Multimodal assessment of neonatal pain



Marianne Louise van der Vaart

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Abstract

Pain assessment is critical to prevent suffering and harm in infants admitted to the neonatal care unit. As pain is a subjective experience, its assessment in non-verbal infants relies on surrogate measures. Current infant pain assessment tools that are based on behaviour and autonomic nervous system measurements lack face validity — they are unlikely to reflect pain in all its dimensions. In recent years, EEG-derived measures of pain have been developed in late preterm and term infants. Multimodal tools which include these cerebral measurements are conceptually more appropriate to measure pain. Yet, their use is still limited to specific research applications. This thesis focuses on outstanding questions that need to be addressed in order to advance the development of multimodal pain assessment tools that incorporate cerebral measurements.

In the first part of this thesis, I focus on the characterisation of preterm infants' noxious-evoked responses and their development. Across several modalities, premature infants have dampened or altered responsiveness compared to term infants, and it is uncertain if these responses can be reliably discriminated from tactile-evoked responses. In particular, a discriminative pattern of noxious-evoked EEG activity that is present in term infants, is unlikely to be present in preterm infants. In addition, it is unclear how noxious-evoked responses, especially brain-derived responses, change with age. In this thesis, I use a classification model to show that infants aged 28–40 weeks postmenstrual age display discriminable multimodal responses to a noxious clinical procedure and a tactile control procedure, and I provide examples of how a such a model could be used in clinical trials of analgesics. I show that noxious-evoked responses change magnitude and morphology across this age range, and that discriminative brain activity emerges in early prematurity. In the second part of this thesis, I focus on improving the neuroscientific validity of a noxious-evoked EEG response measured at the cot-side, as the spatial neural correlates of these responses are still poorly understood. I present an EEG-fMRI pilot study to investigate the spatial neural correlates of inter-individual differences in noxious-evoked EEG responses and provide recommendations for a larger follow-up study.

Overall, this thesis provides a characterisation of infants' noxious-evoked responses and their development across multiple modalities, a crucial next step in improving multimodal neonatal pain assessment.

Acknowledgements

First of all, I am most grateful to the families who took the time to consider participating in research. Their willingness to help at such a unique moment in their life is truly exceptional.

Heartfelt thanks go out to my supervisors Rebeccah, Sean and Ana. Rebeccah, thank you for pouring your intelligence, energy and creativity into my projects and being such an important role model for women in science. I hope that you continue to inspire little people to become professors. Sean, your honest critiques have made my work a million times better and I will thoroughly miss our weekly meetings. Ana, thank you for welcoming me into your group and introducing me to the wonders of ultrasound.

Thank you to all the members of the Paediatric Neuroimaging Group for being wonderful friends and colleagues, in particular to Luke and Caroline for their advice and guidance on data analysis; Gabi, Foteini, Ria, Fiona and the other clinical members of the team for their invaluable role in data collection; and to Deniz, Maria and Miranda, for being a listening ear, especially in those first rockiest months. Thank you also to David Parker, Michael Sanders and Jon Campbell who made the MRI studies in this thesis possible.

Thank you to the Moat House family for instantly making me feel at home in Oxford and putting up with my incessant mug hoarding while working from home. Thank you Brooke, Gareth and Rory for a wonderful year in Abingdon Road and providing me with an education in cinematography, and thank you to my friends from Leiden, especially Julie and Julia, who never quite understood my struggles with thesis writing, but somehow always had the best advice.

Thank you to my family in the Netherlands and Canada; especially my parents and Pascal. Your positivity and support made me feel like home was never far away.

Finally, this thesis is dedicated to Jacob, who motivates me to do better every day without ever making me feel I'm not already good enough. Jacob, I can honestly say that I would have never embarked on this journey if it wasn't for you — and I truly wouldn't have wanted to do it any other way.

Research in verse

Babies born yesterday:
Tricky to chatter with
Crying means hunger or
Pain, little tyke?

Misunderstandings are
Unequivocally
Painful for newborns and
Carers alike.

Solving this problem is
Not a no-brainer but
Many will grant it is
Close to the heart

Brain waves plus metrics of
Cardiovascular
Functions tell painful and
Tactile apart.

So many measurements
One tiny neonate
How do we notice the
Wood through the trees?

Grow random forests that
Combinatorically
Pick the right features and
Classes with ease.

Classification trees:
Quite theoretical
How might this impact the
Clinic you ask?

Pain relief trials in
Neonatology
Hopefully will be an
Easier task!

Marianne van der Vaart, 2021

I wrote this poem about my DPhil for the "Research in Verse" competition organised by St Cross College. The poem is a combination of four 'Ollekebollekes' [319], a type of verse popularised in the 1970s in my native language, Dutch, by the writer and artist Drs. P (1919–2015). Ollekebollekes are a modification of a rhyme form called the Double Dactyl and follow a strict pattern: each poem has eight lines, where lines one, two, three, five, six and seven are double dactyls (a dactyl consists of three syllables where the emphasis is placed on the first syllable, e.g. *Yes*-ter-day). Lines four and eight consist of a dactyl followed by one stressed syllable (e.g. Close to the *heart*) and must rhyme. Line six is again a double dactyl, but this time made up of exactly one six-syllable word. This format dictates the singsong metre at which the poems should be recited. Often, Ollekebollekes contain a play on words. During his lifetime, Drs. P wrote hundreds of Ollekebollekes on various topics, from geography to religion, and from art to science [319]. In my opinion they are also delightfully suitable to summarise research findings with.

Contents

List of Abbreviations **vi**

List of Definitions **ix**

List of Figures **x**

List of Tables **xi**

I General principles 1

1 Introduction **2**
- 1.1 Do newborn infants feel pain? . . . 2
- 1.2 Defining pain . . . 3
- 1.3 Pain in neonatal care . . . 4
- 1.4 Infant pain assessment: a focus on behaviour and vital signs . . . 4
- 1.5 Neonatal pain as a multidimensional construct . . . 5
- 1.6 Physiology of nociception and pain . . . 7
  - 1.6.1 Historical views on pain: from labelled lines to brain-centered 7
  - 1.6.2 Peripheral encoding of noxious stimuli . . . 9
  - 1.6.3 The spinal dorsal horn: projection, modulation and reflex activity . . . 9
    - 1.6.3.1 Nociceptive reflex circuits in the spinal cord . . . 10
    - 1.6.3.2 Descending modulation of spinal dorsal horn activity 10
  - 1.6.4 Supraspinal regions . . . 10
  - 1.6.5 Comparison to non-nociceptive touch . . . 11
  - 1.6.6 Summary . . . 12
- 1.7 Development of nociceptive pathways . . . 12
  - 1.7.1 Development of nociceptors and inhibition in the spinal dorsal horn . . . 12
  - 1.7.2 Development of thalamocortical projections and supraspinal areas . . . 13
  - 1.7.3 Development of the processing of non-nociceptive stimuli . . 15
  - 1.7.4 Summary . . . 15
- 1.8 Reviewing measurements for infant pain assessment . . . 15
  - 1.8.1 Behavioural measurements . . . 16
  - 1.8.2 Autonomic nervous system measurements (vital signs) . . . 17
  - 1.8.3 Summary . . . 17

1.9 Brain-derived measures of pain . . . 18
1.9.1 EEG-derived measures . . . 18
1.9.1.1 Physiological basis of EEG and ERP recordings . . . 18
1.9.1.2 Adult noxious-evoked ERPs . . . 19
1.9.1.3 Infant noxious-evoked ERPs . . . 19
1.9.1.4 Noxious-evoked spectral EEG changes . . . 21
1.9.2 Functional MRI . . . 22
1.9.2.1 Physiological basis of the BOLD signal . . . 22
1.9.2.2 Adult fMRI signatures of pain . . . 22
1.9.2.3 Inferring cognitive states . . . 23
1.9.3 Summary . . . 25
1.10 Multimodal measurements . . . 25
1.11 Thesis overview . . . 25

