedy. The Satisfactory Growth and Development at 2 Years of Age of the INTERGROWTH-21st Fetal Growth Standards Cohort Support Its Appropriateness for Constructing International Standards. *American Journal of Obstetrics and Gynecology*, 218:S841–S854.e2, 2 Suppl, 2018. DOI: 10.1016/j.ajog.2017.11.564.

[2]A. Blasi, S. Lloyd-Fox, L. Katus, and C. E. Elwell. fNIRS for Tracking Brain Development in the Context of Global Health Projects. *Photonics*, 6(3):89, 3, 2019. DOI: 10.3390/photonics6030089.

[3]WHO. Infant and young child feeding. URL: https://www.who.int/news-room/fact-sheets/detail/infant-and-young-child-feeding (visited on Apr. 21, 2022).

[4]Institute of Medicine (US) Committee on Improving Birth Outcomes. Glossary. In *Improving Birth Outcomes: Meeting the Challenge in the Developing World*, pages 324–33. National Academies Press (US), 2003. DOI: 10.17226/10841.

[5]WHO. Newborn health. URL: https://www.who.int/westernpacific/health-topics/newborn-health (visited on Apr. 21, 2022).

[6]WHO. WHO: Recommended Definitions, Terminology and Format for Statistical Tables Related to The Perinatal Period And Use of A New Certificate For Cause of Perinatal Deaths. *Acta Obstetricia et Gynecologica Scandinavica*, 56(3):247–253, 1977. DOI: 10.3109/00016347709162009.

[7]H. Blencowe, S. Cousens, D. Chou, M. Oestergaard, L. Say, A.-B. Moller, M. Kinney, J. Lawn, and the Born Too Soon Preterm Birth Action Group (see acknowledgement for full list). Born Too Soon: The global epidemiology of 15 million preterm births. *Reproductive Health*, 10(1):S2, 2013. DOI: 10.1186/1742-4755-10-S1-S2.

[8]N. Marlow. Full term; an artificial concept. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 97(3):F158–F159, 2012. DOI: 10.1136/fetalneonatal-2011-301507.

[9]Committee on Fetus and Newborn. Age Terminology During the Perinatal Period. *Pediatrics*, 114(5):1362–1364, 2004. DOI: 10.1542/peds.2004-1915.

[10]L. E. Jones. *First Steps: The Early Years of IASP 1973–1984*. IASP Press, Seattle, Washington, 2010.

[11]IASP Subcommittee on Taxonomy. Pain Terms: A List With Definitions and Notes on Usage. *PAIN*, 6(3):249–252, 1979. DOI: 10.1016/0304-3959(79)90046-0.

[12]S. N. Raja, D. B. Carr, M. Cohen, N. B. Finnerup, H. Flor, S. Gibson, F. J. Keefe, J. S. Mogil, M. Ringkamp, K. A. Sluka, X.-J. Song, B. Stevens, M. D. Sullivan, P. R. Tutelman, T. Ushida, and K. Vader. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *PAIN*, 161(9):1976–1982, 2020. DOI: 10.1097/j.pain.0000000000001939.

[13]B. Stevens. Revisions to the IASP definition of pain—What does this mean for children? *Paediatric & Neonatal Pain*, 3(3):101–105, 2021. DOI: 10.1002/pne2.12047.

[14]C.-W. Woo, L. Schmidt, A. Krishnan, M. Jepma, M. Roy, M. A. Lindquist, L. Y. Atlas, and T. D. Wager. Quantifying cerebral contributions to pain beyond nociception. *Nature Communications*, 8:14211, 2017.

[15]A. E. Dubin and A. Patapoutian. Nociceptors: the sensors of the pain pathway. *The Journal of Clinical Investigation*, 120(11):3760–3772, 2010. DOI: 10.1172/JCI42843.

[16]K. J. S. Anand. Defining pain in newborns: need for a uniform taxonomy? *Acta Paediatrica (Oslo, Norway : 1992)*, 106(9):1438–1444, 2017. DOI: 10.1111/apa.13936.

[17]K. J. S. Anand, W. G. Sippell, and A. Aynsley-Green. Randomised Trial of Fentanyl Anaesthesia in Preterm Babies Undergoing Surgery: Effects on the Stress Response. *The Lancet*, 329(8527):243–248, 1987. DOI: 10.1016/S0140-6736(87)90065-1.

[18]P. J. McGrath. Science is not enough: The modern history of pediatric pain. *PAIN*, 152(11):2457–2459, 2011. DOI: 10.1016/j.pain.2011.07.018.

[19]N. Kuratani. The cutting edge of neonatal anesthesia: the tide of history is changing. *Journal of Anesthesia*, 29(1):1–3, 2015. DOI: 10.1007/s00540-014-1817-7.

[20]A. T. Costarino and J. J. Downes. Pediatric Anesthesia Historical Perspective. *Anesthesiology Clinics of North America*, 23(4):573–595, 2005. DOI: 10.1016/j.atc.2005.08.005.

[21]J. R. Lawson. Letters. *Birth*, 13(2):124–125, 1986. DOI: 10.1111/j.1523-536X.1986.tb01024.x.

[22]J. R. Lawson. More on Newborn Surgery Without Anesthesia. *Birth*, 15(1):36–36, 1988. DOI: 10.1111/j.1523-536X.1988.tb01083.x.

[23]H. Harrison. Letters. *Birth*, 13(2):124–124, 1986. DOI: 10.1111/j.1523-536X.1986.tb01023.x.

[24]M. H. Shearer. Surgery on the Paralyzed, Unanesthetized Newborn. *Birth*, 13(2):79–79, 1986. DOI: 10.1111/j.1523-536X.1986.tb01012.x.

[25]K. Anand and P. Hickey. Pain and Its Effects in the Human Neonate and Fetus. *New England Journal of Medicine*, 317(21):1321–1329, 1987. DOI: 10.1056/NEJM198711193172105.

[26]Committee on Fetus and Newborn, Committee on Drugs, Section on Anesthesiology, and Section on Surgery. Neonatal Anesthesia. *Pediatrics*, 80(3):446, 1987. DOI: 10.1542/peds.80.3.446.

[27]A. B. Fletcher. Pain in the neonate. *The New England Journal of Medicine*, 317(21):1347–1348, 1987. DOI: 10.1056/NEJM198711193172110.

[28]F. A. Berry and G. A. Gregory. Do Premature Infants Require Anesthesia for Surgery? *Anesthesiology*, 67(3):291–293, 1987. DOI: 10.1097/00000542-198709000-00002.

[29]M. Yaster. Analgesia and anesthesia in neonates. *The Journal of Pediatrics*, 111(3):394–396, 1987. DOI: 10.1016/S0022-3476(87)80461-4.

[30]C. H. Robson. Anesthesia in Children. *Anesthesia and Analgesia*, 4(4):235–240, 1925.

[31]P. S. Williamson and M. L. Williamson. Physiologic Stress Reduction by a Local Anesthetic During Newborn Circumcision. *Pediatrics*, 71(1):36–40, 1983. DOI: 10.1542/peds.71.1.36.

[32]S. Dixon, J. Snyder, R. Holve, and P. Bromberger. Behavioral effects of circumcision with and without anesthesia. *Journal of Developmental and Behavioral Pediatrics*, 5(5):246–250, 1984. DOI: 10.1097/00004703-198410000-00004.

[33]M. H. Shearer. Newborn Surgery Without Anesthesia: Is the Problem Solved? *Birth*, 15(1):36–38, 1988. DOI: 10.1111/j.1523-536X.1988.tb01084.x.

[34]N. Wellington and M. J. Rieder. Attitudes and Practices Regarding Analgesia for Newborn Circumcision. *Pediatrics*, 92(4):541–543, 1993. DOI: 10.1542/peds.92.4.541.

[35]C. R. Howard, F. M. Howard, L. C. Garfunkel, E. A. de Blieck, and M. Weitzman. Neonatal Circumcision and Pain Relief: Current Training Practices. *Pediatrics*, 101(3):423–428, 1998. DOI: 10.1542/peds.101.3.423.

[36]A. Taddio, J. Katz, A. L. Ilersich, and G. Koren. Effect of neonatal circumcision on pain response during subsequent routine vaccination. *Lancet*, 349(9052):599–603, 1997. DOI: 10.1016/S0140-6736(96)10316-0.

[37]C. Kirya and M. W. Werthmann. Neonatal circumcision and penile dorsal nerve block—a painless procedure. *The Journal of Pediatrics*, 92(6):998–1000, 1978. DOI: 10.1016/S0022-3476(78)80386-2.

[38]D. O' Sullivan, M. Heal, and C. Powell. Circumcision: how do urologists do it? *British Journal of Urology*, 78(2):265–270, 1996. DOI: 10.1046/j.1464-410X.1996.10822.x.

[39]C. R. McLaughlin, J. G. Hull, W. H. Edwards, C. P. Cramer, and W. L. Dewey. Neonatal pain: A comprehensive survey of attitudes and practices. *Journal of Pain and Symptom Management*, 8(1):7–16, 1993. DOI: 10.1016/0885-3924(93)90114-B.

[40]N. Rutter and J. Richardson. A survey of the use of analgesia in newborn intensive care. *International Journal of Pharmacy Practice*, 1(4):220–222, 1992. DOI: 10.1016/S1084-2756(98)80084-5.

[41]N. Rutter and L. Doyal. Neonatal care and management of pain: Historical and ethical issues. *Seminars in Neonatology*, 3(4):297–302, 1998. DOI: 10.1016/S1084-2756(98)80084-5.

[42]R. Carbajal, A. Rousset, C. Danan, S. Coquery, P. Nolent, S. Ducrocq, C. Saizou, A. Lapillonne, M. Granier, P. Durand, R. Lenclen, A. Coursol, P. Hubert, L. de Saint Blanquat, P.-Y. Boëlle, D. Annequin, P. Cimerman, K. J. S. Anand, and G. Bréart. Epidemiology and Treatment of Painful Procedures in Neonates in Intensive Care Units. *JAMA*, 300(1):60–70, 2008. DOI: 10.1001/jama.300.1.60.

[43]M. D. Cruz, A. M. Fernandes, and C. R. Oliveira. Epidemiology of painful procedures performed in neonates: A systematic review of observational studies. *European Journal of Pain*, 20(4):489–498, 2016. DOI: 10.1002/ejp.757.

[44]E. Courtois, P. Cimerman, V. Dubuche, M.-F. Goiset, C. Orfèvre, A. Lagarde, B. Sgaggero, C. Guiot, M. Goussot, E. Huraux, M.-C. Nanquette, C. Butel, A.-M. Ferreira, S. Lacoste, S. Séjourné, V. Jolly, G. Lajoie, V. Maillard, R. Guedj, H. Chappuy, and R. Carbajal. The burden of venipuncture pain in neonatal intensive care units: EPIPPAIN 2, a prospective observational study. *International Journal of Nursing Studies*, 57:48–59, 2016. DOI: 10.1016/j.ijnurstu.2016.01.014.

[45]E. Courtois, S. Droutman, J.-F. Magny, Z. Merchaoui, X. Durrmeyer, C. Roussel, V. Biran, S. Eleni, G. Vottier, S. Renolleau, L. Desfrere, F. Castela, N. Boimond, D. Mellah, P. Bolot, A. Coursol, D. Brault, H. Chappuy, P. Cimerman, K. J. S. Anand, and R. Carbajal. Epidemiology and neonatal pain management of heelsticks in intensive care units: EPIPPAIN 2, a prospective observational study. *International Journal of Nursing Studies*, 59:79–88, 2016. DOI: 10.1016/j.ijnurstu.2016.03.014.

[46]C. C. Johnston, R. L. Stremler, B. J. Stevens, and L. J. Horton. Effectiveness of oral sucrose and simulated rocking on pain response in preterm neonates. *PAIN*, 72(1):193–199, 1997. DOI: 10.1016/S0304-3959(97)00033-X.

[47]C. C. Johnston, F. Filion, L. Snider, A. Majnemer, C. Limperopoulos, C.-D. Walker, A. Veilleux, E. Pelausa, H. Cake, S. Stone, A. Sherrard, and K. Boyer. Routine sucrose analgesia during the first week of life in neonates younger than 31 weeks' postconceptional age. *Pediatrics*, 110(3):523–528, 2002. DOI: 10.1542/peds.110.3.523.

[48]A. Taddio, V. Shah, E. Atenafu, and J. Katz. Influence of repeated painful procedures and sucrose analgesia on the development of hyperalgesia in newborn infants. *Pain*, 144(1-2):43–48, 2009. DOI: 10.1016/j.pain.2009.02.012.

[49]R. Slater, L. Cornelissen, L. Fabrizi, D. Patten, J. Yoxen, A. Worley, S. Boyd, J. Meek, and M. Fitzgerald. Oral sucrose as an analgesic drug for procedural pain in newborn infants: a randomised controlled trial. *The Lancet*, 376(9748):1225–1232, 2010. DOI: 10.1016/S0140-6736(10)61303-7.

[50]R. Carbajal, M. Eriksson, E. Courtois, E. Boyle, A. Avila-Alvarez, R. D. Andersen, K. Sarafidis, T. Polkki, C. Matos, P. Lago, T. Papadouri, S. A. Montalto, M.-L. Ilmoja, S. Simons, R. Tameliene, B. van Overmeire, A. Berger, A. Dobrzanska, M. Schroth, L. Bergqvist, H. Lagercrantz, and K. J. S. Anand. Sedation and analgesia practices in neonatal intensive care units (EUROPAIN): results from a prospective cohort study. *The Lancet Respiratory Medicine*, 3(10):796–812, 2015. DOI: 10.1016/S2213-2600(15)00331-8.

[51]E. Agakidou, K. Tsoni, T. Stathopoulou, A. Thomaidou, M. Farini, A. Kontou, P. Karagianni, and K. Sarafidis. Changes in Physicians' Perceptions and Practices on Neonatal Pain Management Over the Past 20 Years. A Survey Conducted at Two Time-Points. *Frontiers in Pediatrics*, 9:667806, 2021. DOI: 10.3389/fped.2021.667806.

[52]C. Johnston, M. Campbell-Yeo, T. Disher, B. Benoit, A. Fernandes, D. Streiner, D. Inglis, and R. Zee. Skin-to-skin care for procedural pain in neonates. *Cochrane Database of Systematic Reviews*, (2), 2017. DOI: 10.1002/14651858.CD008435.pub3.

[53]M. Ranger, S. Tremblay, C. M. Y. Chau, L. Holsti, R. E. Grunau, and D. Goldowitz. Adverse Behavioral Changes in Adult Mice Following Neonatal Repeated Exposure to Pain and Sucrose. *Frontiers in Psychology*, 9, 2019.

[54]C. McPherson, S. P. Miller, M. El-Dib, A. N. Massaro, and T. E. Inder. The influence of pain, agitation, and their management on the immature brain. *Pediatric Research*, 88(2):168–175, 2020. DOI: 10.1038/s41390-019-0744-6.

[55]R. E. Grunau, M. F. Whitfield, J. Petrie-Thomas, A. R. Synnes, I. L. Cepeda, A. Keidar, M. Rogers, M. MacKay, P. Hubber-Richard, and D. Johannesen. Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. *Pain*, 143(1-2):138–146, 2009. DOI: 10.1016/j.pain.2009.02.014.