2 General methods **27**
2.1 Retrospective datasets . . . 27
2.1.1 Oxford Heel Lance Dataset . . . 27
2.1.1.1 Data inclusion criteria . . . 28
2.1.1.2 Noxious and non-noxious stimuli . . . 30
2.1.1.3 Recording techniques . . . 31
2.1.2 Oxford Immunisation Dataset . . . 32
2.1.3 UCL Dataset . . . 33
2.2 Prospective datasets . . . 34
2.2.1 Experimental noxious stimuli . . . 34
2.2.2 Infant EEG-MRI study . . . 34
2.2.2.1 Recruitment . . . 35
2.2.2.2 EEG session . . . 36
2.2.2.3 MRI session . . . 36
2.2.3 Adult EEG-MRI study . . . 38
2.2.3.1 EEG session . . . 39
2.2.3.2 MRI session . . . 39
2.3 Data analysis principles . . . 39
2.3.1 EEG . . . 40
2.3.2 fMRI . . . 40
2.4 Software . . . 41

II Investigating multimodal noxious-evoked responses in premature infants **44**

3 Discrimination of responses to noxious and non-noxious stimuli **45**
3.1 Introduction . . . 45
3.2 Methods . . . 48
3.2.1 Study design . . . 48
3.2.2 Datasets . . . 49
3.2.3 Features of interest . . . 52

3.2.4 Data acquisition . . . 53
3.2.5 Feature extraction . . . 54
3.2.5.1 EEG preprocessing . . . 54
3.2.5.2 Template of noxious-evoked EEG activity . . . 55
3.2.5.3 Derivation of age-dependent PCs . . . 56
3.2.5.4 ERSP analysis . . . 60
3.2.5.5 EMG analysis . . . 61
3.2.5.6 ECG analysis . . . 61
3.2.5.7 Facial expression analysis . . . 62
3.2.6 Classification model . . . 62
3.2.6.1 Theory . . . 62
3.2.6.2 Implementation and evaluation . . . 63
3.2.6.3 Feature importance . . . 64
3.2.7 Examples of classifier applications . . . 65
3.2.7.1 Exploring within-infant discrimination: paired comparisons . . . 65
3.2.7.2 Testing the classification model in a pharmacological study . . . 67
3.3 Results . . . 67
3.3.1 Noxious stimulation evokes changes in cerebral, autonomic and behavioural modalities in infants 28–40 weeks . . . 67
3.3.2 Development of a classification model . . . 68
3.3.3 Generalisation of classification results . . . 69
3.3.4 Influence of PMA on classification results . . . 70
3.3.5 Feature importance . . . 72
3.3.6 Examples of classifier applications . . . 74
3.3.6.1 Exploring within-infant discrimination: paired comparisons . . . 74
3.3.6.2 Application of the classifier to a pharmacological study . . . 77
3.4 Discussion . . . 79
3.4.1 Even very preterm infants display discriminable responses to noxious and non-noxious stimuli . . . 79
3.4.2 Multimodal features contribute to classification performance 80
3.4.3 Limitations of feature selection and extraction procedure . . 81
3.4.4 Exploring clinical applications . . . 83
3.4.4.1 Establishing a trade-off between false positives and false negatives . . . 83
3.4.4.2 Evaluation of the classifier within the COSMIN framework . . . 84
3.4.5 Summary and next steps . . . 86

4 Development of noxious-evoked responses 88
4.1 Introduction . . . 88
4.2 Methods . . . 90
4.2.1 Chapter overview . . . 90

4.2.2 Datasets . . . 91
4.2.3 Part 1: Developmental trajectories of multimodal noxious-evoked responses . . . 92
4.2.4 Part 2: Investigation of morphological changes in noxious-evoked ERPs and their relationships . . . 95
4.2.4.1 Discriminative ability of noxious-evoked ERPs . . . 95
4.2.4.2 Multimodal patterns of noxious-evoked responses . 96
4.2.4.3 Relationship between discriminative PCs and behavioural and autonomic domains . . . 96
4.2.5 Part 3: Exploratory Analyses: Reproducibility of developmental trajectories and effect of PNA and other clinical factors 96
4.3 Results . . . 97
4.3.1 Part 1: Developmental trajectories of multimodal noxious-evoked responses . . . 97
4.3.1.1 Development of noxious-evoked changes in facial expressions, heart rate, reflex withdrawal and ERSP features . . . 98
4.3.1.2 Development of noxious-evoked ERP morphology . 99
4.3.2 Part 2: Investigation of morphological changes in noxious-evoked ERPs and their relationships . . . 102
4.3.2.1 Differences in ERPs evoked by noxious and non-noxious stimuli were present throughout development 102
4.3.2.2 Multimodal patterns of noxious-evoked activity changed throughout development . . . 104
4.3.2.3 Noxious-evoked PCs were not significantly correlated with cardiovascular and behavioural measures 105
4.3.3 Part 3: Exploratory Analyses: Reproducibility of developmental trajectories and effect of PNA and other clinical factors 105
4.4 Discussion . . . 107
4.4.1 Noxious-evoked ERPs change morphology . . . 107
4.4.2 Discriminatory ERPs are present in premature infants . . . 110
4.4.3 Developmental changes in noxious-evoked activity are also present across other modalities . . . 112
4.4.4 Developmental trajectories are reproducible across datasets . 113
4.4.5 Limitations and future work . . . 113
4.4.6 Summary and next steps . . . 115

III Investigating spatial correlates of a measure of noxious-evoked EEG activity 118

5 Introduction 119
5.1 A template of noxious-evoked EEG activity . . . 119
5.2 Neuroscientific validity . . . 120
5.3 Understanding noxious-evoked EEG activity . . . 121
5.4 Spatial organisation of noxious-evoked EEG activity . . . 121

5.5 Inter-individual variability in noxious-evoked brain activity 123
5.6 Aims and study overview 124

6 Noxious-evoked EEG-BOLD relationships in adults: derivation of ROIs 128
6.1 Methods 128
6.1.1 Participants and design 128
6.1.2 Experimental procedures 129
6.1.2.1 EEG and psychophysics 129
6.1.2.2 MRI 129
6.1.3 Pyschophysics analysis 130
6.1.4 EEG preprocessing 130
6.1.5 Summarising pinprick-evoked potentials (PEP) 131
6.1.5.1 Principal component analysis 131
6.1.5.2 Comparison of PCs with amplitude-based and behavioural measures 133
6.1.6 fMRI Single-subject analysis 134
6.1.7 fMRI Group level analysis 135
6.1.7.1 Mean activation 135
6.1.7.2 PEP-BOLD relationships: identifying ROIs 137
6.2 Results 138
6.2.1 Behavioural responses to pinprick stimuli 138
6.2.2 Pinprick-evoked potentials 139
6.2.3 Derivation of a summary measure of pinprick-evoked potentials 141
6.2.4 Noxious-evoked BOLD activity 146
6.2.5 PEP-BOLD relationships 149
6.3 Discussion 154
6.3.1 PEP-BOLD relationships in adults 154
6.3.2 Ambiguous interpretation of adult PEP amplitude 155
6.3.3 Inconsistencies in PEP-BOLD relationships across intensities 156
6.3.4 Summary and next steps 157

7 Exploring fMRI correlates of noxious-evoked EEG responses in infants 158
7.1 Methods 158
7.1.1 Participants and design 158
7.1.2 Experimental procedures 159
7.1.3 EEG analysis 159
7.1.3.1 Preprocessing and extracting EEG template magnitude 159
7.1.3.2 Statistics 159
7.1.4 fMRI Single-subject analysis 160
7.1.4.1 Preprocessing 160
7.1.4.2 Single-subject statistics 161
7.1.5 fMRI Group level analysis (mean activation) 161
7.1.6 EEG template magnitude–BOLD relationships 161

7.1.6.1 Approach 1: Neurosynth signature 162
7.1.6.2 Approach 2: ROI analysis 163
7.1.6.3 Approach 3: Voxelwise analysis 163
7.2 Results 164
7.2.1 Feasibility of an EEG-fMRI study in infants 164
7.2.2 Infant EEG and BOLD responses to clinical procedures and pinprick stimuli 164
7.2.3 Exploring analytical methods to evaluate EEG template magnitude–BOLD relationships in infants 169
7.2.4 Summary 174
7.3 Discussion 175
7.3.1 Feasibility and data quality 175
7.3.2 Preliminary results 176
7.3.3 Recommendations for follow-up studies 179
7.3.3.1 Sample size and data quality checks 179
7.3.3.2 Design 179
7.3.3.3 Analysis 181
7.3.4 Summary and next steps 182