[56]M. Ranger and R. E. Grunau. Early repetitive pain in preterm infants in relation to the developing brain. *Pain management*, 4(1):57–67, 2014. DOI: 10.2217/pmt.13.61.

[57]B. J. Stevens, S. Gibbins, J. Yamada, K. Dionne, G. Lee, C. Johnston, and A. Taddio. The Premature Infant Pain Profile-Revised (PIPP-R): Initial Validation and Feasibility. *The Clinical Journal of Pain*, 30(3):238–243, 2014. DOI: 10.1097/AJP.0b013e3182906aed.

[58]R. V. E. Grunau and K. D. Craig. Pain expression in neonates: facial action and cry. *PAIN*, 28(3):395–410, 1987. DOI: 10.1016/0304-3959(87)90073-X.

[59]R. V. E. Grunau, C. C. Johnston, and K. D. Craig. Neonatal facial and cry responses to invasive and non-invasive procedures. *PAIN*, 42(3):295–305, 1990. DOI: 10.1016/0304-3959(90)91142-6.

[60]J. W. B. Peters, H. M. Koot, R. E. Grunau, J. de Boer, M. J. van Druenen, D. Tibboel, and H. J. Duivenvoorden. Neonatal Facial Coding System for assessing postoperative pain in infants: item reduction is valid and feasible. *The Clinical Journal of Pain*, 19(6):353–363, 2003. DOI: 10.1097/00002508-200311000-00003.

[61]S. Gibbins, B. J. Stevens, J. Yamada, K. Dionne, M. Campbell-Yeo, G. Lee, K. Caddell, C. Johnston, and A. Taddio. Validation of the Premature Infant Pain Profile-Revised (PIPP-R). *Early Human Development*, 90(4):189–193, 2014. DOI: 10.1016/j.earlhumdev.2014.01.005.

[62]M. Ranger, C. C. Johnston, and K. J. S. Anand. Current Controversies Regarding Pain Assessment in Neonates. *Seminars in Perinatology*, 31(5):283–288, 2007. DOI: 10.1053/j.semperi.2007.07.003.

[63]R. G. Barr. Reflections on measuring pain in infants: dissociation in responsive systems and "honest signalling". *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 79(2):F152–F156, 1998. DOI: 10.1136/fn.79.2.F152.

[64]G. Sellaml, E. L. Cignaccol, K. D. Craigl, and S. Engbergl. Review: Contextual factors influencing pain response to heelstick procedures in preterm infants: What do we know? A systematic review. *European Journal of Pain*, 15(7):661.e1–661.e15, 2011. DOI: 10.1016/j.ejpain.2011.01.002.

[65]R. Slater, A. Cantarella, L. Franck, J. Meek, and M. Fitzgerald. How Well Do Clinical Pain Assessment Tools Reflect Pain in Infants? *PLoS Medicine*, 5(6):e129, 2008. DOI: 10.1371/journal.pmed.0050129.

[66]D. Gursul, C. Hartley, and R. Slater. Nociception and the neonatal brain. *Seminars in Fetal and Neonatal Medicine*, 2019. DOI: 10.1016/j.siny.2019.05.008.

[67]S. Ahola Kohut and R. Pillai Riddell. Does the Neonatal Facial Coding System Differentiate Between Infants Experiencing Pain-Related and Non-Pain-Related Distress? *The Journal of Pain*, 10(2):214–220, 2009. DOI: 10.1016/j.jpain.2008.08.010.

[68]M. M. Cobo, G. Green, F. Andritsou, L. Baxter, R. Evans Fry, A. Grabbe, D. Gursul, A. Hoskin, G. S. Mellado, M. van der Vaart, E. Adams, A. Bhatt, F. Denk, C. Hartley, and R. Slater. Early life inflammation is associated with spinal cord excitability and nociceptive sensitivity in human infants. *Nature Communications*, 13(1):3943, 1, 2022. DOI: 10.1038/s41467-022-31505-y.

[69]L. B. Mokkink, C. B. Terwee, D. L. Patrick, J. Alonso, P. W. Stratford, D. L. Knol, L. M. Bouter, and H. C. W. de Vet. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7):737–745, 2010. DOI: 10.1016/j.jclinepi.2010.02.006.

[70]H. C. W. de Vet. *Measurement in Medicine: A Practical Guide*. Practical Guides to Biostatistics and Epidemiology. Cambridge University Press, Cambridge ; New York, 2011.

[71]V. Giordano, J. Edobor, P. Deindl, B. Wildner, K. Goeral, P. Steinbauer, T. Werther, A. Berger, and M. Olischar. Pain and Sedation Scales for Neonatal and Pediatric Patients in a Preverbal Stage of Development: A Systematic Review. *JAMA Pediatrics*, 173(12):1186–1197, 2019. DOI: 10.1001/jamapediatrics.2019.3351.

[72]N. Meesters, T. Dilles, S. Simons, and M. van Dijk. Do Pain Measurement Instruments Detect the Effect of Pain-Reducing Interventions in Neonates? A Systematic Review on Responsiveness. *The Journal of Pain*, 20(7):760–770, 2019. DOI: 10.1016/j.jpain.2018.12.005.

[73]L. B. Mokkink, C. B. Terwee, D. L. Patrick, J. Alonso, P. W. Stratford, D. L. Knol, L. M. Bouter, and H. C. W. de Vet. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, 19(4):539–549, 2010. DOI: 10.1007/s11136-010-9606-8.

[74]C. B. Terwee, L. B. Mokkink, D. L. Knol, R. W. J. G. Ostelo, L. M. Bouter, and H. C. W. de Vet. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*, 21(4):651–657, 2012. DOI: 10.1007/s11136-011-9960-1.

[75]S. M. Walker, L. S. Franck, M. Fitzgerald, J. Myles, J. Stocks, and N. Marlow. Long-term impact of neonatal intensive care and surgery on somatosensory perception in children born extremely preterm. *PAIN*, 141(1):79–87, 2009. DOI: 10.1016/j.pain.2008.10.012.

[76]S. Walker, A. Melbourne, H. O'Reilly, J. Beckmann, Z. Eaton-Rosen, S. Ourselin, and N. Marlow. Somatosensory function and pain in extremely preterm young adults from the UK EPICure cohort: sex-dependent differences and impact of neonatal surgery. *British Journal of Anaesthesia*, 121(3):623–635, 2018. DOI: 10.1016/j.bja.2018.03.035.

[77]R. E. Grunau. Neonatal pain in very preterm infants: long-term effects on brain, neurodevelopment and pain reactivity. *Rambam Maimonides Medical Journal*, 4(4):e0025, 2013. DOI: 10.5041/RMMJ.10132.

[78]B. O. Valeri, M. Ranger, C. M. Y. Chau, I. L. Cepeda, A. Synnes, M. B. M. Linhares, and R. E. Grunau. Neonatal Invasive Procedures Predict Pain Intensity at School Age in Children Born Very Preterm. *The Clinical Journal of Pain*, 32(12):1086–1093, 2016. DOI: 10.1097/AJP.0000000000000353.

[79]J. Vinall, S. P. Miller, B. H. Bjornson, K. P. Fitzpatrick, K. J. Poskitt, R. Brant, A. R. Synnes, I. L. Cepeda, and R. E. Grunau. Invasive Procedures in Preterm Children: Brain and Cognitive Development at School Age. *Pediatrics*, 133(3):412–421, 2014. DOI: 10.1542/peds.2013-1863.

[80]E. G. Duerden, R. E. Grunau, T. Guo, J. Foong, A. Pearson, S. Au-Young, R. Lavoie, M. M. Chakravarty, V. Chau, A. Synnes, and S. P. Miller. Early Procedural Pain Is Associated with Regionally-Specific Alterations in Thalamic Development in Preterm Neonates. *The Journal of Neuroscience*, 38(4):878–886, 2018. DOI: 10.1523/JNEUROSCI.0867-17.2017.

[81]J. Hohmeister, A. Kroll, I. Wollgarten-Hadamek, K. Zohsel, S. Demirakça, H. Flor, and C. Hermann. Cerebral processing of pain in school-aged children with neonatal nociceptive input: An exploratory fMRI study. *Pain*, 150(2):257–267, 2010. DOI: 10.1016/j.pain.2010.04.004.

[82]J. N. van den Anker, M. D. Reed, K. Allegaert, and G. L. Kearns. Developmental Changes in Pharmacokinetics and Pharmacodynamics. *The Journal of Clinical Pharmacology*, 58:S10–S25, 2018. DOI: 10.1002/jcph.1284.

[83]K. O'Hara, J. H. Martin, and J. J. Schneider. Barriers and Challenges in Performing Pharmacokinetic Studies to Inform Dosing in the Neonatal Population. *Pharmacy: Journal of Pharmacy Education and Practice*, 8(1):16, 2020. DOI: 10.3390/pharmacy8010016.

[84]J. N. van den Anker and K. Allegaert. Acetaminophen in the Neonatal Intensive Care Unit: Shotgun Approach or Silver Bullet. *The Journal of Pediatrics*, 198:10–11, 2018. DOI: 10.1016/j.jpeds.2018.02.046.

[85]C. Eccleston, E. Fisher, R. F. Howard, R. Slater, P. Forgeron, T. M. Palermo, K. A. Birnie, B. J. Anderson, C. T. Chambers, G. Crombez, G. Ljungman, I. Jordan, Z. Jordan, C. Roberts, N. Schechter, C. B. Sieberg, D. Tibboel, S. M. Walker, D. Wilkinson, and C. Wood. Delivering transformative action in paediatric pain: a Lancet Child & Adolescent Health Commission. *The Lancet Child & Adolescent Health*, 5(1):47–87, 2021. DOI: 10.1016/S2352-4642(20)30277-7.

[86]European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP), and Paediatric Committee (PDCO). Guideline on investigation of medicinal products in the term and preterm neonate (EMEA/536810/2008). European Medicines Agency. 2010. URL: https://www.ema.europa.eu/en/investigation-medicinal-products-term-preterm-neonate (visited on Aug. 24, 2022).

[87]European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP), and Paediatric Committee (PDCO). Concept paper on the need for revision of the guideline on the investigation of medicinal products in the term and preterm neonate (Adopted, EMA/PDCO/362462/2016). European Medicines Agency. 2019. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/adopted-concept-paper-need-revision-guideline-investigation-medicinal-products-term-preterm-neonate_en.pdf (visited on Aug. 24, 2022).

[88]FDA, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research. E11 Clinical Investigation of Medicinal Products in the Pediatric Population (FDA-2000-D-0942). U.S. Food and Drug Administration (FDA). 2000. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e11-clinical-investigation-medicinal-products-pediatric-population (visited on Aug. 24, 2022).

[89]FDA, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research. E11(R1) Addendum: Clinical Investigation of Medicinal Products in the Pediatric Population (FDA-2016-D-3848). U.S. Food and Drug Administration (FDA). 2018. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e11r1-addendum-clinical-investigation-medicinal-products-pediatric-population (visited on Aug. 24, 2022).

[90]R. M. Ward and C. M. Sherwin. Newborns still lack drug data to guide therapy. *British Journal of Clinical Pharmacology*, 82(6):1410–1411, 2016. DOI: 10.1111/bcp.13074.

[91]R. Slater, F. Moultrie, R. Bax, J. N. van den Anker, and A. Bhatt. Preterm health: time to bridge the evidence gap. *The Lancet*, 396(10255):872–873, 2020. DOI: 10.1016/S0140-6736(20)31977-2.

[92]European Medicines Agency. Workshop on regulatory and scientific issues related to the investigation of medicinal products intended for neonatal use (EMEA/484678/2006). In collaboration with Chairperson: Prof. John van den Anker. European Medicines Agency. 2006. URL: https://www.ema.europa.eu/en/documents/minutes/workshop-regulatory-scientific-issues-related-investigation-medicinal-productsintended-neonatal-use_en-0.pdf (visited on Aug. 11, 2022).

[93]FDA, FDA Division of Pediatrics and Maternal Health (DPMH), FDA Division of Anesthesiology, Addition Medicine, and Pain Medicine (DAAP), and University of Maryland CERSI. FDA-M CERSI: Analgesic Clinical Trial Designs, Extrapolation, and Endpoints in Patients from Birth to Less Than Two Years of Age Public Workshop - 10/13/2021 - 10/14/2021. U.S. Food and Drug Administration (FDA). 2021. URL: https://www.fda.gov/drugs/news-events-human-drugs/fda-m-cersi-analgesic-clinical-trial-designs-extrapolation-and-endpoints-patients-birth-less-two (visited on Aug. 11, 2022).

[94]R. B. Flint, F. van Beek, P. Andriessen, L. J. Zimmermann, K. D. Liem, I. K. Reiss, R. de Groot, D. Tibboel, D. M. Burger, S. H. Simons, and DINO Research group. Large differences in neonatal drug use between NICUs are common practice: time for consensus?: Large differences in drug use between NICUs. *British Journal of Clinical Pharmacology*, 84(6):1313–1323, 2018. DOI: 10.1111/bcp.13563.

[95]H. T. M. d. L. Costa, T. X. Costa, R. R. Martins, and A. G. Oliveira. Use of off-label and unlicensed medicines in neonatal intensive care. *PLOS ONE*, 13(9):e0204427, 2018. DOI: 10.1371/journal.pone.0204427.

[96]C. P. F. O'Donnell, R. J. Stone, and C. J. Morley. Unlicensed and off-label drug use in an Australian neonatal intensive care unit. *Pediatrics*, 110(5):e52, 2002. DOI: 10.1542/peds.110.5.e52.

[97]J. L. Lee, A. M. Redzuan, and N. M. Shah. Unlicensed and off-label use of medicines in children admitted to the intensive care units of a hospital in Malaysia. *International Journal of Clinical Pharmacy*, 35(6):1025–1029, 2013. DOI: 10.1007/s11096-013-9846-0.

[98]J. Lass, R. Käär, K. Jõgi, H. Varendi, T. Metsvaht, and I. Lutsar. Drug utilisation pattern and off-label use of medicines in Estonian neonatal units. *European Journal of Clinical Pharmacology*, 67(12):1263–1271, 2011. DOI: 10.1007/s00228-011-1072-x.

[99]A. Neubert, K. Lukas, T. Leis, H. Dormann, K. Brune, and W. Rascher. Drug utilisation on a preterm and neonatal intensive care unit in Germany: a prospective, cohort-based analysis. *European Journal of Clinical Pharmacology*, 66(1):87–95, 2010. DOI: 10.1007/s00228-009-0722-8.

[100]E. A. Kieran, N. O'Callaghan, and C. P. F. O'Donnell. Unlicensed and off-label drug use in an Irish neonatal intensive care unit: a prospective cohort study. *Acta Paediatrica (Oslo, Norway: 1992)*, 103(4):e139–142, 2014. DOI: 10.1111/apa.12541.

[101]A. Dessì, C. Salemi, V. Fanos, and L. Cuzzolin. Drug treatments in a neonatal setting: focus on the off-label use in the first month of life. *Pharmacy World & Science*, 32(2):120–124, 2010. DOI: 10.1007/s11096-009-9356-2.