IV Discussion 183

8 General discussion 184
8.1 Thesis summary 184
8.2 Towards clinical tools 185
8.2.1 Responsiveness and clinical significance 185
8.2.2 Implementing multimodal pain assessment in clinical practice 188
8.3 A better understanding of noxious-evoked brain activity in the context of brain development 189
8.3.1 The role of the subplate in noxious-evoked ERP transitions . 189
8.3.2 Anatomical source of developing noxious-evoked ERPs . . . 190
8.3.3 The use of experimental noxious stimuli to study infant nociception 191
8.4 What does it feel like to be in pain? 192

V Appendix and bibliography 194

A Appendix: Parent information leaflets and consent forms (infant studies) 195

B Appendix: Participant information leaflet and consent form (adult study) 208

C Appendix: Publications, funding and awards **216**
C.1 Publications relevant to this thesis 216
C.2 Other publications 216
C.3 Funding and awards 217

Bibliography **219**

List of Abbreviations

2DFT . . . . . Two-Dimensional Fourier Transform
a.u. . . . . . Arbitrary units
ACC . . . . . Anterior Cingulate Cortex
ANS . . . . . Autonomic Nervous System
ANTs . . . . . Advanced Normalization Tools
AUC . . . . . Area Under the Curve
BOLD . . . . . Blood Oxygen Level Dependent
CI . . . . . Confidence Interval
COPE . . . . . Contrast of Parameter Estimates
COSMIN . . . COnsensus-based Standards for the selection of health Measurement INstruments
df . . . . . Degrees of freedom
dHCP . . . . . Developing Human Connectome Project
EPI . . . . . Echo Planar Imaging
GA . . . . . Gestational age
GRE . . . . . Gradient Echo
ECG . . . . . Electrocardiography
EEG . . . . . Electroencephalography
EMG . . . . . Electromyography
ERP . . . . . Event-related Potential
ERSP . . . . . Event-related Spectral Perturbations
FIX . . . . . FMRIB's ICA-based Xnoiseifier
FLIRT . . . . . FMRIB's Linear Image Registration Tool
fMRI . . . . . Functional Magnetic Resonance Imaging
fNIRS . . . . . Functional Near-Infrared Spectroscopy
FNR . . . . . False Negative Rate
FPR . . . . . False Positive Rate
FSL . . . . . FMRIB Software Library
FWE(R) . . . Family-Wise Error (Rate)

GABA . . . . . Gamma Aminobutyric Acid
GLM . . . . . General Linear Model
GRAPPA . . . GeneRalized Autocalibrating Partial Parallel Acquisition
GRE . . . . . Gradient Response Echo
HIE . . . . . Hypoxic Ischemic Encephalopathy
HRF . . . . . Haemodynamic Response Function
IASP . . . . . International Association for the Study of Pain
ICA . . . . . Independent Component Analysis
IVH . . . . . Intraventricular Haemorrhage
MCC . . . . . Midcingulate Cortex
MCFLIRT . . Motion Correction FMRIB's Linear Image Registration Tool
NICU . . . . . Neonatal Intensive Care Unit
NFCS . . . . . Neonatal Facial Coding System
NIPS . . . . . Neonatal Infant Pain Scale
NHS . . . . . National Health Service
NPS . . . . . Neonatal Pain Scale
NRS . . . . . Numerical Rating Scale
OFC . . . . . Orbitofrontal Cortex
PALM . . . . . Permutation Analysis of Linear Models
PC . . . . . Principal Component
PCA . . . . . Principal Component Analysis
PEP . . . . . Pinprick-evoked Potential
PIPP(-R) . . . Premature Infant Pain Profile (-Revised)
PMA . . . . . Postmenstrual Age
PNA . . . . . Postnatal Age
RCT . . . . . Randomised Controlled Trial
REC . . . . . Research Ethics Committee
RMS . . . . . Root Mean Square
ROC . . . . . Receiver Operator Characteristic
ROI . . . . . Region of Interest
SyN . . . . . Symmetric Normalization
SMA . . . . . Supplementary Motor Area
SNR . . . . . Signal-to-Noise Ratio
TE . . . . . Echo Time

TFCE . . . . . Threshold-Free Cluster Enhancement
TR . . . . . Repetition Time
TSE . . . . . Turbo Spin Echo
WIN . . . . . Wellcome Centre for Integrative Neuroimaging

List of Definitions

Gestational age . . . . . Duration of gestation based on best obstetric estimate [50]

Postnatal age . . . . . Time elapsed since birth [50]

Postmenstrual age . . . Gestational age at birth + postnatal age [50]

False positive rate . . . Number of false positives divided by number of true negatives (1 - specificity)

False negative rate . . . Number of false negatives divided by number of true positives (1 - sensitivity)

List of Figures

2.1 Overview of data included in the Oxford Heel Lance Dataset. 30
2.2 Overview of infant EEG preprocessing steps. 41

3.1 Graphical representation of Chapter 3. 50
3.2 PMA and PNA distributions in the three main datasets. 51
3.3 Gaussian windows used to construct age-weighted averages. 58
3.4 The nine principal components derived from the EEG data. 59
3.5 Example EEG traces from two individual infants overlaid with EEG templates. 59
3.6 Average $ERSP_{log}$ plot at Cz in the Oxford Training Dataset. 60
3.7 Average responses to the control heel lance and heel lance in the Oxford Training Dataset. 68
3.8 Performance of the classifier in the Oxford Training Dataset and two independent datasets. 70
3.9 Performance of the classifier in the Oxford Training Dataset and two independent datasets, in the premature (PMA $<$ 37 weeks) infants only 71
3.10 Classifier performance by PMA. 72
3.11 Feature importance estimates. 74
3.12 Probability distribution for the difference in classification score between the noxious and non-noxious stimulus, split by age group and dataset. 76
3.13 Scores and probability distribution for the control group (who received paracetamol after the immunisation) and the paracetamol group (who received paracetamol before the immunisation). 78

4.1 Graphical overview of the analysis steps in Chapter 4. 91
4.2 Construction of age-weighted averages. 94
4.3 Developmental trajectories of the five ERSP features. 98
4.4 Amplitude of individual baseline-corrected trials for each infant at Cz. 100
4.5 Characterisation of the three PCs that explained most of the variance in the noxious-evoked EEG data. 101
4.6 Comparison of noxious and non-noxious evoked ERPs in different age groups. 104
4.7 Normalised magnitude of noxious-related responses at 29–39 weeks PMA. 105

4.8 Exploration of the influence of clinical factors on PMA-feature relationships. . . . 107
4.9 Overview of the development of multimodal noxious-evoked responses. 116
4.10 Overview of the development of multimodal responses evoked by a non-noxious stimulus. . . . 117

5.1 Graphical representation of the analysis steps and figures in Chapter 6. 126
5.2 Graphical representation of the analysis steps and figures in Chapter 7. 127

6.1 Numerical rating scale scores for the 64, 128 and 512 mN pinpricks. 139
6.2 Group EEG responses to pinprick stimuli in adults. . . . 140
6.3 Derivation of a summary measure of pinprick-evoked potentials in adults (128 mN). . . . 142
6.4 Derivation of a summary measure of pinprick-evoked potentials in adults (512 mN). . . . 143
6.5 Within-participant averages for the 128 mN data sorted by PC1-128 magnitude. . . . 144
6.6 Within-participant averages for the 512 mN data sorted by PC1-512 magnitude. . . . 145
6.7 Average BOLD activation to the pinprick stimuli. . . . 147
6.8 Direct comparisons of BOLD activity evoked by the 512 mN stimulus and the lower intensity stimuli. . . . 148
6.9 Mean weighted Neurosynth activity for the 64, 128 and 512 mN pinpricks increases with intensity. . . . 148
6.10 Unthresholded t-statistic maps showing areas in which BOLD activity evoked by the 128 mN stimuli co-varies with PC1-128 magnitude. . 150
6.11 Unthresholded t-statistic maps showing areas in which BOLD activity evoked by the 512 mN stimuli co-varies with PC1-512 magnitude. . 151
6.12 Unthresholded t-statistic maps showing areas in which BOLD activity evoked by the 512 mN stimuli co-varies with pinprick-evoked EEG activity. . . . 152
6.13 Correlations between PC1-128 magnitude and PC1-512 magnitude and mean BOLD activity in all identified clusters with FWE corrected $p < 0.6$. . . . 153