[102]S. Chatterjee, A. Mandal, N. Lyle, S. Mukherjee, and A. K. Singh. Drug utilization study in a neonatology unit of a tertiary care hospital in eastern India. *Pharmacoepidemiology and Drug Safety*, 16(10):1141–1145, 2007. DOI: 10.1002/pds.1469.

[103]M. Eerdekens, C. Beuter, C. Lefeber, and J. N. van den Anker. The challenge of developing pain medications for children: therapeutic needs and future perspectives. *Journal of Pain Research*, Volume 12:1649–1664, 2019. DOI: 10.2147/JPR.S195788.

[104]R. Rao, J. S. Sampers, S. S. Kronsberg, J. V. Brown, N. S. Desai, and K. J. S. Anand. Neurobehavior of Preterm Infants at 36 Weeks Postconception as a Function of Morphine Analgesia. *American Journal of Perinatology*, 24(09):511–517, 2007. DOI: 10.1055/s-2007-986675.

[105]R. Carbajal, R. Lenclen, M. Jugie, A. Paupe, B. A. Barton, and K. J. S. Anand. Morphine Does Not Provide Adequate Analgesia for Acute Procedural Pain Among Preterm Neonates. *Pediatrics*, 115(6):1494–1500, 2005. DOI: 10.1542/peds.2004-1425.

[106]C. Hartley, F. Moultrie, A. Hoskin, G. Green, V. Monk, J. L. Bell, A. R. King, M. Buckle, M. van der Vaart, D. Gursul, S. Goksan, E. Juszczak, J. E. Norman, R. Rogers, C. Patel, E. Adams, and R. Slater. Analgesic efficacy and safety of morphine in the Procedural Pain in Premature Infants (Poppi) study: randomised placebo-controlled trial. *The Lancet*, 392(10164):2595–2605, 2018. DOI: 10.1016/S0140-6736(18)31813-0.

[107]K. Bauer, J. Ketteler, M. Hellwig, M. Laurenz, and H. Versmold. Oral glucose before venepuncture relieves neonates of pain, but stress is still evidenced by increase in oxygen consumption, energy expenditure, and heart rate. *Pediatric Research*, 55(4):695–700, 2004. DOI: 10.1203/01.PDR.0000113768.50419.CD.

[108]S. M. Walker. Long-term effects of neonatal pain. *Seminars in Fetal and Neonatal Medicine*. Pain Management In The Neonate, 24(4):101005, 2019. DOI: 10.1016/j.siny.2019.04.005.

[109]G. L. Kearns, S. M. Abdel-Rahman, S. W. Alander, D. L. Blowey, J. S. Leeder, and R. E. Kauffman. Developmental Pharmacology — Drug Disposition, Action, and Therapy in Infants and Children. *New England Journal of Medicine*, 349(12):1157–1167, 2003. DOI: 10.1056/NEJMra035092.

[110]N. J. Bouwmeester, J. N. van den Anker, W. C. J. Hop, K. J. S. Anand, and D. Tibboel. Age- and therapy-related effects on morphine requirements and plasma concentrations of morphine and its metabolites in postoperative infants. *BJA: British Journal of Anaesthesia*, 90(5):642–652, 2003. DOI: 10.1093/bja/aeg121.

[111]N. J. Bouwmeester, B. J. Anderson, D. Tibboel, and N. H. G. Holford. Developmental pharmacokinetics of morphine and its metabolites in neonates, infants and young children. *BJA: British Journal of Anaesthesia*, 92(2):208–217, 2004. DOI: 10.1093/bja/aeh042.

[112]M. Campbell-Yeo, M. Eriksson, and B. Benoit. Assessment and Management of Pain in Preterm Infants: A Practice Update. *Children (Basel, Switzerland)*, 9(2):244, 2022. DOI: 10.3390/children9020244.

[113]European Medicines Agency and Committee for Medicinal Products for Human Use (CHMP). Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population (EMEA/CHMP/EWP/147013/2004). European Medicines Agency. 2006. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-role-pharmacokinetics-development-medicinal-products-paediatric-population_en.pdf (visited on Aug. 24, 2022).

[114]European Medicines Agency and Committee for Medicinal Products for Human Use (CHMP). Concept paper on the revision of the Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population (EMA/CHMP/448599/2016). European Medicines Agency. 2017. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-revision-guideline-role-pharmacokinetics-development-medicinal-products-paediatric_en.pdf (visited on Aug. 24, 2022).

[115]K. Anand, J. V. Aranda, C. B. Berde, S. Buckman, E. V. Capparelli, W. A. Carlo, P. Hummel, J. Lantos, C. Celeste Johnston, V. Tutag Lehr, A. M. Lynn, L. G. Maxwell, T. F. Oberlander, T. N. Raju, S. G. Soriano, A. Taddio, and G. A. Walco. Analgesia and anesthesia for neonates: Study design and ethical issues. *Clinical Therapeutics*, 27(6):814–843, 2005. DOI: 10.1016/j.clinthera.2005.06.021.

[116]G. A. Walco, E. A. Kopecky, S. J. Weisman, J. Stinson, B. Stevens, P. J. Desjardins, C. B. Berde, E. J. Krane, K. J. Anand, M. Yaster, C. D. Dampier, R. H. Dworkin, I. Gilron, A. M. Lynn, L. G. Maxwell, S. Raja, B. Schachtel, and D. C. Turk. Clinical trial designs and models for analgesic medications for acute pain in neonates, infants, toddlers, children, and adolescents: ACTTION recommendations. *Pain*, 159(2):193–205, 2018. DOI: 10.1097/j.pain.0000000000001104.

[117]S. A. Cooper, P. J. Desjardins, D. C. Turk, R. H. Dworkin, N. P. Katz, H. Kehlet, J. C. Ballantyne, L. B. Burke, E. Carragee, P. Cowan, S. Croll, R. A. Dionne, J. T. Farrar, I. Gilron, D. B. Gordon, S. Iyengar, G. W. Jay, E. A. Kalso, R. D. Kerns, M. P. McDermott, S. N. Raja, B. A. Rappaport, C. Rauschkolb, M. A. Royal, M. Segerdahl, J. W. Stauffer, K. H. Todd, G. F. Vanhove, M. S. Wallace, C. West, R. E. White, and C. Wu. Research design considerations for single-dose analgesic clinical trials in acute pain: IMMPACT recommendations. *PAIN*, 157(2):288–301, 2016. DOI: 10.1097/j.pain.0000000000000375.

[118]K. J. S. Anand, R. W. Hall, N. Desai, B. Shephard, L. L. Bergqvist, T. E. Young, E. M. Boyle, R. Carbajal, V. K. Bhutani, M. B. Moore, S. S. Kronsberg, B. A. Barton, and NEOPAIN Trial Investigators Group. Effects of morphine analgesia in ventilated preterm neonates: primary outcomes from the NEOPAIN randomised trial. *Lancet*, 363(9422):1673–1682, 2004. DOI: 10.1016/S0140-6736(04)16251-X.

[119]S. H. P. Simons, M. van Dijk, R. A. van Lingen, D. Roofthooft, H. J. Duivenvoorden, N. Jongeneel, C. Bunkers, E. Smink, K. J. S. Anand, J. N. van den Anker, and D. Tibboel. Routine morphine infusion in preterm newborns who received ventilatory support: a randomized controlled trial. *JAMA*, 290(18):2419–2427, 2003. DOI: 10.1001/jama.290.18.2419.

[120]B. Lemyre, R. Sherlock, D. Hogan, I. Gaboury, C. Blanchard, and D. Moher. How effective is tetracaine 4% gel, before a peripherally inserted central catheter, in reducing procedural pain in infants: a randomized double-blind placebo controlled trial [ISRCTN75884221]. *BMC Medicine*, 4(1):11, 2006. DOI: 10.1186/1741-7015-4-11.

[121]M. van der Vaart, E. Duff, N. Raafat, R. Rogers, C. Hartley, and R. Slater. Multimodal pain assessment improves discrimination between noxious and non-noxious stimuli in infants. *Paediatric & Neonatal Pain*, 1(1):21–30, 2019. DOI: 10.1002/pne2.12007.

[122]M. M. Cobo, C. Hartley, D. Gursul, F. Andritsou, M. van der Vaart, G. Schmidt Mellado, L. Baxter, E. P. Duff, M. Buckle, R. Evans Fry, G. Green, A. Hoskin, R. Rogers, E. Adams, F. Moultrie, and R. Slater. Quantifying noxious-evoked baseline sensitivity in neonates to optimise analgesic trials. *eLife*, 10:e65266, 2021. DOI: 10.7554/eLife.65266.

[123]E. Olsson, H. Ahl, K. Bengtsson, D. N. Vejayaram, E. Norman, M. Bruschettini, and M. Eriksson. The use and reporting of neonatal pain scales: a systematic review of randomized trials. *PAIN*, 162(2):353–360, 2021. DOI: 10.1097/j.pain.0000000000002046.

[124]B. Stevens and S. Gibbins. Clinical utility and clinical significance in the assessment and management of pain in vulnerable infants. *Clinics in Perinatology*, 29(3):459–468, 2002. DOI: 10.1016/S0095-5108(02)00016-7.

[125]V. Shah, M. Ipp, J. Sam, T. Einarson, and A. Taddio. Eliciting the Minimal Clinically Important Difference in the Pain Response from Parents of Newborn Infants and Nurses. *Paediatrics & Child Health*, 9:44A, suppl_a, 2004. DOI: 10.1093/pch/9.suppl_a.44aa.

[126]S. M. Cepeda, J. M. Africano, R. Polo, R. Alcala, and D. B. Carr. What decline in pain intensity is meaningful to patients with acute pain? *PAIN*, 105(1):151–157, 2003. DOI: 10.1016/S0304-3959(03)00176-3.

[127]J. T. Farrar, J. P. J. Young, L. LaMoreaux, J. L. Werth, and M. R. Poole. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *PAIN*, 94(2):149–158, 2001. DOI: 10.1016/S0304-3959(01)00349-9.

[128]R. H. Dworkin, D. C. Turk, K. W. Wyrwich, D. Beaton, C. S. Cleeland, J. T. Farrar, J. A. Haythornthwaite, M. P. Jensen, R. D. Kerns, D. N. Ader, N. Brandenburg, L. B. Burke, D. Cella, J. Chandler, P. Cowan, R. Dimitrova, R. Dionne, S. Hertz, A. R. Jadad, N. P. Katz, H. Kehlet, L. D. Kramer, D. C. Manning, C. McCormick, M. P. McDermott, H. J. McQuay, S. Patel, L. Porter, S. Quessy, B. A. Rappaport, C. Rauschkolb, D. A. Revicki, M. Rothman, K. E. Schmader, B. R. Stacey, J. W. Stauffer, T. von Stein, R. E. White, J. Witter, and S. Zavisic. Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *The Journal of Pain*, 9(2):105–121, 2008. DOI: 10.1016/j.jpain.2007.09.005.

[129]V. S. Shah, A. Taddio, R. Hancock, P. Shah, and A. Ohlsson. Topical amethocaine gel 4% for intramuscular injection in term neonates: A double-blind, placebo-controlled, randomized trial. *Clinical Therapeutics*, 30(1):166–174, 2008. DOI: 10.1016/j.clinthera.2008.01.018.

[130]S. Gibbins, B. Stevens, E. Hodnett, J. Pinelli, A. Ohlsson, and G. Darlington. Efficacy and Safety of Sucrose for Procedural Pain Relief in Preterm and Term Neonates. *Nursing Research*, 51(6):375–382, 2002. DOI: 10.1097/00006199-200211000-00005.

[131]M. M. Cobo, F. Moultrie, A. G. V. Hauck, D. Crankshaw, V. Monk, C. Hartley, R. Evans Fry, S. Robinson, M. van der Vaart, L. Baxter, E. Adams, R. Poorun, A. Bhatt, and R. Slater. Multicentre, randomised controlled trial to investigate the effects of parental touch on relieving acute procedural pain in neonates (Petal). *BMJ Open*, 12(7):e061841, 2022. DOI: 10.1136/bmjopen-2022-061841.

[132]M. Kassab, A. Sheehy, M. King, C. Fowler, and M. Foureur. A double-blind randomised controlled trial of 25% oral glucose for pain relief in 2-month old infants undergoing immunisation. *International Journal of Nursing Studies*, 49(3):249–256, 2012. DOI: 10.1016/j.ijnurstu.2011.09.013.

[133]P. M. Lavoie, A. Stritzke, J. Ting, M. Jabr, A. Jain, E. Kwan, E. Chakkarapani, P. Brooks, R. Brant, P. J. McNamara, and L. Holsti. A Randomized Controlled Trial of the Use of Oral Glucose with or without Gentle Facilitated Tucking of Infants during Neonatal Echocardiography. *PLoS ONE*, 10(10):e0141015, 2015. DOI: 10.1371/journal.pone.0141015.

[134]C. V. Bellieni and C. C. Johnston. Analgesia, nil or placebo to babies, in trials that test new analgesic treatments for procedural pain. *Acta Paediatrica*, 105(2):129–136, 2016. DOI: 10.1111/apa.13210.

[135]L. F. Ross. Placebo-controlled analgesia research in healthy newborns is unethical. Comment on: C. V. Bellieni, A. Taddio, J. S. Linebarger, and J. D. Lantos. Should an IRB Approve a Placebo-Controlled Randomized Trial of Analgesia for Procedural Pain in Neonates? Pediatrics. 2012. URL: https://doi.org/10.1542/peds.2011-2910. Comment posted on Sep. 9, 2012.

[136]C. V. Bellieni, A. Taddio, J. S. Linebarger, and J. D. Lantos. Should an IRB Approve a Placebo-Controlled Randomized Trial of Analgesia for Procedural Pain in Neonates? *Pediatrics*, 130(3):550–553, 2012. DOI: 10.1542/peds.2011-2910.

[137]M.-C. Ou-Yang, I.-L. Chen, C.-C. Chen, M.-Y. Chung, F.-S. Chen, and H.-C. Huang. Expressed breast milk for procedural pain in preterm neonates: a randomized, double-blind, placebo-controlled trial. *Acta Paediatrica*, 102(1):15–21, 2013. DOI: 10.1111/apa.12045.

[138]J.-J. Liaw, L. Yang, C.-M. Lee, H.-C. Fan, Y.-C. Chang, and L.-P. Cheng. Effects of combined use of non-nutritive sucking, oral sucrose, and facilitated tucking on infant behavioural states across heel-stick procedures: A prospective, randomised controlled trial. *International Journal of Nursing Studies*, 50(7):883–894, 2013. DOI: 10.1016/j.ijnurstu.2012.08.021.

[139]R. J. Martin, K. Wang, Ö. Köroğlu, J. Di Fiore, and P. Kc. Intermittent Hypoxic Episodes in Preterm Infants: Do They Matter? *Neonatology*, 100(3):303–310, 2011. DOI: 10.1159/000329922.

[140]F. Moultrie, A. Shriver, C. Hartley, D. Wilkinson, A. K. Ewer, R. Rogers, E. Adams, and R. Slater. A universal right to pain relief: balancing the risks in a vulnerable patient population. *The Lancet Child & Adolescent Health*, 3(2):62–64, 2019. DOI: 10.1016/S2352-4642(18)30269-4.

[141]E. K. S. Louis, L. C. Frey, J. W. Britton, L. C. Frey, J. L. Hopp, P. Korb, M. Z. Koubeissi, W. E. Lievens, E. M. Pestana-Knight, and E. K. S. Louis. *Appendix 6. A Brief History of EEG*. American Epilepsy Society, 2016.

[142]J. R. Smith. The Electroencephalogram During Normal Infancy and Childhood: I. Rhythmic Activities Present in the Neonate and Their Subsequent Development. *The Pedagogical Seminary and Journal of Genetic Psychology*, 53(2):431–453, 1938. DOI: 10.1080/08856559.1938.10533820.

[143]D. Cohen. Magnetoencephalography: Evidence of Magnetic Fields Produced by Alpha-Rhythm Currents. *Science*, 161(3843):784–786, 1968. DOI: 10.1126/science.161.3843.784.

[144]J. E. Brazy, D. V. Lewis, M. H. Mitnick, and F. F. J. vander Vliet. Noninvasive Monitoring of Cerebral Oxygenation in Preterm Infants: Preliminary Observations. *Pediatrics*, 75(2):217–225, 1985. DOI: 10.1542/peds.75.2.217.

[145]M. Ferrari and V. Quaresima. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2):921–935, 2012. DOI: 10.1016/j.neuroimage.2012.03.049.

[146]A. Pollatou, C. A. Filippi, E. Aydin, K. Vaughn, D. Thompson, M. Korom, A. J. Dufford, B. Howell, L. Zöllei, A. D. Martino, A. Graham, D. Scheinost, and M. N. Spann. An ode to fetal, infant, and toddler neuroimaging: Chronicling early clinical to research applications with MRI, and an introduction to an academic society connecting the field. *Developmental Cognitive Neuroscience*, 54:101083, 2022. DOI: 10.1016/j.dcn.2022.101083.

[147]D. B. Lindsley. Heart and Brain Potentials of Human Fetuses in Utero. *The American Journal of Psychology*, 55(3):412–416, 1942. DOI: 10.2307/1417473.

[148]M.-D. Lamblin, M. André, M. J. Challamel, L. Curzi-Dascalova, A. M. d'Allest, E. De Giovanni, F. Moussalli-Salefranque, Y. Navelet, P. Plouin, M. F. Radvanyi-Bouvet, D. Samson-Dollfus, and M.-F. Vecchierini-Blineau. Électroencéphalographie du nouveau-né prématuré et à terme. Aspects maturatifs et glossaire. *Neurophysiologie Clinique/Clinical Neurophysiology*, 29(2):123–219, 1999. DOI: 10.1016/S0987-7053(99)80051-3.

[149]M. André, M.-D. Lamblin, A. M. d'Allest, L. Curzi-Dascalova, F. Moussalli-Salefranque, S. Nguyen The Tich, M.-F. Vecchierini-Blineau, F. Wallois, E. Walls-Esquivel, and P. Plouin. Electroencephalography in premature and full-term infants. Developmental features and glossary. *Neurophysiologie Clinique/Clinical Neurophysiology*, 40(2):59–124, 2010. DOI: 10.1016/j.neucli.2010.02.002.

[150]E. Bourel-Ponchel, S. Gueden, D. Hasaerts, C. Héberlé, G. Malfilâtre, L. Mony, P. Vignolo-Diard, and M.-D. Lamblin. Normal EEG during the neonatal period: maturational aspects from premature to full-term newborns. *Neurophysiologie Clinique*, 51(1):61–88, 2021. DOI: 10.1016/j.neucli.2020.10.004.

[151]L. V. Marcuse, M. C. Fields, and J. Yoo. The normal EEG from neonates to adolescents. In *Rowan's Primer of EEG*, pages 67–86. Elsevier, 2nd edition, 2016. DOI: 10.1016/B978-0-323-35387-8.00003-2.

[152]R. A. Hrachovy and E. M. Mizrahi. *Atlas of Neonatal Electroencephalography*. Springer Publishing Company, Incorporated, New York, United States, 2015.

[153]K. Pillay, A. Dereymaeker, K. Jansen, G. Naulaers, and M. De Vos. Applying a data-driven approach to quantify EEG maturational deviations in preterms with normal and abnormal neurodevelopmental outcomes. *Scientific Reports*, 10(1):7288, 2020. DOI: 10.1038/s41598-020-64211-0.

[154]P. Jezzard, P. M. Matthews, and S. M. Smith. *Functional MRI: An Introduction to Methods*. Oxford University Press, Oxford; New York, 2001.

[155]S. P. Fitzgibbon, S. J. Harrison, M. Jenkinson, L. Baxter, E. C. Robinson, M. Bastiani, J. Bozek, V. Karolis, L. Cordero Grande, A. N. Price, E. Hughes, A. Makropoulos, J. Passerat-Palmbach, A. Schuh, J. Gao, S.-R. Farahibozorg, J. O'Muircheartaigh, J. Ciarrusta, C. O'Keeffe, J. Brandon, T. Arichi, D. Rueckert, J. V. Hajnal, A. D. Edwards, S. M. Smith, E. Duff, and J. Andersson. The developing Human Connectome Project (dHCP) automated resting-state functional processing framework for newborn infants. *NeuroImage*, 223:117303, 2020. DOI: 10.1016/j.neuroimage.2020.117303.

[156]C. D. Smyser, T. E. Inder, J. S. Shimony, J. E. Hill, A. J. Degnan, A. Z. Snyder, and J. J. Neil. Longitudinal Analysis of Neural Network Development in Preterm Infants. *Cerebral Cortex*, 20(12):2852–2862, 2010. DOI: 10.1093/cercor/bhq035.

[157]E. Damaraju, A. Caprihan, J. R. Lowe, E. A. Allen, V. D. Calhoun, and J. P. Phillips. Functional connectivity in the developing brain: A longitudinal study from 4 to 9 months of age. *NeuroImage*, 84:169–180, 2014. DOI: 10.1016/j.neuroimage.2013.08.038.

[158]M. Eyre, S. P. Fitzgibbon, J. Ciarrusta, L. Cordero-Grande, A. N. Price, T. Poppe, A. Schuh, E. Hughes, C. O'Keeffe, J. Brandon, D. Cromb, K. Vecchiato, J. Andersson, E. P. Duff, S. J. Counsell, S. M. Smith, D. Rueckert, J. V. Hajnal, T. Arichi, J. O'Muircheartaigh, D. Batalle, and A. D. Edwards. The Developing Human Connectome Project: typical and disrupted perinatal functional connectivity. *Brain*, 144(7):2199–2213, 2021. DOI: 10.1093/brain/awab118.

[159]B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995. DOI: 10.1002/mrm.1910340409.

[160]P. Fransson, B. Skiöld, S. Horsch, A. Nordell, M. Blennow, H. Lagercrantz, and U. Åden. Resting-state networks in the infant brain. *Proceedings of the National Academy of Sciences*, 104(39):15531–15536, 2007. DOI: 10.1073/pnas.0704380104.

[161]W. Lin, Q. Zhu, W. Gao, Y. Chen, C.-H. Toh, M. Styner, G. Gerig, J. Smith, B. Biswal, and J. Gilmore. Functional Connectivity MR Imaging Reveals Cortical Functional Connectivity in the Developing Brain. *American Journal of Neuroradiology*, 29(10):1883–1889, 2008. DOI: 10.3174/ajnr.A1256.

[162]V. Doria, C. F. Beckmann, T. Arichi, N. Merchant, M. Groppo, F. E. Turkheimer, S. J. Counsell, M. Murgasova, P. Aljabar, R. G. Nunes, D. J. Larkman, G. Rees, and A. D. Edwards. Emergence of resting state networks in the preterm human brain. *Proceedings of the National Academy of Sciences*, 107(46):20015–20020, 2010. DOI: 10.1073/pnas.1007921107.

[163]M. L. Krishnan, L. E. Dyet, J. P. Boardman, O. Kapellou, J. M. Allsop, F. Cowan, A. D. Edwards, M. A. Rutherford, and S. J. Counsell. Relationship between white matter apparent diffusion coefficients in preterm infants at term-equivalent age and developmental outcome at 2 years. *Pediatrics*, 120(3):e604–609, 2007. DOI: 10.1542/peds.2006-3054.

[164]J. Dubois, G. Dehaene-Lambertz, S. Kulikova, C. Poupon, P. S. Hüppi, and L. Hertz-Pannier. The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. *Neuroscience*. Secrets of the CNS White Matter, 276:48–71, 2014. DOI: 10.1016/j.neuroscience.2013.12.044.

[165]S. Alcauter, W. Lin, J. K. Smith, S. J. Short, B. D. Goldman, J. S. Reznick, J. H. Gilmore, and W. Gao. Development of Thalamocortical Connectivity during Infancy and Its Cognitive Correlations. *The Journal of Neuroscience*, 34(27):9067–9075, 2014. DOI: 10.1523/JNEUROSCI.0796-14.2014.

[166]C. D. Smyser, A. Z. Snyder, J. S. Shimony, A. Mitra, T. E. Inder, and J. J. Neil. Resting-State Network Complexity and Magnitude Are Reduced in Prematurely Born Infants. *Cerebral Cortex*, 26(1):322–333, 2016. DOI: 10.1093/cercor/bhu251.

[167]H. Zhang, D. Shen, and W. Lin. Resting-state functional MRI studies on infant brains: A decade of gap-filling efforts. *NeuroImage*, 185:664–684, 2019. DOI: 10.1016/j.neuroimage.2018.07.004.

[168]C. R. L. Mongerson, R. W. Jennings, D. Borsook, L. Becerra, and D. Bajic. Resting-State Functional Connectivity in the Infant Brain: Methods, Pitfalls, and Potentiality. *Frontiers in Pediatrics*, 5, 2017. DOI: 10.3389/fped.2017.00159.

[169]W. Gao, S. Alcauter, A. Elton, C. R. Hernandez-Castillo, J. K. Smith, J. Ramirez, and W. Lin. Functional Network Development During the First Year: Relative Sequence and Socioeconomic Correlations. *Cerebral Cortex (New York, N.Y.: 1991)*, 25(9):2919–2928, 2015. DOI: 10.1093/cercor/bhu088.

[170]W. Gao, S. Alcauter, J. K. Smith, J. H. Gilmore, and W. Lin. Development of human brain cortical network architecture during infancy. *Brain Structure and Function*, 220(2):1173–1186, 2015. DOI: 10.1007/s00429-014-0710-3.

[171]T. Arichi, A. Moraux, A. Melendez, V. Doria, M. Groppo, N. Merchant, S. Combs, E. Burdet, D. J. Larkman, S. J. Counsell, C. F. Beckmann, and A. D. Edwards. Somatosensory cortical activation identified by functional MRI in preterm and term infants. *NeuroImage*, 49(3):2063–2071, 2010. DOI: 10.1016/j.neuroimage.2009.10.038.

[172]A. G. Allievi, A. Melendez-Calderon, T. Arichi, A. D. Edwards, and E. Burdet. An fMRI Compatible Wrist Robotic Interface to Study Brain Development in Neonates. *Annals of Biomedical Engineering*, 41(6):1181–1192, 2013. DOI: 10.1007/s10439-013-0782-x.

[173]A. G. Allievi, T. Arichi, N. Tusor, J. Kimpton, S. Arulkumaran, S. J. Counsell, A. D. Edwards, and E. Burdet. Maturation of Sensori-Motor Functional Responses in the Preterm Brain. *Cerebral Cortex (New York, NY)*, 26(1):402–413, 2016. DOI: 10.1093/cercor/bhv203.

[174]T. Mima, N. Sadato, S. Yazawa, T. Hanakawa, H. Fukuyama, Y. Yonekura, and H. Shibasaki. Brain structures related to active and passive finger movements in man. *Brain*, 122(10):1989–1997, 1999. DOI: 10.1093/brain/122.10.1989.

[175]L. Scheef, J. A. Nordmeyer-Massner, A. P. Smith-Collins, N. Müller, G. Stegmann-Woessner, J. Jankowski, J. Gieseke, M. Born, H. Seitz, P. Bartmann, H. H. Schild, K. P. Pruessmann, A. Heep, and H. Boecker. Functional Laterality of Task-Evoked Activation in Sensorimotor Cortex of Preterm Infants: An Optimized 3 T fMRI Study Employing a Customized Neonatal Head Coil. *PLoS ONE*, 12(1):e0169392, 2017. DOI: 10.1371/journal.pone.0169392.

[176]G. Williams, L. Fabrizi, J. Meek, D. Jackson, I. Tracey, N. Robertson, R. Slater, and M. Fitzgerald. Functional magnetic resonance imaging can be used to explore tactile and nociceptive processing in the infant brain. *Acta Paediatrica (Oslo, Norway : 1992)*, 104(2):158–166, 2015. DOI: 10.1111/apa.12848.

[177]S. Goksan, C. Hartley, F. Emery, N. Cockrill, R. Poorun, F. Moultrie, R. Rogers, J. Campbell, M. Sanders, E. Adams, S. Clare, M. Jenkinson, I. Tracey, and R. Slater. fMRI reveals neural activity overlap between adult and infant pain. *eLife*, 4:e06356, 2015. DOI: 10.7554/eLife.06356.

[178]E. P. Duff, F. Moultrie, M. van der Vaart, S. Goksan, A. Abos, S. P. Fitzgibbon, L. Baxter, T. D. Wager, and R. Slater. Inferring pain experience in infants using quantitative whole-brain functional MRI signatures: a cross-sectional, observational study. *The Lancet Digital Health*, 2(9):e458–e467, 2020. DOI: 10.1016/S2589-7500(20)30168-0.

[179]J. J. Tuulari, N. M. Scheinin, S. Lehtola, H. Merisaari, J. Saunavaara, R. Parkkola, I. Sehlstedt, L. Karlsson, H. Karlsson, and M. Björnsdotter. Neural correlates of gentle skin stroking in early infancy. *Developmental Cognitive Neuroscience*. Social Touch: A New Vista for Developmental Cognitive Neuroscience?, 35:36–41, 2019. DOI: 10.1016/j.dcn.2017.10.004.

[180]B. Deen, H. Richardson, D. D. Dilks, A. Takahashi, B. Keil, L. L. Wald, N. Kanwisher, and R. Saxe. Organization of high-level visual cortex in human infants. *Nature Communications*, 8(1):13995, 1, 2017. DOI: 10.1038/ncomms13995.

[181]G. Dehaene-Lambertz, A. Montavont, A. Jobert, L. Allirol, J. Dubois, L. Hertz-Pannier, and S. Dehaene. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, 114(2):53–65, 2010. DOI: 10.1016/j.bandl.2009.09.003.

[182]G. Dehaene-Lambertz, S. Dehaene, and L. Hertz-Pannier. Functional neuroimaging of speech perception in infants. *Science*, 298(5600):2013–2015, 2002. DOI: 10.1126/science.1077066.

[183]J. Dubois, P. Adibpour, C. Poupon, L. Hertz-Pannier, and G. Dehaene-Lambertz. MRI and M/EEG studies of the White Matter Development in Human Fetuses and Infants: Review and Opinion. *Brain Plasticity (Amsterdam, Netherlands)*, 2(1):49–69, 2016. DOI: 10.3233/BPL-160031.