7.1 The Neurosynth signature in infant space. . . . 163
7.2 Mean infant EEG responses to the control heel lance (n = 7), heel lance (n = 7) and 128 mN pinprick stimuli (n = 5). . . . 166
7.3 Individual infant EEG responses to the noxious heel lance. . . . 167
7.4 Mean EEG response to the 128 mN pinprick in infants and adults. . 167
7.5 Comparison of adult and infant BOLD activity evoked by the pinprick stimuli. . . . 169
7.6 Correlations between Neurosynth activity and EEG template magnitudes. . . . 170
7.7 The 11 regions of interest (ROI) identified in the adult data in infant space. . . . 171

7.8 Correlations between heel lance EEG template magnitudes and activity in the adult ROIs. 172
7.9 Voxelwise analysis of areas in which BOLD activity co-varies with the heel lance EEG template magnitude. 174

List of Tables

3.1 Demographic variables for the participants in Oxford Training Dataset, the Oxford Held-out Test Dataset and the UCL Dataset. 51
3.2 Response modalities and features used in this study, organised by response domain. 53
3.3 Overview of missing data for each feature. 55
3.4 Time-frequency windows of interest. 61
3.5 Forced choice accuracy, mean difference in score for the heel lance and the control heel lance and p-value for the paired t-test comparing the scores in the Oxford Training Dataset. 77
3.6 Forced choice accuracy, mean difference in score for the heel lance and the control heel lance and p-value for the paired t-test comparing the scores in the in the UCL Dataset. 77

6.1 Clusters in which BOLD activity evoked by the 128 mN pinprick was associated with variation in PC1-128 magnitude. 150
6.2 Clusters in which BOLD activity evoked by the 512 mN pinprick was associated with variation in PC1-512 magnitude. 151
6.3 Clusters in which BOLD activity evoked by the 128 mN pinprick was associated with variation in mean PC1-128 and PC1-512 magnitude. 152

Part I

General principles

1 Introduction

1.1 Do newborn infants feel pain?

A survey distributed amongst all paediatric anaesthetists in the United Kingdom in 1988 asked its respondents whether they believed newborn infants could feel pain. Although 80% of the 60 respondents agreed, the vast majority reported that analgesics were rarely prescribed pre-, peri- and post-operatively [257]. The authors of the survey reasoned that this might be because "*Often, the decision to give analgesia is made on a very subjective assessment (...)*" [257].

This statement remains highly relevant today, as the care for hospitalised infants often involves painful procedures. Robust pain assessment tools are therefore of critical importance. Although an abundance of such tools are available, many have been criticised for being subjective or non-specific. Rapid developments in the field of neonatal neuroscience have paved the way for the use of more objective brain-derived measures of pain. Yet, several important knowledge gaps remain. Little is known about the development of brain responses to pain in very preterm infants, who bear the largest burden of noxious clinical procedures, and the neurological basis of brain-derived measures is still poorly understood. The work in this thesis aims to address some of these outstanding questions in order to improve the multimodal assessment of pain in neonates.

This introduction outlines the definition, epidemiology and physiology of pain and the ways it can be measured in infants.

1.2 Defining pain

Conceptual theories of pain date back hundreds to thousands of years. The ancient Greek physician Galen considered pain a rapid and intense negative disruption of the body's natural state, brought upon by a stimulus that damages the body's continuity [31]. Such a stimulus that damages or threatens damage to normal tissues was later termed "noxious", and its neural encoding "nociception" [151]. Pain, however, is not merely the encoding of a noxious stimulus. The International Association for the Study of Pain (IASP) acknowledged this in their first definition of pain, released in 1979. Here, they defined pain as:

> *An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage* [150]

Pain, in contrast to nociception, thus refers to a subjective experience which is not necessarily correlated with the intensity of noxious input. In extreme cases, pain can even exist without nociception, for instance in certain chronic pain conditions [247].

Discussions on the definition of pain are ongoing. The IASP definition was updated in 2020 [258] to read:

> *An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage* [258]

Although this change may seem trivial at first sight, the removal of the word "described" shifts the emphasis away from verbal report and makes it abundantly clear that pain need not be described in order to be real. It does leave the reader with a burning question: if pain is inherently subjective, how can it be inferred in those who cannot describe it?

1.3 Pain in neonatal care

This problem is pertinent in infants in the neonatal intensive care unit (NICU). Studies in European NICUs showed that neonatal in-patients required multiple noxious clinical procedures a day, including tracheal suctioning, blood draws, and cannula insertions [38, 284, 264]. Similar procedures are associated with heightened verbal pain reports in adult ICU patients [256]. Noxious procedures are associated with acute distress and physiological instability in infants [222] and increased levels of parental concern [101, 188]. In addition, exposure to a high number of noxious procedures in early life is associated with long-term alterations in somatosensory processing [138, 338], brain structure [259, 261, 332] and cognitive function [76, 332]. While the number of noxious clinical procedures experienced by infants in the NICU has decreased over the years [264], it is likely that some noxious procedures will remain essential for neonatal care [264]. It is thus vital that pain is adequately detected and that effective pain management strategies are developed and implemented.

Yet, the provision of analgesics in neonates is still far from optimal. Analgesics are often provided to neonates outside the terms of the manufacturer's license ("off-label") [109, 100, 55] and there is a lack of consensus in prescribing strategies between NICUs [39, 100]. As a result, there is a risk both of side effects and under-treatment [289]. In order to establish guidelines for safe and effective pain management, robust measures for neonatal pain are needed [289].

1.4 Infant pain assessment: a focus on behaviour and vital signs

The search for infant pain assessment tools gained momentum in the 1980s. Grunau and Craig [124] described a characteristic cluster of facial expressions that was evoked in infants after a clinically required skin-breaking blood test. This facial pattern included "brow bulge" (frowning), "eye squeeze" (squeezing the eyes shut) and "nasolabial furrow" (deepening of the grooves between the mouth and the nose)

[124]. Their study was followed up by investigations of noxious-evoked changes in body movements, heart rate and oxygen saturation [56].

These observations of noxious-evoked responses led to the development of the first infant pain assessment scales, such as the Neonatal Facial Coding Scale (NFCS, incorporating facial expressions, [123]), the Premature Infant Pain Profile (PIPP, incorporating facial expressions, heart rate and oxygen saturation and contextual factors [296]) and the Neonatal Infant Pain Scale (NIPS, incorporating behavioural responses, breathing pattern and arousal state, [191]). Since the development of these scales, numerous studies have confirmed the observations that noxious procedures are related to changes in vital signs and behaviour in infants [135] and over 40 different pain scales based on these modalities have been developed [51].

Nonetheless, the search for pain assessment tools is by no means complete. Recent meta-analyses have highlighted gaps in the validity [114] and responsiveness [213] of several scales. In addition, increased pain scores can also be present in response to non-noxious care procedures [230, 141], which has prompted a debate on the usefulness of behavioural and physiological measures in distinguishing pain from distress [287]. On a more fundamental level, it is questionable whether behavioural and physiological measurements are the most appropriate surrogate measures of pain [19].

In the remainder of this chapter, I will argue that they are not, and delineate ways to improve pain assessment by incorporating brain-derived measurements.

1.5 Neonatal pain as a multidimensional construct

The design of infant pain assessment tools is conceptually challenging. In contrast to some other neonatal problems, such as low blood sugar, pain is “non-observable”. That is, whereas it is possible to measure blood glucose levels, it is impossible to make a direct observation of pain. In the field of psychology, such non-observable characteristics are called “constructs” [61]. The COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiative aims to improve measurement instruments for constructs in health care and provides

guidelines for their development [227]. This thesis focuses on knowledge gaps that need to be addressed before brain-derived infant pain assessment measurements can be widely implemented as part of multimodal neonatal pain assessment. For this reason it is useful to consider existing and potential tools to measure neonatal pain within the COSMIN framework. According to COSMIN, the first step in designing any measurement tool is defining the construct of interest and the population of interest [62]. In this thesis, the prime interest is procedural pain in preterm and term infants up to 44 weeks postmenstrual age (PMA), where pain is defined according to the IASP definition.