[184]M. Jenkinson and M. Chappell. *Introduction to Neuroimaging Analysis*. Oxford Neuroimaging Primers. Oxford University Press, Oxford, UK, first edition edition, 2018.

[185]D. Le Bihan and H. Johansen-Berg. Diffusion MRI at 25: Exploring brain tissue structure and function. *NeuroImage*, 61(2):324–341, 2012. DOI: 10.1016/j.neuroimage.2011.11.006.

[186]D. Le Bihan and M. Iima. Diffusion Magnetic Resonance Imaging: What Water Tells Us about Biological Tissues. *PLoS Biology*, 13(7):e1002203, 2015. DOI: 10.1371/journal.pbio.1002203.

[187]Z. Eaton-Rosen, A. Melbourne, E. Orasanu, M. J. Cardoso, M. Modat, A. Bainbridge, G. S. Kendall, N. J. Robertson, N. Marlow, and S. Ourselin. Longitudinal measurement of the developing grey matter in preterm subjects using multi-modal MRI. *NeuroImage*, 111:580–589, 2015. DOI: 10.1016/j.neuroimage.2015.02.010.

[188]S. Kulikova, L. Hertz-Pannier, G. Dehaene-Lambertz, A. Buzmakov, C. Poupon, and J. Dubois. Multi-parametric evaluation of the white matter maturation. *Brain Structure and Function*, 220(6):3657–3672, 6, 2015. DOI: 10.1007/s00429-014-0881-y.

[189]P. S. Hüppi and J. Dubois. Diffusion tensor imaging of brain development. *Seminars in Fetal and Neonatal Medicine*. Assessing Brain Function in the Perinatal Period, 11(6):489–497, 2006. DOI: 10.1016/j.siny.2006.07.006.

[190]M. P. van den Heuvel, K. J. Kersbergen, M. A. de Reus, K. Keunen, R. S. Kahn, F. Groenendaal, L. S. de Vries, and M. J. Benders. The Neonatal Connectome During Preterm Brain Development. *Cerebral Cortex*, 25(9):3000–3013, 2015. DOI: 10.1093/cercor/bhu095.

[191]K. Keunen, S. J. Counsell, and M. J. N. L. Benders. The emergence of functional architecture during early brain development. *NeuroImage*. Functional Architecture of the Brain, 160:2–14, 2017. DOI: 10.1016/j.neuroimage.2017.01.047.

[192]G. Ball, P. Aljabar, S. Zebari, N. Tusor, T. Arichi, N. Merchant, E. C. Robinson, E. Ogundipe, D. Rueckert, A. D. Edwards, and S. J. Counsell. Rich-club organization of the newborn human brain. *Proceedings of the National Academy of Sciences*, 111(20):7456–7461, 2014. DOI: 10.1073/pnas.1324118111.

[193]C. J. Brown, S. P. Miller, B. G. Booth, S. Andrews, V. Chau, K. J. Poskitt, and G. Hamarneh. Structural network analysis of brain development in young preterm neonates. *NeuroImage*, 101:667–680, 2014. DOI: 10.1016/j.neuroimage.2014.07.030.

[194]T. Zhao, V. Mishra, T. Jeon, M. Ouyang, Q. Peng, L. Chalak, J. L. Wisnowski, R. Heyne, N. Rollins, N. Shu, and H. Huang. Structural network maturation of the preterm human brain. *NeuroImage*, 185:699–710, 2019. DOI: 10.1016/j.neuroimage.2018.06.047.

[195]P.-T. Yap, Y. Fan, Y. Chen, J. H. Gilmore, W. Lin, and D. Shen. Development Trends of White Matter Connectivity in the First Years of Life. *PLoS ONE*, 6(9):e24678, 2011. DOI: 10.1371/journal.pone.0024678.

[196]A. Pfefferbaum, D. H. Mathalon, E. V. Sullivan, J. M. Rawles, R. B. Zipursky, and K. O. Lim. A Quantitative Magnetic Resonance Imaging Study of Changes in Brain Morphology From Infancy to Late Adulthood. *Archives of Neurology*, 51(9):874–887, 1994. DOI: 10.1001/archneur.1994.00540210046012.

[197]M. Ouyang, J. Dubois, Q. Yu, P. Mukherjee, and H. Huang. Delineation of early brain development from fetuses to infants with diffusion MRI and beyond. *NeuroImage*, 185:836–850, 2019. DOI: 10.1016/j.neuroimage.2018.04.017.

[198]S. Wilson, M. Pietsch, L. Cordero-Grande, A. N. Price, J. Hutter, J. Xiao, L. McCabe, M. A. Rutherford, E. J. Hughes, S. J. Counsell, J.-D. Tournier, T. Arichi, J. V. Hajnal, A. D. Edwards, D. Christiaens, and J. O'Muircheartaigh. Development of human white matter pathways in utero over the second and third trimester. *Proceedings of the National Academy of Sciences*, 118(20), 2021. DOI: 10.1073/pnas.2023598118.

[199]E. Zanin, J.-P. Ranjeva, S. Confort-Gouny, M. Guye, D. Denis, P. J. Cozzone, and N. Girard. White matter maturation of normal human fetal brain. An in vivo diffusion tensor tractography study. *Brain and Behavior*, 1(2):95–108, 2011. DOI: 10.1002/brb3.17.

[200]J. F. Schneider, S. Confort-Gouny, Y. Le Fur, P. Viout, M. Bennathan, F. Chapon, C. Fogliarini, P. Cozzone, and N. Girard. Diffusion-weighted imaging in normal fetal brain maturation. *European Radiology*, 17(9):2422–2429, 2007. DOI: 10.1007/s00330-007-0634-x.

[201]S. A. Back, N. L. Luo, N. S. Borenstein, J. J. Volpe, and H. C. Kinney. Arrested Oligodendrocyte Lineage Progression During Human Cerebral White Matter Development: Dissociation Between the Timing of Progenitor Differentiation and Myelinogenesis. *Journal of Neuropathology & Experimental Neurology*, 61(2):197–211, 2002. DOI: 10.1093/jnen/61.2.197.

[202]M. J. Barkovich, Y. Li, R. S. Desikan, A. J. Barkovich, and D. Xu. Challenges in pediatric neuroimaging. *NeuroImage*, 185:793–801, 2019. DOI: 10.1016/j.neuroimage.2018.04.044.

[203]M. Korom, M. C. Camacho, C. A. Filippi, R. Licandro, L. A. Moore, A. Dufford, L. Zöllei, A. M. Graham, M. Spann, B. Howell, S. Shultz, and D. Scheinost. Dear reviewers: Responses to common reviewer critiques about infant neuroimaging studies. *Developmental Cognitive Neuroscience*, 53:101055, 2022. DOI: 10.1016/j.dcn.2021.101055.

[204]A. Copeland, E. Silver, R. Korja, S. J. Lehtola, H. Merisaari, E. Saukko, S. Sinisalo, J. Saunavaara, T. Lähdesmäki, R. Parkkola, S. Nolvi, L. Karlsson, H. Karlsson, and J. J. Tuulari. Infant and Child MRI: A Review of Scanning Procedures. *Frontiers in Neuroscience*, 15, 2021.

[205]J. Dubois, M. Alison, S. J. Counsell, L. Hertz-Pannier, P. S. Hüppi, and M. J. Benders. MRI of the Neonatal Brain: A Review of Methodological Challenges and Neuroscientific Advances. *Journal of Magnetic Resonance Imaging*, 53(5):1318–1343, 2021. DOI: 10.1002/jmri.27192.

[206]T. Arichi, K. Whitehead, G. Barone, R. Pressler, F. Padormo, A. D. Edwards, and L. Fabrizi. Localization of spontaneous bursting neuronal activity in the preterm human brain with simultaneous EEG-fMRI. *eLife*, 6:e27814, 2017. DOI: 10.7554/eLife.27814.

[207]E. J. Hughes, T. Winchman, F. Padormo, R. Teixeira, J. Wurie, M. Sharma, M. Fox, J. Hutter, L. Cordero-Grande, A. N. Price, J. Allsop, J. Bueno-Conde, N. Tusor, T. Arichi, A. D. Edwards, M. A. Rutherford, S. J. Counsell, and J. V. Hajnal. A dedicated neonatal brain imaging system. *Magnetic Resonance in Medicine*, 78(2):794–804, 2017. DOI: 10.1002/mrm.26462.

[208]R. Slater. Cortical Pain Responses in Human Infants. *Journal of Neuroscience*, 26(14):3662–3666, 2006. DOI: 10.1523/JNEUROSCI.0348-06.2006.

[209]M. Bartocci, L. L. Bergqvist, H. Lagercrantz, and K. J. Anand. Pain activates cortical areas in the preterm newborn brain. *PAIN*, 122(1):109–117, 2006. DOI: 10.1016/j.pain.2006.01.015.

[210]J. Chen, G. U. Gadi, A. Panigrahy, and E. W. Y. Tam. Using Neuroimaging to Study the Effects of Pain, Analgesia, and Anesthesia on Brain Development. *Journal of Neurosurgical Anesthesiology*, 31(1):119–121, 2019. DOI: 10.1097/ANA.0000000000000549.

[211]R. Slater, A. Worley, L. Fabrizi, S. Roberts, J. Meek, S. Boyd, and M. Fitzgerald. Evoked potentials generated by noxious stimulation in the human infant brain. *European Journal of Pain*, 14(3):321–326, 2010. DOI: 10.1016/j.ejpain.2009.05.005.

[212]L. Fabrizi, R. Slater, A. Worley, J. Meek, S. Boyd, S. Olhede, and M. Fitzgerald. A Shift in Sensory Processing that Enables the Developing Human Brain to Discriminate Touch from Pain. *Current Biology*, 21(18):1552–1558, 2011. DOI: 10.1016/j.cub.2011.08.010.

[213]G. Green, C. Hartley, A. Hoskin, E. Duff, A. Shriver, D. Wilkinson, E. Adams, R. Rogers, F. Moultrie, and R. Slater. Behavioural discrimination of noxious stimuli in infants is dependent on brain maturation. *Pain*, 160(2):493–500, 2019. DOI: 10.1097/j.pain.0000000000001425.

[214]N. L. Maitre, A. P. Key, O. D. Chorna, J. C. Slaughter, P. J. Matusz, M. T. Wallace, and M. M. Murray. The Dual Nature of Early-Life Experience on Somatosensory Processing in the Human Infant Brain. *Current Biology*, 27(7):1048–1054, 2017. DOI: 10.1016/j.cub.2017.02.036.

[215]D. Gursul, S. Goksan, C. Hartley, G. S. Mellado, F. Moultrie, A. Hoskin, E. Adams, G. Hathway, S. Walker, F. McGlone, and R. Slater. Stroking modulates noxious-evoked brain activity in human infants. *Current Biology*, 28(24):R1380–R1381, 2018. DOI: 10.1016/j.cub.2018.11.014.

[216]L. Jones, M. P. Laudiano-Dray, K. Whitehead, J. Meek, M. Fitzgerald, L. Fabrizi, and R. Pillai Riddell. The impact of parental contact upon cortical noxious-related activity in human neonates. *European Journal of Pain*, 25(1):149–159, 2021. DOI: 10.1002/ejp.1656.

[217]C. Hartley, F. Moultrie, D. Gursul, A. Hoskin, E. Adams, R. Rogers, and R. Slater. Changing Balance of Spinal Cord Excitability and Nociceptive Brain Activity in Early Human Development. *Current Biology*, 26(15):1998–2002, 2016. DOI: 10.1016/j.cub.2016.05.054.

[218]M. van der Vaart, C. Hartley, L. Baxter, G. S. Mellado, F. Andritsou, M. M. Cobo, R. E. Fry, E. Adams, S. Fitzgibbon, and R. Slater. Premature Infants Display Discriminable Behavioral, Physiological, and Brain Responses to Noxious and Nonnoxious Stimuli. *Cerebral Cortex*:bhab449, 2021. DOI: 10.1093/cercor/bhab449.

[219]C. Hartley, E. P. Duff, G. Green, G. S. Mellado, A. Worley, R. Rogers, and R. Slater. Nociceptive brain activity as a measure of analgesic efficacy in infants. *Science Translational Medicine*, 9(388):eaah6122, 2017. DOI: 10.1126/scitranslmed.aah6122.

[220]B. Benoit, A. Newman, R. Martin-Misener, M. Latimer, and M. Campbell-Yeo. The influence of breastfeeding on cortical and bio-behavioural indicators of procedural pain in newborns: Findings of a randomized controlled trial. *Early Human Development*, 154:105308, 2021. DOI: 10.1016/j.earlhumdev.2021.105308.

[221]M. Verriotis, L. Fabrizi, A. Lee, S. Ledwidge, J. Meek, and M. Fitzgerald. Cortical activity evoked by inoculation needle prick in infants up to one-year old. *PAIN*, 156(2):222–230, 2015. DOI: 10.1097/01.j.pain.0000460302.56325.0c.

[222]G. Green. *Measuring Pain in the Newborn Infant*, University of Oxford, 2018. [DPhil Thesis]. Available from: https://ora.ox.ac.uk/objects/uuid:5647e78c-48fb-4b1d-a54f-146803bd7037 (visited on Aug. 5, 2022).

[223]S. Kasser, C. Hartley, H. Rickenbacher, N. Klarer, A. Depoorter, A. N. Datta, M. M. Cobo, S. Goksan, A. Hoskin, W. Magerl, E. A. Huhn, G. Green, R. Slater, and S. Wellmann. Birth experience in newborn infants is associated with changes in nociceptive sensitivity. *Scientific Reports*, 9(1):4117, 1, 2019. DOI: 10.1038/s41598-019-40650-2.

[224]M. Campbell-Yeo, B. Benoit, A. Newman, C. Johnston, T. Bardouille, B. Stevens, and A. Jiang. The influence of skin-to-skin contact on Cortical Activity during Painful procedures in preterm infants in the neonatal intensive care unit (iCAP mini): study protocol for a randomized control trial. *Trials*, 23(1):512, 2022. DOI: 10.1186/s13063-022-06424-4.

[225]C. Hartley, S. Goksan, R. Poorun, K. Brotherhood, G. S. Mellado, F. Moultrie, R. Rogers, E. Adams, and R. Slater. The relationship between nociceptive brain activity, spinal reflex withdrawal and behaviour in newborn infants. *Scientific Reports*, 5:12519, 2015. DOI: 10.1038/srep12519.

[226]M. Verriotis, L. Jones, K. Whitehead, M. Laudiano-Dray, I. Panayotidis, H. Patel, J. Meek, L. Fabrizi, and M. Fitzgerald. The distribution of pain activity across the human neonatal brain is sex dependent. *NeuroImage*, 178:69–77, 2018. DOI: 10.1016/j.neuroimage.2018.05.030.

[227]L. Jones, L. Fabrizi, M. Laudiano-Dray, K. Whitehead, J. Meek, M. Verriotis, and M. Fitzgerald. Nociceptive Cortical Activity Is Dissociated from Nociceptive Behavior in Newborn Human Infants under Stress. *Current Biology*, 27(24):3846–3851.e3, 2017. DOI: 10.1016/j.cub.2017.10.063.