Several theoretical models exist for designing a tool [61]. In a reflective model, the measurements ("items") that make up the tool are a reflection of the construct to be measured. In other words, the measurements change when the construct changes [61]. An example of a reflective model for measuring pain in adults is the Numerical Rating Scale (NRS), where pain intensity is verbally indicated on a scale of 0 to 10. If an adult's pain experience (the construct) changes, we would expect their verbal report (the measurement) to change as well.

Once the construct is defined, the validity of the measurement instrument should be assessed. COSMIN defines validity as "the degree to which an instrument truly measures the constructs it purports to measure" [227, 66]. Several sub-types of validity exist. For instance, construct validity is defined as "the degree to which the scores of a measurement instrument are consistent with hypotheses, e.g. with regard to internal relationships, relationships with scores of other instruments or differences between relevant groups" [227, 66]. Evidence for the construct validity of a neonatal pain scale can be gathered e.g. by testing whether a pain score is higher after a noxious procedure than after a non-noxious procedure, as was done in the development of the PIPP score [296].

This exercise is futile, however, if a tool is inherently not a suitable reflection of the construct. The measurement property that relates to this concern, is called "content validity" and is defined as "the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured" [227, 66].

The establishment of content validity is a qualitative assessment based on existing knowledge of the construct [66]. It should consider whether the items included in the measurement instrument are relevant (i.e. relate to the construct) and comprehensive (i.e. cover all aspects of the construct). Establishing content validity is the first and most important step in the validation process [66, 62]. Face validity is "the degree to which a measurement instrument, indeed, looks as though it is an adequate reflection of the construct to be measured" [227, 66] and is a first, quick assessment of content validity. Returning to the example of the NRS for adult pain, if we consider that pain is a subjective experience, it seems reasonable to assume that the verbal report given by an adult is relevant and could provide an adequate reflection of the overall sensory and emotional aspects of the experience. Thus, we can conclude the NRS has face validity.

Having defined neonatal pain according to the IASP definition, an infant pain assessment tool would also need to cover the emotional and sensory aspects of this experience to have content validity. Arguably, the strongest case for a tool's validity can be made if it can be established that changes in the construct are causal to changes in the measurement, that is, the measurements are taken downstream from the construct [28]. In order to further understand which measurements could be downstream to the construct of infant pain, we need to understand the physiology of nociception and pain, and how these relate to measurement items.

1.6 Physiology of nociception and pain

As described in Section 1.2, nociception is not always implicated in pain. However, this thesis focuses on acute procedural pain, which necessarily involves a noxious stimulus and therefore nociception. The sections below briefly review historical perspectives on pain and nociception, followed by a description of nociceptive pathways.

1.6.1 Historical views on pain: from labelled lines to brain-centered

In the twentieth century, three theories dominated the field of pain research. The discovery of nerves that specifically transmitted sensory stimuli led to support for the "specificity theory" [223, 246]. This theory argued that pain is a unique sensation, distinct from tactile and other sensory sensations, with its own pain-specific neuronal pathways ("labelled lines") leading all the way from the affected organ to the brain [223, 246]. This was diametrically opposed to the "intensity theory", propagated by Bell, who argued that specific pain circuitry was non-existent and pain was nothing more than high-intensity sensory activity [223, 246]. Finally, "pattern theory" was centred around the idea that patterns of peripheral neuronal firing were the source of differentiation between pain and other sensory experiences. The commonality between these views of pain, is that the brain was assumed to passively receive input from the periphery [223, 246].

The specificity, intensity and pattern theories were brought together in the influential Gate Control Theory proposed by Melzack and Wall [217]. The authors proposed a model where small- and large-diameter peripheral neurons with varying thresholds to sensory stimuli project onto the same transmission cells in the spinal cord. Activity of the transmission cells is controlled by a "gate", which consists of inhibitory cells located in the spinal cord. The activation of these neurons, and therefore the closure of the gate, is controlled by the relative input from large-diameter and small-diameter neurons. Activation of the large-diameter neurons closes the gate and thus blocks nociceptive signals from reaching the brain, while stimulation of the small-diameter neurons opens the gate. Importantly, the theory also acknowledged the presence of brain-driven (top-down) modulation of nociceptive signalling on the spinal dorsal horn [217]. The Gate Control Theory proved to have some shortcomings. For instance, it is now commonly accepted that nociceptive-specific neurons do exist in the periphery as well as in the spinal dorsal horn ([246, 17, 218], see Section 1.6.2). Nonetheless, this model shaped our view of

the spinal dorsal horn and the brain as active areas for pain processing instead of passive relay stations [216].

In 1990, Melzack emphasised the role of the brain even more in his neuromatrix theory [215]. He proposed that "*the anatomical substratum of the physical self is a network of neurons that extends throughout widespread areas of the brain*" [215] and that the brain can thus generate the experience of pain. The emergence of functional MRI as an imaging technique not long after allowed *in vivo* human investigations of such pain-related brain activity [8] and the central role of the brain in pain perception is now well established [308]. Current views on nociception and pain are detailed in the sections below.

1.6.2 Peripheral encoding of noxious stimuli

It is now known that noxious stimuli are encoded by nociceptors, which are the free nerve endings of specialised nociceptive neurons in the skin [78, 282]. There are two main classes of nociceptive fibers: Aδ and C-fibers. Aδ fibers are medium-diameter myelinated fibers with a fast conductive speed which generate localised sharp pain sensations, while C-fibers are slow, small-diameter non-myelinated fibers which are involved in dull, aching pain [78, 282, 25]. Nociceptors can be further divided by the responsiveness of their terminal to particular noxious stimuli and to the irritant substance capsaicin [78]. The sensory experience evoked by mechanical pinprick stimuli, such as the ones studied in this thesis, are mediated by capsaicin-insensitive A-fibers, which include high-threshold mechanoreceptors (A-M, or HTM) and type I mechano-heat receptors (A-MH type I) [203, 147]. Other subtypes of A-fibers are more sensitive to thermal stimuli than mechanical stimuli, while most C-fibers are polymodal and can be activated by mechanical, thermal and chemical noxious stimulation [78].

1.6.3 The spinal dorsal horn: projection, modulation and reflex activity

The cell bodies of nociceptive fibers lie in the dorsal root ganglia, and their central projections extend to the spinal dorsal horn, predominantly to laminae I-II [78]. Nociceptive neurons synapse onto modulatory interneurons in the dorsal horn to form connections with second-order projection neurons [283, 305]. A subset of these projection neurons is nociceptive-specific, whereas wide dynamic range neurons are activated by a variety of stimuli [75]. The spinothalamic tract is the major ascending tract that mediates pain sensation. The tract crosses the midline to project to the contralateral thalamus, which is in turn connected to various cortical areas [283, 282]. Other projection neurons arising from the spinal dorsal horn form connections with the brain stem, which are likely to relay the autonomic, endocrine, motor and arousal effects of noxious stimulation [283, 305, 282].

1.6.3.1 Nociceptive reflex circuits in the spinal cord

Noxious stimuli evoke a spinally-mediated nocifensive flexor reflex of the ipsilateral limb [184, 273]. This nociceptive withdrawal reflex is mediated by a poly-synaptic connection of inter-neurons between nociceptive fibers and alpha-motorneurons. Activation of alpha-motorneurons results in the contraction of the limb flexors and relaxation of the limb extensors, causing a withdrawal response [273].

1.6.3.2 Descending modulation of spinal dorsal horn activity

Neuronal activity in the spinal dorsal horn is modulated through descending pathways stemming from the brainstem [283, 282]. Stimulation of the periaqueductal grey (PAG) induces endogenous analgesia by signalling to the raphe nuclei, which in turn activate inhibitory interneurons in the spinal dorsal horn [301, 283, 282]. The interneurons inhibit the central projections of nociceptors as well as the spinothalamic neurons in the spinal dorsal horn [283, 282]. Projections from the locus ceruleus act through the same mechanism [283, 282].

1.6.4 Supraspinal regions

The spinothalamic tract projects to the thalamus, which is in turn connected to various cortical and subcortical brain areas [47]. It is thought that activation at the cerebral level ultimately generates the experience of pain [308].

In functional magnetic resonance imaging (fMRI) studies, noxious stimulation induces a widespread pattern of brain activity. In a recent meta-analysis [357] of 222 fMRI studies of experimentally induced pain in healthy humans, four areas were consistently activated across a range of noxious stimulus modalities and locations. These are the thalamus, insula, midcingulate cortex (MCC, in some studies labelled as part of the anterior cingulate cortex (ACC)), and secondary somatosensory cortex (SII). Activity in the primary somatosensory cortex (SI), prefrontal cortex, amygdala, basal ganglia, primary motor cortex, supplementary motor area, cerebellum and brainstem was also frequently reported, but their activity varied more with the experimental paradigm [357]. The importance of the thalamus, insula, MCC and SII in sensory and emotional dimensions of pain is also apparent from studies in animals and neurosurgery patients [301]. The thalamus is the target of the spinothalamic tract and relays sensory information from the spinal cord to cortical areas [283, 282]. A study in monkeys showed that the insula, SII and areas in the cingulate sulcus are major onwards targets of the spinothalamic tract [81]. The insula is thought to play a pivotal role in pain processing [301]. For instance, direct stimulation of the insula and SII can elicit painful sensations in humans [211], lesions in the insula can trigger painful seizures [155] and the intensity of experienced tonic pain is coupled to insular blood flow [279]. The MCC is thought to integrate pain-related sensory, affective and cognitive input to contribute to behavioural responses [281].

The pain system is highly distributed [47]. Multiple brain areas are activated in parallel rather than in series [47] and sensory and limbic areas are activated in tandem [18]. While many (if not all) pain-responsive brain areas can also be activated by non-painful tactile stimuli [299], painful and non-painful responses can be distinguished by taking into account the magnitude [299] and spatial pattern of activity [196]. This approach is further discussed in Section 1.9.2.2.

1.6.5 Comparison to non-nociceptive touch

While this thesis focuses on pain and nociception, a vibrotactile stimulus is used as a control procedure in Chapter 3 and Chapter 4 and thus some important differences between the nociceptive system and the non-nociceptive somatosensory system should be highlighted.

Non-nociceptive touch, pressure and vibratory sensations are encoded by specialised non-neural receptors in the skin, including Meissner's corpuscles (sensitive to touch and vibration) and pacinian corpuscles (sensitive to high-frequency vibration). Signals are conducted by low-threshold Aβ and Aδ fibers [226]. Like nociceptors, the cell bodies of these fibers are located in the dorsal root ganglion, but their axons terminate in laminae III-V of the spinal dorsal horn [226]. The main ascending projections for fine touch and vibration are located in the ipsilateral dorsal columns, which decussate at the level of the brainstem to terminate in the contralateral thalamus. Some signals are also transmitted through the spinothalamic tract [283, 282]. In the thalamus, projection neurons synapse onto third-order neurons which extend to the primary somatosensory cortex and other cortical areas [283, 282].

1.6.6 Summary

Nociception involves highly specialised fibers in the periphery, which synapse onto ascending projections in the spinal dorsal horn to transmit nociceptive signals to a wide range of brain areas. The brain plays an active role in nociception and pain, both by modulating incoming signals in the dorsal horn and by ultimately generating the conscious pain experience.

1.7 Development of nociceptive pathways

Having outlined the physiology of the adult nociceptive system, the following section will focus on the developmental changes that occur in nociceptive pathways during fetal and early postnatal development.

1.7.1 Development of nociceptors and inhibition in the spinal dorsal horn

In mice, Aδ and C-fibers develop between embryonic day 10.5 and 11.5 (total gestation 19-20 days [212]) [189] and action potentials can be recorded *in vitro* from mechanosensitive dorsal root ganglion neurons not long after [192]. In the rat fetus, A- and C-fibers arising from the dorsal root ganglion innervate the skin on embryonic day 13-14 (total gestation 21.5 days) [156] and the central projections of A-fibers penetrate the dorsal horn soon after, followed several days later by the C-fibers [156]. These connections to the spinal cord allow the formation of sensory reflex loops [95]. Nociceptive reflexes can be observed in rat pups at birth [94] and in preterm human infants as young as 29 weeks PMA [53, 133].

In rodents, inhibitory tone in the spinal dorsal horn develops postnatally [95, 34]. In the superficial laminae of the spinal dorsal horn, low-threshold A-fibers become sparser while C-fiber input strengthens. Meanwhile, the function of the neurotransmitter gamma-aminobutyric acid (GABA) evolves from being both excitatory and inhibitory to being solely inhibitory [34], and descending modulation from the brainstem emerges [136]. These changes are thought to contribute to the pronounced age-related decrease in magnitude of nociceptive reflexes in rat pups [94] and in human infants [133, 53].

1.7.2 Development of thalamocortical projections and supraspinal areas

Ascending spinothalamic projections are present before birth in mice [58]. Thalamocortical connections start forming in the second trimester of gestation in human fetuses. Projection fibers from the thalamus initially synapse onto neurons located in the subplate [180], an intermediary structure that lies underneath the cortical plate [181] and reaches its maximal thickness around 30 weeks gestational age (GA) [327]. After a waiting period of several weeks, projection fibers from the subplate innervate their cortical targets in the somatosensory cortex around 26–28 weeks gestation, allowing signals transmitted through the spinothalamic tract to reach

the cortex [180]. After direct thalamocortical connections are formed between 31–34 weeks PMA, the subplate gradually dissolves from approximately 35 weeks gestation until the postnatal period [180, 179]. These anatomical findings suggest that nociceptive signals can reach the cortex in infants from 26–28 weeks PMA. Using functional infrared spectroscopy (fNIRS) and electroencephalography (EEG), cortical responses to noxious stimuli have been observed in infants of this age [288, 89]. Meanwhile, the cortical plate undergoes maturational changes between 31–34 weeks gestation to establish its adult-like configuration [180] and long cortico-cortical connections start forming from 30 weeks gestation [179]. At birth, an abundance of white matter connections are present, and many of these are pruned during the postnatal period [154].

The activity of the brain is also rapidly changing. Functional connectivity in the somatosensory resting state network can be observed in premature infants as young as 30 weeks PMA, although connections are weaker and the network is fragmented compared to that in term infants [77]. Cortical activity as measured with EEG also changes in nature [7]. In very preterm infants, the EEG is discontinuous and consists of high-amplitude bursts of activity interspersed with quiescent intervals [7]. Delta brushes form an important hallmark of the preterm EEG [346]. These activity bursts consist of a slow wave with superimposed high frequency activity. Delta brushes occur spontaneously as well as in response to various external stimuli [346, 219, 11]. They are not modality-specific and have been described in response to noxious stimuli and tactile stimuli [89], but also in response to auditory, visual and motor events [171, 219, 49]. Delta brushes generated by spontaneous body movements are thought to shape the mapping of the developing somatosensory cortex [219]. Rodent pups display a similar type of immature brain activity, termed spindle bursts [339]. Depriving the brain of peripheral somatosensory input decreases the occurrence of these spindle bursts, which further suggests that they are partly endogenously generated and partly evoked by sensory feedback [174]. The subplate is the hypothesised source of delta brush activity [175, 176]. The developmental period in which the subplate is present in human preterm neonates coincides with

the period during which delta brushes form an important part of the neonatal EEG [175, 172]. In addition, human fetal subplate cells generate spontaneous activity bursts *in vitro* [228], which has led authors to draw parallels with the intermittent bursting pattern of the neonatal EEG *in vivo* [171]. Importantly, in neonatal rats, the ablation of subplate neurons leads to the disappearance of spindle bursts [306]. This suggests that the subplate is essential in driving spindle bursts and their human equivalent. With increasing postmenstrual age, brain activity activity gradually becomes more continuous in nature, while its amplitude decreases [7] and its frequency content increases [240].