[228]G. Schmidt Mellado, K. Pillay, E. Adams, A. Alarcon, F. Andritsou, M. M. Cobo, R. Evans Fry, S. Fitzgibbon, F. Moultrie, L. Baxter, and R. Slater. The impact of premature extrauterine exposure on infants' stimulus-evoked brain activity across multiple sensory systems. *NeuroImage: Clinical*, 33:102914, 2022. DOI: 10.1016/j.nicl.2021.102914.

[229]T. D. Wager, L. Y. Atlas, M. A. Lindquist, M. Roy, C.-W. Woo, and E. Kross. An fMRI-Based Neurologic Signature of Physical Pain. *New England Journal of Medicine*, 368(15):1388–1397, 2013. DOI: 10.1056/NEJMoa1204471.

[230]S. Goksan, L. Baxter, F. Moultrie, E. Duff, G. Hathway, C. Hartley, I. Tracey, and R. Slater. The influence of the descending pain modulatory system on infant pain-related brain activity. *eLife*, 7:e37125, 2018. DOI: 10.7554/eLife.37125.

[231]L. Baxter, F. Moultrie, S. Fitzgibbon, M. Aspbury, R. Mansfield, M. Bastiani, R. Rogers, S. Jbabdi, E. Duff, and R. Slater. Functional and diffusion MRI reveal the neurophysiological basis of neonates' noxious-stimulus evoked brain activity. *Nature Communications*, 12(1):2744, 2021. DOI: 10.1038/s41467-021-22960-0.

[232]S. Goksan, C. Hartley, S. A. Hurley, A. M. Winkler, E. P. Duff, M. Jenkinson, R. Rogers, S. Clare, and R. Slater. Optimal echo time for functional MRI of the infant brain identified in response to noxious stimulation. *Magnetic Resonance in Medicine*, 78(2):625–631, 2017. DOI: 10.1002/mrm.26455.

[233]L. Baxter, S. Fitzgibbon, F. Moultrie, S. Goksan, M. Jenkinson, S. Smith, J. Andersson, E. Duff, and R. Slater. Optimising neonatal fMRI data analysis: Design and validation of an extended dHCP preprocessing pipeline to characterise noxious-evoked brain activity in infants. *NeuroImage*, 186:286–300, 2019. DOI: 10.1016/j.neuroimage.2018.11.006.

[234]R. R. Pillai Riddell, D. B. Flora, S. A. Stevens, B. Stevens, L. L. Cohen, S. Greenberg, and H. Garfield. Variability in infant acute pain responding meaningfully obscured by averaging pain responses. *PAIN*, 154(5):714–721, 2013. DOI: 10.1016/j.pain.2013.01.015.

[235]M. S. Salekin, P. R. Mouton, G. Zamzmi, R. Patel, D. Goldgof, M. Kneusel, S. L. Elkins, E. Murray, M. E. Coughlin, D. Maguire, T. Ho, and Y. Sun. Future roles of artificial intelligence in early pain management of newborns. *Paediatric and Neonatal Pain*, 3(3):134–145, 2021. DOI: 10.1002/pne2.12060.

[236]N. L. Maitre, A. R. Stark, C. C. M. Menser, O. D. Chorna, D. J. France, A. F. Key, K. Wilkens, M. Moore-Clingenpeel, D. M. Wilkes, and S. Bruehl. Cry presence and amplitude do not reflect cortical processing of painful stimuli in newborns with distinct responses to touch or cold. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 102(5):F428–F433, 2017. DOI: 10.1136/archdischild-2016-312279.

[237]N. Maimon, R. E. Grunau, I. L. Cepeda, M. Friger, L. Selnovik, S. Gilat, and E. Shany. Electroencephalographic Activity in Response to Procedural Pain in Preterm Infants Born at 28 and 33 Weeks Gestational Age. *The Clinical Journal of Pain*, 29(12):1044–1049, 2013. DOI: 10.1097/AJP.0b013e318284e525.

[238]E. P. Duff, W. Vennart, R. G. Wise, M. A. Howard, R. E. Harris, M. Lee, K. Wartolowska, V. Wanigasekera, F. J. Wilson, M. Whitlock, I. Tracey, M. W. Woolrich, and S. M. Smith. Learning to identify CNS drug action and efficacy using multistudy fMRI data. *Science Translational Medicine*, 7(274):274ra16–274ra16, 2015. DOI: 10.1126/scitranslmed.3008438.

[239]A. M. Graham, J. H. Pfeifer, P. A. Fisher, W. Lin, W. Gao, and D. A. Fair. The potential of infant fMRI research and the study of early life stress as a promising exemplar. *Developmental Cognitive Neuroscience*, 12:12–39, 2015. DOI: 10.1016/j.dcn.2014.09.005.

[240]L. Gui, S. Loukas, F. Lazeyras, P. S. Hüppi, D. E. Meskaldji, and C. Borradori Tolsa. Longitudinal study of neonatal brain tissue volumes in preterm infants and their ability to predict neurodevelopmental outcome. *NeuroImage*, 185:728–741, 2019. DOI: 10.1016/j.neuroimage.2018.06.034.

[241]Open Science Collaboration. An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science*, 7(6):657–660, 2012. DOI: 10.1177/1745691612462588.

[242]B. A. Nosek, J. R. Spies, and M. Motyl. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, 7(6):615–631, 2012. DOI: 10.1177/1745691612459058.

[243]E. B. Liem, C.-M. Lin, M.-I. Suleman, A. G. Doufas, R. G. Gregg, J. M. Veauthier, G. Loyd, and D. I. Sessler. Anesthetic Requirement Is Increased in Redheads. *Anesthesiology*, 101(2):279–283, 2004. DOI: 10.1097/00000542-200408000-00006.

[244]E. B. Liem, T. V. Joiner, K. Tsueda, and D. I. Sessler. Increased Sensitivity to Thermal Pain and Reduced Subcutaneous Lidocaine Efficacy in Redheads. *Anesthesiology*, 102(3):509–514, 2005. DOI: 10.1097/00000542-200503000-00006.

[245]J. A. Johnson. Ethnic Differences in Cardiovascular Drug Response: Potential Contribution of Pharmacogenetics. *Circulation*, 118(13):1383–1393, 2008. DOI: 10.1161/CIRCULATIONAHA.107.704023.

[246]J. Y. Shin, R. Martin, and J. Suls. Meta-analytic evaluation of gender differences and symptom measurement strategies in acute coronary syndromes. *Heart & Lung: The Journal of Critical Care*, 39(4):283–295, 2010 Jul-Aug. DOI: 10.1016/j.hrtlng.2009.10.010.

[247]P. J. Colvonen, P. N. DeYoung, N.-O. A. Bosompra, and R. L. Owens. Limiting racial disparities and bias for wearable devices in health science research. *Sleep*, 43(10):zsaa159, 2020. DOI: 10.1093/sleep/zsaa159.

[248]R. Courtland. Bias detectives: the researchers striving to make algorithms fair. *Nature*, 558(7710):357–360, 7710, 2018. DOI: 10.1038/d41586-018-05469-3.

[249]J. Angwin, J. Larson, S. Mattu, L. Kirchner, and ProPublica. Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. 2016. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=Tu5C70R2pCBv8Yj33AkMh2E-mHz3d6iu (visited on Aug. 18, 2022).

[250]Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019. DOI: 10.1126/science.aax2342.

[251]S. N. Goodman, D. Fanelli, and J. P. A. Ioannidis. What does research reproducibility mean? *Science Translational Medicine*, 8(341):341ps12, 2016. DOI: 10.1126/scitranslmed.aaf5027.

[252]B. A. Nosek and T. M. Errington. What is replication? *PLOS Biology*, 18(3):e3000691, 2020. DOI: 10.1371/journal.pbio.3000691.

[253]M. Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, 7604, 2016. DOI: 10.1038/533452a.

[254]D. Fanelli. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115(11):2628–2631, 2018. DOI: 10.1073/pnas.1708272114.

[255]B. A. Spellman. A Short (Personal) Future History of Revolution 2.0. *Perspectives on Psychological Science*, 10(6):886–899, 2015. DOI: 10.1177/1745691615609918.

[256]J. P. Simmons, L. D. Nelson, and U. Simonsohn. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11):1359–1366, 2011. DOI: 10.1177/0956797611417632.

[257]U. Simonsohn, L. D. Nelson, and J. P. Simmons. P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2):534–547, 2014. DOI: 10.1037/a0033242.

[258]N. L. Kerr. HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 2(3):196–217, 1998. DOI: 10.1207/s15327957pspr0203_4.

[259]J. P. A. Ioannidis. Why Most Discovered True Associations Are Inflated. *Epidemiology*, 19(5):640–648, 2008. DOI: 10.1097/EDE.0b013e31818131e7.

[260]K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, 5, 2013. DOI: 10.1038/nrn3475.

[261]D. Bishop. Rein in the four horsemen of irreproducibility. *Nature*, 568(7753):435–435, 7753, 2019. DOI: 10.1038/d41586-019-01307-2.

[262]A. Franco, N. Malhotra, and G. Simonovits. Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203):1502–1505, 2014. DOI: 10.1126/science.1255484.

[263]D. Fanelli. Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3):891–904, 2011. DOI: 10.1007/s11192-011-0494-7.

[264]M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. P. du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware, and J. P. A. Ioannidis. A manifesto for reproducible science. *Nature human behaviour*, 1:0021, 2017. DOI: 10.1038/s41562-016-0021.

[265]B. Fischhoff and R. Beyth. I knew it would happen. *Organizational Behavior and Human Performance*, 13(1):1–16, 1975. DOI: 10.1016/0030-5073(75)90002-1.

[266]J. J. J. Christensen-Szalanski and C. F. Willham. The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48(1):147–168, 1991. DOI: 10.1016/0749-5978(91)90010-Q.

[267]R. S. Nickerson. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2):175–220, 1998. DOI: 10.1037/1089-2680.2.2.175.

[268]B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor. The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11):2600–2606, 2018. DOI: 10.1073/pnas.1708274114.

[269]J. A. C. Sterne and G. D. Smith. Sifting the evidence—what's wrong with significance tests? *Physical Therapy*, 81(8):1464–1469, 2001. DOI: 10.1093/ptj/81.8.1464.

[270]J. P. A. Ioannidis. Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8):e124, 2005. DOI: 10.1371/journal.pmed.0020124.

[271]J. P. A. Ioannidis, R. Tarone, and J. K. McLaughlin. The False-positive to False-negative Ratio in Epidemiologic Studies. *Epidemiology*, 22(4):450–456, 2011. DOI: 10.1097/EDE.0b013e31821b506e.

[272]A. Tversky and D. Kahneman. Belief in the law of small numbers. *Psychological Bulletin*, 76(2):105–110, 1971. DOI: 10.1037/h0031322.

[273]L. D. Nelson, J. Simmons, and U. Simonsohn. Psychology's Renaissance. *Annual Review of Psychology*, 69(1):511–534, 2018. DOI: 10.1146/annurev-psych-122216-011836.

[274]J. P. Simmons, L. D. Nelson, and U. Simonsohn. False-Positive Citations. *Perspectives on Psychological Science*, 13(2):255–259, 2018. DOI: 10.1177/1745691617698146.

[275]J. IntHout, J. P. Ioannidis, and G. F. Borm. Obtaining evidence by a single well-powered trial or several modestly powered trials. *Statistical Methods in Medical Research*, 25(2):538–552, 2016. DOI: 10.1177/0962280212461098.

[276]P. T. Quinlan. Misuse of power: in defence of small-scale science. *Nature Reviews Neuroscience*, 14(8):585–585, 8, 2013. DOI: 10.1038/nrn3475-c1.

[277]P. Bacchetti, C. E. McCulloch, and M. R. Segal. Simple, Defensible Sample Sizes Based on Cost Efficiency. *Biometrics*, 64(2):577–585, 2008. DOI: 10.1111/j.1541-0420.2008.01004_1.x.

[278]K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò. Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience*, 14(8):585–585, 8, 2013. DOI: 10.1038/nrn3475-c4.

[279]C. D. Chambers. Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49(3):609–610, 2013. DOI: 10.1016/j.cortex.2012.12.016.

[280]N. Havron, C. Bergmann, and S. Tsuji. Preregistration in infant research—A primer. *Infancy*, 25(5):734–754, 2020. DOI: 10.1111/infa.12353.

[281]M. Aspbury, L. Baxter, M. M. Cobo, M. van der Vaart, R. Poorun, A. Bhatt, R. Slater, C. Hartley, K. Pillay, S. Fitzgibbon, and A. Hauck. Establishing a standardised approach for the measurement of neonatal noxious-evoked brain activity in response to an acute somatic nociceptive heel lance stimulus. *Cortex*, 2022. DOI: 10.17605/OSF.IO/ZY9MS.

[282]G. M. de Melo, A. L. P. d. A. Lélis, A. F. de Moura, M. V. L. M. L. Cardoso, and V. M. da Silva. Pain assessment scales in newborns: integrative review. *Revista Paulista de Pediatria*, 32(4):395–402, 2014. DOI: 10.1590/S0103-05822014000400017.

[283]B. S. Carter and J. Brunkhorst. Neonatal pain management. *Seminars in Perinatology*. Palliative Care in the Perinatal/Neonatal Period, 41(2):111–116, 2017. DOI: 10.1053/j.semperi.2016.11.001.

[284]R. Slater. A Randomised Controlled Trial to Investigate the Effects of Parental Touch on Relieving Acute Procedural Pain in Neonates. Clinical Trial Registration ClinicalTrials.gov Identifier: NCT04901611, ClinicalTrials.gov, 2021.

[285]B. Benoit, R. Martin-Misener, A. Newman, M. Latimer, and M. Campbell-Yeo. Neurophysiological assessment of acute pain in infants: a scoping review of research methods. *Acta Paediatrica*, 106(7):1053–1066, 2017. DOI: 10.1111/apa.13839.

[286]L. Jones, M. P. Laudiano-Dray, K. Whitehead, M. Verriotis, J. Meek, M. Fitzgerald, and L. Fabrizi. EEG, behavioural and physiological recordings following a painful procedure in human neonates. *Scientific Data*, 5(1):180248, 1, 2018. DOI: 10.1038/sdata.2018.248.

[287]M. K. Jones, K. Tsugawa, A. S. Tarnawski, and D. Baatar. Dual actions of nitric oxide on angiogenesis: possible roles of PKC, ERK, and AP-1. *Biochemical and Biophysical Research Communications*, 318(2):520–528, 2004. DOI: 10.1016/j.bbrc.2004.04.055.

[288]A. Worley, L. Fabrizi, S. Boyd, and R. Slater. Multi-modal pain measurements in infants. *Journal of Neuroscience Methods*, 205(2):252–257, 2012. DOI: 10.1016/j.jneumeth.2012.01.009.

[289]A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004. DOI: 10.1016/j.jneumeth.2003.10.009.

[290]F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, 2011:e879716, 2011. DOI: 10.1155/2011/879716.

[291]E. Maris and R. Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, 2007. DOI: 10.1016/j.jneumeth.2007.03.024.

[292]A. Canty and B. Ripley. Boot: Bootstrap Functions, version 1.3-28, 2021. Available from: https://CRAN.R-project.org/package=boot (visited on Dec. 9, 2021).

[293]A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge ; New York, NY, USA, 1997.