In term infants, a noxious pinprick stimulus that does not pierce the skin or cause behavioural distress, activates largely the same brain regions as in adults [349, 116]. The fMRI Neurologic Signature of Pain (NPS), which in adults is associated with stimulus intensity and verbal reports of pain intensity [333], is expressed by infants and graded with stimulus intensity [80]. Yet, an adult fMRI signature designed to explain additional variation in subjective pain reports related to factors such as motivation and expectation [352], is not expressed in infants [80]. This suggests that the term infant brain is capable of encoding nociceptive information, but that the experience associated with it may differ from that in adults.

1.7.3 Development of the processing of non-nociceptive stimuli

The circuitry for tactile processing equally undergoes widespread changes. A-fibers are abundantly present in the neonate [95], but the vibrotactile sensors in the skin and their afferents are not fully mature at term age [92]. Pacinian corpuscles are neurochemically mature at 36 weeks [92], but in neonatal cats, their afferents do not have the same responsiveness as in adults [93]. fMRI and EEG responses to tactile stimuli have been recorded in preterm infants, which confirms that tactile signals can be transmitted from the periphery to the cortex at this stage of development [89, 57].

1.7.4 Summary

Taken together, these findings from anatomical and functional studies indicate that the infant nervous system has the capacity to transmit noxious input from the periphery to the cerebral cortex well before term age. However, nervous system maturation continues into the postnatal period.

1.8 Reviewing measurements for infant pain assessment

As described in Section 1.5, ideally, there should be a causal relationship between a construct and the measurements taken to observe it. Without verbal report, it is difficult to assess what measurements would be considered downstream to the pain experience. However, based on the knowledge about the physiology of nociception and pain, a measurement may be more likely to be representative of pain if 1) similar measurements in adults are consistently related to verbal reports; and 2) the measurement is affected by cerebral cortex activity, which generates the experience of pain. Measures reflecting nociception are located upstream to experience, but can still provide useful information about the intensity of the signal that reaches the brain. With this in mind, we can qualitatively review the evidence base for commonly used surrogate measures of procedural pain in neonates.

1.8.1 Behavioural measurements

As described previously, noxious-evoked facial expressions form the basis of many pain assessment tools [51]. However, it is unclear if they reflect pain experience. In adults, facial expressions co-vary with the experimental modulation of the sensory and affective qualities of experimentally induced pain [186] and eye squeeze is linked to increased activity in the primary somatosensory cortex and decreased activity in the prefrontal cortex [185]. However, neonates with widespread parenchymal brain injury can still mount facial responses that are similar to their counterparts without brain damage [242]. This suggests that noxious-evoked facial expressions

in the infant are reflexes which are not (yet) under cerebral control [96] and may thus be reflective of nociception rather than pain. In addition, factors such as muscle weakness, lethargy and prematurity can suppress noxious-evoked facial movements [120], potentially without altering nociception. Indeed, in infants under stress, noxious-evoked brain responses (see Section 1.9.1.3) are dissociated from noxious-evoked facial expressions [167].

In contrast to facial expressions, nociceptive withdrawal reflexes have a clear biological substrate. Reflexes can be quantified by using electromyography (EMG) recorded from the flexors in the limb. In adults, the nociceptive reflex response has been used in studies of nociception for decades [273]. Electrically evoked nociceptive reflex threshold are related to subjective pain thresholds [347] and laser-evoked nociceptive reflexes are related to both heat intensity and subjective pain ratings [229]. In infants, nociceptive reflexes are evoked by noxious events [132, 133, 53] and are graded with stimulus intensity [132]. However, nociceptive reflexes can occur without input from the brain, and are therefore not likely to reflect experience [96]. The observation of nociceptive reflexes does, however, provide evidence that noxious signals are transmitted to the spinal dorsal horn [96]. As the composition and top-down modulation of the spinal dorsal horn undergoes rapid development during the preterm period and reflexes become progressively more refined [95], infant age should be taken into account when interpreting these responses.

1.8.2 Autonomic nervous system measurements (vital signs)

Noxious-evoked autonomic responses, such as changes in skin conductance and heart rate, have been widely observed in infants [267] and adults [187]. The autonomic nervous system (ANS) is integrated with nociceptive circuitry at the level of the spinal cord, brain stem and brain [54], which implies that these responses may be a consequence of nociception, i.e. reflexive, or modulated by activity in higher brain centres, i.e. potentially downstream to pain perception [220]. Behavioural studies in adults are inconclusive on this matter. For instance, a study of tonic heat pain reported that autonomic responses better encode stimulus intensity (nociception)

than verbal report (experience) [238], while other studies found that autonomic responses are mediated by pain perception [220] or correlated with both perceived affective and sensory intensity [197]. Autonomic nervous system tone is not fully developed in preterm infants [40], therefore infant age is likely to be an important moderator for the magnitude of ANS responses.

1.8.3 Summary

In summary, evidence from infant and adult studies suggests that behavioural and ANS measurements can be useful in the study of nociception. However, because it is uncertain if they are downstream to the pain experience, they are unlikely to be a comprehensive reflection of the construct of neonatal pain. For this reason, infant pain assessment should include measures derived directly from the brain.

1.9 Brain-derived measures of pain

In 2006, two independent studies [288, 16] used fNIRS to investigate cortical responses to a clinically required skin-breaking heel lance in infants. In both investigations, noxious-related haemodynamic changes were observed in the somatosensory cortex. These landmark studies demonstrated the feasibility of recording noxious-evoked brain activity in infants and paved the way for further investigations using EEG and functional MRI.

1.9.1 EEG-derived measures

EEG is a relatively low-cost, safe and portable tool which can be used to measure brain activity at the bed- or cot-side of a patient [7, 200]. In contrast to NIRS, which is dependent on cerebral haemodynamics, it directly measures neuronal activity. The temporal resolution of EEG is high, and allows the investigation of changes in brain activity that occur within tens of milliseconds after an event [200].

1.9.1.1 Physiological basis of EEG and ERP recordings

In brief, the EEG signal measured at the scalp is the sum of post-synaptic potentials in a large population of neurons which are activated at the same time and have a similar orientation [177]. It is expressed as the voltage between one or more recording electrodes anywhere on the scalp and a reference electrode in a different position [201]. Ongoing brain activity, such as the intermittent bursting pattern in preterm infants (see Section 1.7.2), can be detected at rest.

A temporal change in activity which is related to a (time-locked) stimulus is called an event-related potential (ERP) [201]. In adults, ERPs are generally investigated by applying tens to hundreds of stimuli and averaging across data epochs, so that ongoing brain oscillations, which have different phase across trials, are averaged out and the stimulus-related activity surfaces [201]. However, single-trial ERPs can also be recorded, which can be exploited in studies of clinical procedures, which are necessarily single events. Pyramidal cells in the cerebral cortex are the most likely source of ERPs, as these cells are oriented in such a way that summation of their voltage can take place [177, 201]. ERPs are often characterised by their polarity, topography, latency and amplitude to allow comparisons across studies or conditions. However, the potentials measured at the scalp are the summation of activity across one or more brain regions. Therefore these characteristics are not necessarily representative of the underlying neuronal activity [201]. Polarity is influenced by whether the post-synaptic potentials are excitatory or inhibitory, but also by the orientation of the neurons relative to the recording and reference electrodes [177, 201]. The summation of temporally overlapping activity from multiple sources can lead to the measurement of scalp potentials with different latency and amplitude compared to those of the neuronal activity in the individual sources [201]. This low spatial resolution is the main disadvantage of the ERP technique.

1.9.1.2 Adult noxious-evoked ERPs

In adults, ERPs can be reliably recorded in response to noxious laser and contact heat stimuli [194] as well as pinprick stimuli which probe Aδ fibers [203, 147]. The

adult noxious-evoked ERP at the vertex consists of a negative-positive complex (N2-P2). Latency to the P2 peak varies by stimulus modality and location (hand or foot), but is generally less than 500 ms post-stimulus [194, 147]. Source localisation EEG studies have placed the generator of the N2-P2 complex in the suprasylvian area and the ACC [106], both areas that are implicated in pain in fMRI studies [357, 301]. Noxious-related ERP amplitudes are related to subjective pain scores [149, 104, 304, 144], although expectations may differentially modulate subjective reports and ERPs [148, 239] and disrupt this relationship [148].