[294]B. Derrick, A. Dobson-Mckittrick, D. Toher, and P. White. Test statistics for comparing two proportions with partially overlapping samples. *Journal of Applied Quantitative Methods*, 10(3):14, 2015.

[295]B. Derrick, B. Russ, D. Toher, and P. White. Test statistics for the comparison of means for two samples that include both paired and independent observations. *Journal of Modern Applied Statistical Methods*, 16(1):137–157, 2017. DOI: 10.22237/jmasm/1493597280.

[296]B. Derrick, D. Toher, and P. White. How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017). *The Quantitative Methods for Psychology*, 13(2):120–126, 2017. DOI: 10.20982/tqmp.13.2.p120.

[297]R. Slater, L. Fabrizi, A. Worley, J. Meek, S. Boyd, and M. Fitzgerald. Premature infants display increased noxious-evoked neuronal activity in the brain compared to healthy age-matched term-born infants. *NeuroImage*, 52(2):583–589, 2010. DOI: 10.1016/j.neuroimage.2010.04.253.

[298]D. Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4:863, 2013. DOI: 10.3389/fpsyg.2013.00863.

[299]F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4):1149–1160, 2009. DOI: 10.3758/BRM.41.4.1149.

[300]Center for Drug Evaluation and Research, FDA. Guidance for Industry: Statistical Approaches to Establishing Bioequivalence (01D-0027), U.S. Food and Drug Administration (FDA), 2001.

[301]B. M. Davit, P. E. Nwakama, G. J. Buehler, D. P. Conner, S. H. Haidar, D. T. Patel, Y. Yang, L. X. Yu, and J. Woodcock. Comparing Generic and Innovator Drugs: A Review of 12 Years of Bioequivalence Data from the United States Food and Drug Administration. *Annals of Pharmacotherapy*, 43(10):1583–1597, 2009. DOI: 10.1345/aph.1M141.

[302]D. J. Schuirmann. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15(6):657–680, 1987. DOI: 10.1007/BF01068419.

[303]K. Kelley. Confidence Intervals for Standardized Effect Sizes: Theory, Application, and Implementation. *Journal of Statistical Software*, 20(8), 2007. DOI: 10.18637/jss.v020.i08.

[304]K. Kelley. Methods for the Behavioral, Educational, and Social Sciences: An R package. *Behavior Research Methods*, 39(4):979–984, 2007. DOI: 10.3758/BF03192993.

[305]K. Kelley. MBESS: The MBESS R Package, version 4.8.1, 2021. Available from: https://CRAN.R-project.org/package=MBESS (visited on Dec. 9, 2021).

[306]K. D. Craig, M. F. Whitfield, R. V. E. Grunau, J. Linton, and H. D. Hadjistavropoulos. Pain in the preterm neonate: behavioural and physiological indices. *Pain*, 52(3):287–299, 1993. DOI: 10.1016/0304-3959(93)90162-I.

[307]C. Davatzikos. Machine learning in neuroimaging: Progress and challenges. *NeuroImage*, 197:652–656, 2019. DOI: 10.1016/j.neuroimage.2018.10.003.

[308]J. Lee, I. Mawla, J. Kim, M. L. Loggia, A. Ortiz, C. Jung, S.-T. Chan, J. Gerber, V. J. Schmithorst, R. R. Edwards, A. D. Wasan, C. Berna, J. Kong, T. J. Kaptchuk, R. L. Gollub, B. R. Rosen, and V. Napadow. Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. *Pain*, 160(3):550–560, 2019. DOI: 10.1097/j.pain.0000000000001417.

[309]A. N. Santana, I. Cifre, C. N. de Santana, and P. Montoya. Using Deep Learning and Resting-State fMRI to Classify Chronic Pain Conditions. *Frontiers in Neuroscience*, 13, 2019. DOI: 10.3389/fnins.2019.01313.

[310]J. Boissoneault, L. Sevel, J. Letzen, M. Robinson, and R. Staud. Biomarkers for Musculoskeletal Pain Conditions: Use of Brain Imaging and Machine Learning. *Current rheumatology reports*, 19(1):5, 2017. DOI: 10.1007/s11926-017-0629-9.

[311]S. Brahnam, C.-F. Chuang, F. Y. Shih, and M. R. Slack. Machine recognition and representation of neonatal facial displays of acute pain. *Artificial Intelligence in Medicine*, 36(3):211–222, 2006. DOI: 10.1016/j.artmed.2004.12.003.

[312]G. Zamzmi, C.-Y. Pai, D. Goldgof, R. Kasturi, T. Ashmeade, and Y. Sun. An approach for automated multimodal analysis of infants' pain. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016 23rd International Conference on Pattern Recognition (ICPR), pages 4148–4153, 2016. DOI: 10.1109/ICPR.2016.7900284.

[313]I. Tavor, O. P. Jones, R. B. Mars, S. M. Smith, T. E. Behrens, and S. Jbabdi. Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352(6282):216–220, 2016. DOI: 10.1126/science.aad8127.

[314]C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1001–1013, 2005. DOI: 10.1098/rstb.2005.1634.

[315]B. A. Seitzman, A. Z. Snyder, E. C. Leuthardt, and J. S. Shimony. The State of Resting State Networks. *Topics in magnetic resonance imaging : TMRI*, 28(4):189–196, 2019. DOI: 10.1097/RMR.0000000000000214.

[316]S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31):13040–13045, 2009. DOI: 10.1073/pnas.0905267106.

[317]O. Parker Jones, N. Voets, J. Adcock, R. Stacey, and S. Jbabdi. Resting connectivity predicts task activation in pre-surgical populations. *NeuroImage : Clinical*, 13:378–385, 2016. DOI: 10.1016/j.nicl.2016.12.028.

[318]T. Spisak, B. Kincses, F. Schlitt, M. Zunhammer, T. Schmidt-Wilcke, Z. T. Kincses, and U. Bingel. Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. *Nature Communications*, 11(1):187, 1, 2020. DOI: 10.1038/s41467-019-13785-z.

[319]E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671, 11, 2015. DOI: 10.1038/nn.4135.

[320]R. A. Poldrack, G. Huckins, and G. Varoquaux. Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, 77(5):534–540, 2020. DOI: 10.1001/jamapsychiatry.2019.3671.

[321]M. Hadders-Algra and H. Forssberg. Development of motor functions in health and disease. In *The Newborn Brain: Neuroscience and Clinical Applications*, pages 345–360. Cambridge University Press, Cambridge, 2nd edition, 2010. DOI: 10.1017/CBO9780511711848.023.

[322]S. Moeller, E. Yacoub, C. A. Olman, E. Auerbach, J. Strupp, N. Harel, and K. Uğurbil. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic resonance in medicine*, 63(5):1144–1153, 2010. DOI: 10.1002/mrm.22361.

[323]J. Xu, S. Moeller, E. J. Auerbach, J. Strupp, S. M. Smith, D. A. Feinberg, E. Yacoub, and K. Uğurbil. Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *NeuroImage*, 83:991–1001, 2013.

[324]S. J. Harrison, M. W. Woolrich, E. C. Robinson, M. F. Glasser, C. F. Beckmann, M. Jenkinson, and S. M. Smith. Large-scale Probabilistic Functional Modes from resting state fMRI. *NeuroImage*, 109:217–231, 2015. DOI: 10.1016/j.neuroimage.2015.01.013.

[325]T. Arichi, G. Fagiolo, M. Varela, A. Melendez-Calderon, A. Allievi, N. Merchant, N. Tusor, S. J. Counsell, E. Burdet, C. F. Beckmann, and A. D. Edwards. Development of BOLD signal hemodynamic responses in the human brain. *NeuroImage*, 63(2):663–673, 2012. DOI: 10.1016/j.neuroimage.2012.06.054.

[326]S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219, 2004. DOI: 10.1016/j.neuroimage.2004.07.051.

[327]M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith. FSL. *NeuroImage*, 62(2):782–790, 2012. DOI: 10.1016/j.neuroimage.2011.09.015.

[328]L. Snoek, S. Miletić, and H. S. Scholte. How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, 184:741–760, 2019. DOI: 10.1016/j.neuroimage.2018.09.074.

[329]M. F. Glasser, T. S. Coalson, J. D. Bijsterbosch, S. J. Harrison, M. P. Harms, A. Anticevic, D. C. Van Essen, and S. M. Smith. Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. *NeuroImage*, 181:692–717, 2018. DOI: 10.1016/j.neuroimage.2018.04.076.

[330]A. Eklund, H. Knutsson, and T. E. Nichols. Cluster failure revisited: Impact of first level design and physiological noise on cluster false positive rates. *Human Brain Mapping*, 40(7):2017–2032, 2019. DOI: 10.1002/hbm.24350.

[331]A. Schuh, A. Makropoulos, E. C. Robinson, L. Cordero-Grande, E. Hughes, J. Hutter, A. N. Price, M. Murgasova, R. P. A. G. Teixeira, N. Tusor, J. K. Steinweg, S. Victor, M. A. Rutherford, J. V. Hajnal, A. D. Edwards, and D. Rueckert. Unbiased construction of a temporally consistent morphological atlas of neonatal brain development. *bioRxiv*:251512, 2018. DOI: 10.1101/251512.

[332]A. P. Holmes, R. C. Blair, J. D. Watson, and I. Ford. Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, 16(1):7–22, 1996. DOI: 10.1097/00004647-199601000-00002.

[333]A. V. Apkarian, M. C. Bushnell, R.-D. Treede, and J.-K. Zubieta. Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain*, 9(4):463–463, 2005. DOI: 10.1016/j.ejpain.2004.11.001.

[334]L. Baxter. *Neuroimaging of Neonatal Nociception*, University of Oxford, 2020. [DPhil Thesis]. Available from: https://ora.ox.ac.uk/objects/uuid:d708d937-e82e-44e1-b519-d03bcb619855 (visited on Aug. 29, 2022).

[335]L. J. Chang, P. J. Gianaros, S. B. Manuck, A. Krishnan, and T. D. Wager. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLOS Biology*, 13(6):1–28, 2015. DOI: 10.1371/journal.pbio.1002180.

[336]H. Eisenbarth, L. J. Chang, and T. D. Wager. Multivariate Brain Prediction of Heart Rate and Skin Conductance Responses to Social Threat. *Journal of Neuroscience*, 36(47):11987–11998, 2016. DOI: 10.1523/JNEUROSCI.3672-15.2016.

[337]A. Krishnan, C.-W. Woo, L. J. Chang, L. Ruzic, X. Gu, M. López-Solà, P. L. Jackson, J. Pujol, J. Fan, and T. D. Wager. Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife*, 5:e15166, 2016. DOI: 10.7554/eLife.15166.

[338]Z. Cui and G. Gong. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*, 178:622–637, 2018. DOI: 10.1016/j.neuroimage.2018.06.001.

[339]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[340]L. Snoek, S. Miletic, and H. S. Scholte. GitHub Repositry: How to control for confounds in decoding analyses of neuroimaging data ("MultiVoxel Confound Analysis", MVCA), GitHub, 2018. Available from: https://github.com/lukassnoek/MVCA (visited on July 1, 2019).

[341]T. O. Kvalseth. Cautionary Note about R2. *The American Statistician*, 39(4):279–285, 1985. DOI: 10.2307/2683704.

[342]M. J. Rosa and B. Seymour. Decoding the matrix: Benefits and limitations of applying machine learning algorithms to pain neuroimaging. *PAIN*, 155(5):864–867, 2014. DOI: 10.1016/j.pain.2014.02.013.

[343]C. S. Nielsen, R. Staud, and D. D. Price. Individual Differences in Pain Sensitivity: Measurement, Causation, and Consequences. *The Journal of Pain*, 10(3):231–237, 2009. DOI: 10.1016/j.jpain.2008.09.010.

[344]D. Batalle, A. D. Edwards, and J. O'Muircheartaigh. Annual Research Review: Not just a small adult brain: understanding later neurodevelopment through imaging the neonatal brain. *Journal of Child Psychology and Psychiatry*, 59(4):350–371, 2018. DOI: 10.1111/jcpp.12838.

[345]F. A. Lenz. The human pain system : experimental and clinical perspectives [electronic resource]. In *The Human Pain System : Experimental and Clinical Perspectives*, Cambridge Core. Cambridge, 2010.

[346]K. Wiech. Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science (American Association for the Advancement of Science)*, 354(6312):584–587, 2016. DOI: 10.1126/science.aaf8934.

[347]J. F. de Kieviet, L. Zoetebier, R. M. van Elburg, R. J. Vermeulen, and J. Oosterlaan. Brain development of very preterm and very low-birthweight children in childhood and adolescence: a meta-analysis: Review. *Developmental Medicine & Child Neurology*, 54(4):313–323, 2012. DOI: 10.1111/j.1469-8749.2011.04216.x.

[348]S. Tsuneishi and P. Casaer. Effects of preterm extrauterine visual experience on the development of the human visual system: a flash VEP study. *Developmental Medicine & Child Neurology*, 42(10):663–668, 2000. DOI: 10.1111/j.1469-8749.2000.tb00676.x.

[349]G. Jandó, E. Mikó-Baráth, K. Markó, K. Hollódy, B. Török, and I. Kovacs. Early-onset binocularity in preterm infants reveals experience-dependent visual development in humans. *Proceedings of the National Academy of Sciences*, 109(27):11049–11052, 2012. DOI: 10.1073/pnas.1203096109.

[350]S. Johnson and N. Marlow. Early and long-term outcome of infants born extremely preterm. *Archives of Disease in Childhood*, 102(1):97–102, 2017. DOI: 10.1136/archdischild-2015-309581.

[351]N. Botting, A. Powls, R. W. Cooke, and N. Marlow. Cognitive and educational outcome of very-low-birthweight children in early adolescence. *Developmental Medicine & Child Neurology*, 40(10):652–660, 1998. DOI: 10.1111/j.1469-8749.1998.tb12324.x.

[352]S. J. Counsell, A. D. Edwards, A. T. M. Chew, M. Anjari, L. E. Dyet, L. Srinivasan, J. P. Boardman, J. M. Allsop, J. V. Hajnal, M. A. Rutherford, and F. M. Cowan. Specific relations between neurodevelopmental abilities and white matter microstructure in children born preterm. *Brain*, 131(12):3201–3208, 2008. DOI: 10.1093/brain/awn268.

[353]J. M. Lorenz, D. E. Wooliever, J. R. Jetton, and N. Paneth. A Quantitative Review of Mortality and Developmental Disability in Extremely Premature Newborns. *Archives of Pediatrics & Adolescent Medicine*, 152(5):425–435, 1998. DOI: 10.1001/archpedi.152.5.425.

[354]N. A. Parikh. Advanced neuroimaging and its role in predicting neurodevelopmental outcomes in very preterm infants. *Seminars in Perinatology*, 40(8):530–541, 2016. DOI: 10.1053/j.semperi.2016.09.005.

[355]C. Nosarti, A. Reichenberg, R. M. Murray, S. Cnattingius, M. P. Lambe, L. Yin, J. MacCabe, L. Rifkin, and C. M. Hultman. Preterm Birth and Psychiatric Disorders in Young Adult Life. *Archives of General Psychiatry*, 69(6):610–617, 2012. DOI: 10.1001/archgenpsychiatry.2011.1374.

[356]D. Buskila, L. Neumann, E. Zmora, M. Feldman, A. Bolotin, and J. Press. Pain Sensitivity in Prematurely Born Adolescents. *Archives of Pediatrics & Adolescent Medicine*, 157(11):1079–1082, 2003. DOI: 10.1001/archpedi.157.11.1079.

[357]G. Ball, L. Pazderova, A. Chew, N. Tusor, N. Merchant, T. Arichi, J. M. Allsop, F. M. Cowan, A. D. Edwards, and S. J. Counsell. Thalamocortical Connectivity Predicts Cognition in Children Born Preterm. *Cerebral Cortex*, 25(11):4310–4318, 2015. DOI: 10.1093/cercor/bhu331.

[358]D. M. Romeo, C. Brogna, F. Sini, M. G. Romeo, F. Cota, and D. Ricci. Early psychomotor development of low-risk preterm infants: Influence of gestational age and gender. *European Journal of Paediatric Neurology*, 20(4):518–523, 2016. DOI: 10.1016/j.ejpn.2016.04.011.

[359]J. E. Kline, V. S. P. Illapani, L. He, M. Altaye, J. W. Logan, and N. A. Parikh. Early cortical maturation predicts neurodevelopment in very preterm infants. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 105(5):460–465, 2020. DOI: 10.1136/archdischild-2019-317466.

[360]P. Moeskops, I. Išgum, K. Keunen, N. H. P. Claessens, I. C. van Haastert, F. Groenendaal, L. S. de Vries, M. A. Viergever, and M. J. N. L. Benders. Prediction of cognitive and motor outcome of preterm infants based on automatic quantitative descriptors from neonatal MR brain images. *Scientific Reports*, 7(1):2163, 1, 2017. DOI: 10.1038/s41598-017-02307-w.

[361]B. Böhm, M. Katz-Salamon, A.-C. Smedler, H. Lagercrantz, and H. Forssberg. Developmental risks and protective factors for influencing cognitive outcome at 5½ years of age in very-low-birthweight children. *Developmental Medicine & Child Neurology*, 44(8):508–516, 2007. DOI: 10.1111/j.1469-8749.2002.tb00321.x.

[362]M. B. Resnick, K. Stralka, R. L. Carter, M. Ariet, R. L. Bucciarelli, R. R. Furlough, J. H. Evans, J. S. Curran, and W. W. Ausbon. Effects of birth weight and sociodemographic variables on mental development of neonatal intensive care unit survivors. *American Journal of Obstetrics and Gynecology*, 162(2):374–378, 1990. DOI: 10.1016/0002-9378(90)90389-O.

[363]J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017. DOI: 10.1016/j.neuroimage.2016.09.046.

[364]T. E. Inder, S. K. Warfield, H. Wang, P. S. Hüppi, and J. J. Volpe. Abnormal Cerebral Structure Is Present at Term in Premature Infants. *Pediatrics*, 115(2):286–294, 2005. DOI: 10.1542/peds.2004-0326.

[365]J. P. Boardman and S. J. Counsell. Invited Review: Factors associated with atypical brain development in preterm infants: insights from magnetic resonance imaging. *Neuropathology and Applied Neurobiology*, 46(5):413–421, 2020. DOI: 10.1111/nan.12589.

[366]M. Bastiani, J. L. Andersson, L. Cordero-Grande, M. Murgasova, J. Hutter, A. N. Price, A. Makropoulos, S. P. Fitzgibbon, E. Hughes, D. Rueckert, S. Victor, M. Rutherford, A. D. Edwards, S. M. Smith, J.-D. Tournier, J. V. Hajnal, S. Jbabdi, and S. N. Sotiropoulos. Automated processing pipeline for neonatal diffusion MRI in the developing Human Connectome Project. *NeuroImage*, 185:750–763, 2019. DOI: 10.1016/j.neuroimage.2018.05.064.

[367]N. Bayley. *Bayley Scales of Infant and Toddler Development–Third Edition: Technical Manual*. Harcourt Assessment, Inc., 2006.

[368]E. M. Lennon, J. M. Gardner, B. Z. Karmel, and M. J. Flory. Bayley Scales of Infant Development. In *Encyclopedia of Infant and Early Childhood Development*. Pages 140–146. Elsevier, 2nd edition, 2020. DOI: 10.1016/b978-0-12-809324-5.23376-2.

[369]L. G. Weiss, T. Oakland, and G. P. Aylward. *Bayley-III Clinical Use and Interpretation*. Elsevier Science, Burlington, 2010.

[370]E. E. Rogers and S. R. Hintz. Early neurodevelopmental outcomes of extremely preterm infants. *Seminars in Perinatology*. Outcomes of High-Risk Infants, 40(8):497–509, 2016. DOI: 10.1053/j.semperi.2016.09.002.

[371]S. Johnson, T. Moore, and N. Marlow. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatric Research*, 75(5):670–674, 5, 2014. DOI: 10.1038/pr.2014.10.

[372]B. R. Vohr, B. E. Stephens, R. D. Higgins, C. M. Bann, S. R. Hintz, A. Das, J. E. Newman, M. Peralta-Carcelen, K. Yolton, A. M. Dusick, P. W. Evans, R. F. Goldstein, R. A. Ehrenkranz, A. Pappas, I. Adams-Chapman, D. E. Wilson-Costello, C. R. Bauer, A. Bodnar, R. J. Heyne, Y. E. Vaucher, R. G. Dillard, M. J. Acarregui, E. C. McGowan, G. J. Myers, and J. Fuller. Are Outcomes of Extremely Preterm Infants Improving? Impact of Bayley Assessment on Outcomes. *The Journal of Pediatrics*, 161(2):222–228.e3, 2012. DOI: 10.1016/j.jpeds.2012.01.057.

[373]C.-Y. Lin, C.-H. Hsu, and J.-H. Chang. Neurodevelopmental outcomes at 2 and 5 years of age in very-low-birth-weight preterm infants born between 2002 and 2009: A prospective cohort study in Taiwan. *Pediatrics & Neonatology*, 61(1):36–44, 2020. DOI: 10.1016/j.pedneo.2019.05.006.

[374]Y.-T. Yu, W.-S. Hsieh, C.-H. Hsu, L.-C. Chen, W.-T. Lee, N.-C. Chiu, Y.-C. Wu, and S.-F. Jeng. A psychometric study of the Bayley Scales of Infant and Toddler Development – 3rd Edition for term and preterm Taiwanese infants. *Research in Developmental Disabilities*, 34(11):3875–3883, 2013. DOI: 10.1016/j.ridd.2013.07.006.

[375]P. J. Anderson, C. R. De Luca, E. Hutchinson, G. Roberts, L. W. Doyle, and the Victorian Infant Collaborative Group. Underestimation of Developmental Delay by the New Bayley-III Scale. *Archives of Pediatrics & Adolescent Medicine*, 164(4):352–356, 2010. DOI: 10.1001/archpediatrics.2010.20.

[376]M. M. Spencer-Smith, A. J. Spittle, K. J. Lee, L. W. Doyle, and P. J. Anderson. Bayley-III Cognitive and Language Scales in Preterm Children. *Pediatrics*, 135(5):e1258–e1265, 2015. DOI: 10.1542/peds.2014-3039.

[377]A. J. Spittle, M. M. Spencer-Smith, A. L. Eeles, K. J. Lee, L. E. Lorefice, P. J. Anderson, and L. W. Doyle. Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? *Developmental Medicine & Child Neurology*, 55(5):448–452, 2013. DOI: 10.1111/dmcn.12049.

[378]R. S. Flynn, M. D. Huber, and S. B. DeMauro. Predictive Value of the BSID-II and the Bayley-III for Early School Age Cognitive Function in Very Preterm Infants. *Global Pediatric Health*, 7:2333794X20973146, 2020. DOI: 10.1177/2333794X20973146.

[379]M.-X. Li, J. M. Y. Yeung, S. S. Cherny, and P. C. Sham. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, 131(5):747–756, 2012. DOI: 10.1007/s00439-011-1118-2.

[380]J. Z. Bakdash and L. R. Marusich. Repeated Measures Correlation. *Frontiers in Psychology*, 8, 2017. DOI: 10.3389/fpsyg.2017.00456.

[381]R. Vallat. Pingouin: statistics in Python. *Journal of Open Source Software*, 3(31):1026, 2018. DOI: 10.21105/joss.01026.

[382]H. G. Cavalcanti, A. D. da Silva Nunes, B. K. S. da Cunha, K. de Freitas Alvarenga, S. A. Balen, and A. Pereira. Early exposure to environment sounds and the development of cortical auditory evoked potentials of preterm infants during the first 3 months of life. *BMC Research Notes*, 13(1):303, 2020. DOI: 10.1186/s13104-020-05129-8.

[383]T. Nishida, T. Kusaka, K. Isobe, S. Ijichi, K. Okubo, T. Iwase, K. Kawada, M. Namba, T. Imai, and S. Itoh. Extrauterine environment affects the cortical responses to verbal stimulation in preterm infants. *Neuroscience Letters*, 443(1):23–26, 2008. DOI: 10.1016/j.neulet.2008.07.035.

[384]G. C. Smith, J. Gutovich, C. Smyser, R. Pineda, C. Newnham, T. H. Tjoeng, C. Vavasseur, M. Wallendorf, J. Neil, and T. Inder. Neonatal intensive care unit stress is associated with brain development in preterm infants. *Annals of Neurology*, 70(4):541–549, 2011. DOI: 10.1002/ana.22545.

[385]B. Alexander, C. E. Kelly, C. Adamson, R. Beare, D. Zannino, J. Chen, A. L. Murray, W. Y. Loh, L. G. Matthews, S. K. Warfield, P. J. Anderson, L. W. Doyle, M. L. Seal, A. J. Spittle, J. L. Y. Cheong, and D. K. Thompson. Changes in neonatal regional brain volume associated with preterm birth and perinatal factors. *NeuroImage*, 185:654–663, 2019. DOI: 10.1016/j.neuroimage.2018.07.021.

[386]J. De Asis-Cruz, K. Kapse, S. K. Basu, M. Said, D. Scheinost, J. Murnick, T. Chang, A. du Plessis, and C. Limperopoulos. Functional brain connectivity in ex utero premature infants compared to in utero fetuses. *NeuroImage*, 219:117043, 2020. DOI: 10.1016/j.neuroimage.2020.117043.

[387]M. Bouyssi-Kobar, M. Brossard-Racine, M. Jacobs, J. Murnick, T. Chang, and C. Limperopoulos. Regional microstructural organization of the cerebral cortex is affected by preterm birth. *NeuroImage. Clinical*, 18:871–880, 2018. DOI: 10.1016/j.nicl.2018.03.020.

[388]R. J. Schafer, C. Lacadie, B. Vohr, S. R. Kesler, K. H. Katz, K. C. Schneider, K. R. Pugh, R. W. Makuch, A. L. Reiss, R. T. Constable, and L. R. Ment. Alterations in functional connectivity for language in prematurely born adolescents. *Brain: A Journal of Neurology*, 132:661–670, Pt 3, 2009. DOI: 10.1093/brain/awn353.

[389]S. Scherjon, J. Briët, H. Oosting, and J. Kok. The Discrepancy Between Maturation of Visual-Evoked Potentials and Cognitive Outcome at Five Years in Very Preterm Infants With and Without Hemodynamic Signs of Fetal Brain-Sparing. *Pediatrics*, 105(2):385–391, 2000. DOI: 10.1542/peds.105.2.385.

[390]C. A. Newnham, T. E. Inder, and J. Milgrom. Measuring preterm cumulative stressors within the NICU: The neonatal infant stressor scale. *Early Human Development*, 85(9):549–555, 2009. DOI: 10.1016/j.earlhumdev.2009.05.002.

[391]C. M. Wheatley, J. L. Dickinson, D. A. Mackey, J. E. Craig, and M. M. Sale. Retinopathy of prematurity: recent advances in our understanding. *British Journal of Ophthalmology*, 86(6):696–700, 2002. DOI: 10.1136/bjo.86.6.696.

[392]M. Kozberg and E. Hillman. Neurovascular coupling and energy metabolism in the developing brain. In *Progress in Brain Research*. Volume 225, pages 213–242. Elsevier, 2016. DOI: 10.1016/bs.pbr.2016.02.002.

[393]H. Blencowe, J. E. Lawn, T. Vazquez, A. Fielder, and C. Gilbert. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. *Pediatric Research*, 74:35, Suppl 1, 2013. DOI: 10.1038/pr.2013.205.

[394]H. Blencowe, S. Moxon, and C. Gilbert. Update on Blindness Due to Retinopathy of Prematurity Globally and in India. *Indian pediatrics*, 53 Sup:S89–S92, 2016.

[395]B. Lacoste, C. H. Comin, A. Ben-Zvi, P. S. Kaeser, X. Xu, L. d. F. Costa, and C. Gu. Sensory-Related Neural Activity Regulates the Structure of Vascular Networks in the Cerebral Cortex. *Neuron*, 83(5):1117–1130, 2014. DOI: 10.1016/j.neuron.2014.07.034.

[396]C. Whiteus, C. Freitas, and J. Grutzendler. Perturbed neural activity disrupts cerebral angiogenesis during a postnatal critical period. *Nature*, 505(7483):407–411, 7483, 2014. DOI: 10.1038/nature12821.

[397]E. G. Argandoña and J. V. Lafuente. Effects of dark-rearing on the vascularization of the developmental rat visual cortex. *Brain Research*, 732(1):43–51, 1996. DOI: 10.1016/0006-8993(96)00485-4.

[398]J. B. Ackman, T. J. Burbridge, and M. C. Crair. Retinal waves coordinate patterned activity throughout the developing visual system. *Nature*, 490(7419):219–225, 7419, 2012. DOI: 10.1038/nature11529.

[399]D. K. Warland, A. D. Huberman, and L. M. Chalupa. Dynamics of Spontaneous Activity in the Fetal Macaque Retina during Development of Retinogeniculate Pathways. *Journal of Neuroscience*, 26(19):5190–5197, 2006. DOI: 10.1523/JNEUROSCI.0328-06.2006.

[400]M. B. Feller. Spontaneous Correlated Activity in Developing Neural Circuits. *Neuron*, 22(4):653–656, 1999. DOI: 10.1016/S0896-6273(00)80724-2.

[401]B. Lacoste and C. Gu. Control of cerebrovascular patterning by neural activity during postnatal development. *Mechanisms of Development*. Neurovascular Unit, 138:43–49, 2015. DOI: 10.1016/j.mod.2015.06.003.

[402]L. A. Ridnour, J. S. Isenberg, M. G. Espey, D. D. Thomas, D. D. Roberts, and D. A. Wink. Nitric oxide regulates angiogenesis through a functional switch involving thrombospondin-1. *Proceedings of the National Academy of Sciences*, 102(37):13147–13152, 2005. DOI: 10.1073/pnas.0502979102.

[403]J. Lowe, S. W. Duvall, P. C. MacLean, A. Caprihan, R. Ohls, C. Qualls, and J. Phillips. Comparison of Structural Magnetic Resonance Imaging and Development in Toddlers Born Very Low Birth Weight and Full-term. *Journal of Child Neurology*, 26(5):586–592, 2011. DOI: 10.1177/0883073810388418.