1.9.1.3 Infant noxious-evoked ERPs

Slater and colleagues [290] were the first to investigate noxious-evoked EEG potentials in infants. The authors recorded EEG activity in a group of infants aged 35–39 weeks PMA during a clinically required skin-breaking heel lance and a non-noxious tactile control procedure [290]. While both stimuli evoked an early potential between approximately 150-400 ms post-stimulus, only the noxious stimulus evoked a second negative-positive (N-P) complex around 400-600 ms post-stimulus, which was maximal at the central electrodes Cz and CPz [290]. In a follow-up study, EEG was recorded in a group of preterm and term infants [89]. While term infants displayed this noxious-evoked potential in response to a noxious clinical procedure, infants younger than 32 weeks PMA did not [89]. Instead, noxious-evoked responses consisted of delta brushes [89] (see Section 1.7.2). The likelihood of observing the noxious-evoked potential increased with PMA, while the likelihood of observing a delta brush response decreased with PMA [89]. The period between 35–37 weeks PMA appeared to be a transition period, during which either response was equally likely [89].

Since its discovery, the infant noxious-evoked ERP has been extensively investigated in preterm and term infants. The ERP is not only evoked by heel lances, but also by immunisations [330, 46] and mild experimental noxious pinprick stimuli which activate Aδ fibers without piercing the skin or causing behavioural distress [132], it is graded with stimulus intensity [132] and its magnitude is influenced

by PMA [133], mode of delivery [173], prematurity [286], skin-to-skin contact [168] and infection [45].

In order to standardise the measurement of this noxious-evoked potential, Hartley and colleagues [131] used principal component analysis (PCA) to derive a template of noxious-evoked EEG activity in a group of 18 term infants. This template can be projected to new data to obtain an estimate of the magnitude of an infant's noxious-evoked brain activity [131] (see Fig. 3.5, "Template"). It is designed for use in infants from 34 weeks PMA onwards, and is scaled so that a magnitude of 1 corresponds to the average noxious-evoked response to a heel lance in a term infant. Initial construct validity was established by demonstrating that its magnitude correlates with facial activity [131] and is higher for a noxious heel lance compared to other sensory events [131].

Small studies have shown that template magnitude is modulated by pain management interventions, such as local anaesthetic [131], gentle stroking [125] and paracetamol [46]. It was used as an outcome measure in a randomised controlled trial (RCT) investigating the efficacy of morphine for procedural pain [134], and is currently in use as an outcome measure in a RCT investigating the effects of parental touch on heel lance pain [44].

Yet, important knowledge gaps remain. First, while the template of noxious-evoked EEG activity can be used in clinical trials in neonates aged 34 weeks PMA and over, there is no equivalent measure for infants below 34 weeks. This is addressed in Chapter 3 and Chapter 4 in this thesis. Second, the neuronal activity that underlies the infant noxious-evoked ERP is poorly understood. A simultaneous EEG-fNIRS study found that noxious-evoked ERPs did not always coincide with an increase in blood flow in the somatosensory cortex, which suggests that the somatosensory cortex does not necessarily contribute to the ERP [328]. Better understanding of the spatial neural correlates of variability in infant ERP responses will be key to interpreting changes in ERP magnitude related to analgesic modulation. This is explored in Chapter 7.

1.9.1.4 Noxious-evoked spectral EEG changes

Noxious-evoked EEG responses can also be measured in the spectral domain. As described in Section 1.9.1.1, the averaging procedure used to identify ERPs, averages out ongoing brain oscillations. This means that any stimulus-induced changes in the frequency domain which are not phase-locked to the stimulus are not captured in the ERP, even though they may be informative [206]. Event-related spectral perturbations (ERSP) can be computed to capture these non-phase-locked brain dynamics [206]. The ERSP technique is based on the extraction of the power spectrum of the time-locked EEG epochs in short, overlapping sliding windows, for example by using wavelet transforms [206, 68, 118]. The result is a quantification of EEG power in a set of time-frequency pixels. Thus, while in the ERP technique background activity is "noise" which is removed by averaging, the ERSP technique focuses specifically on the impact of stimuli on ongoing brain activity. The interpretation of time-frequency analyses can be ambiguous, as the presence of change in power in a certain frequency can be a reflection of a true change in brain oscillations and neural processing, but also a short-lived voltage deflection [199].

In infants, noxious-evoked spectral EEG changes have been less well studied than ERPs. Fabrizi and colleagues [90] report that a heel lance evokes a sustained increase in delta and gamma power. In contrast, Norman and colleagues [241] report an increase in beta power (10-30 Hz) following skin-breaking blood tests and experimental noxious stimulation, although the authors attributed these changes to artefact. Further investigations would be of interest, as several adult studies report noxious-evoked changes in the EEG power spectrum which are related to verbal reports [90, 144, 277, 244, 310].

1.9.2 Functional MRI

Functional MRI is another approach by which noxious-evoked brain activity can be measured. MRI does not involve any ionising radiation and is a safe method to assess brain activity in infants [137, 178, 52]. Feeding and swaddling the infant prior to the scan encourages natural sleep, so that scanning can be carried out

without any sedative agents, and hearing protection is used to shield the infant from the loud noise made by the scanner [137, 178, 52]. Technical advances in data acquisition and analysis [99, 117] have led to significant improvements in data quality [20] (see Chapter 7 for more details).

1.9.2.1 Physiological basis of the BOLD signal

In brief, fMRI capitalises on the fact that oxygenated and de-oxygenated haemoglobin have different magnetic properties [107]. When groups of neurons are activated, the demand for oxygen increases, and oxygenated blood is channeled to the active area in a process called the haemodynamic response. This increase in oxygenated blood is the source of the fMRI blood-oxygen-dependent-level (BOLD) signal [107, 12]. Compared to EEG, the BOLD response is slow, in the range of several seconds [10]. Because it relies on the coupling between neurons and cerebral blood vessels [12], it is influenced by developmental changes in vasculature [182]. One of the benefits of fMRI is that it has high spatial resolution and allows the investigation of brain structures on the millimetre scale [162].

1.9.2.2 Adult fMRI signatures of pain

As previously described, fMRI has been used to investigate the brain areas that are activated in response to painful stimuli in adults [357] and infants. These discoveries have paved the way for the development of multivariate MRI signatures of pain [309, 351]. For instance, the Neurologic Signature of Pain is a multivariate spatial map that can be projected onto new data to predict subjective pain reports at the individual level [333]. Large-scale meta-analyses can also be used to derive signatures associated with pain. The online tool Neurosynth automatically conducts large-scale meta-analyses based on the presence of key words in published articles [358]. This tool was used to identify voxels that are preferentially activated in studies which contain the word “pain” compared to those that do not (`https://www.neurosynth.org/analyses/terms/pain/`). Such multivariate spatial maps can be used for the “neuronal decoding” of cognitive states.

1.9.2.3 Inferring cognitive states

Because clinical procedures cannot be conducted inside an MRI scanner, it is unlikely that this technique will be used for the measurement of pain-related brain activity in a clinical setting. However, it is an invaluable and unique tool because it provides a way to make inferences about infant experiences [80]. In fMRI, reverse inference is the process in which the engagement of a given cognitive process is inferred based on activity in a given brain region that has been associated with that cognitive process in previous publications [250]. Neuronal decoding is the extension of reverse inference to voxelwise patterns of activity [251]. Reverse inference can lead to spurious conclusions when misused [250]. A telling example is that of the neuroscientist James Fallon, who diagnosed himself as a psychopath based on an MRI scan of his brain, which according to him, had all the traits of psychopathy [41, 298]. The fact that most psychopaths have certain brain traits (forward inference) does not mean that everyone with those traits is a psychopath (reverse inference). For an individual, the a posteriori likelihood of psychopathy thus depends on the a priori likelihood as well as the specificity of the neuroimaging traits. The posterior likelihood of a cognitive process being engaged is higher when the pattern of activity is more specific or when the a priori likelihood of the cognitive state is increased [250]. In the James Fallon example, the a priori probability cannot be increased. However, in task-based fMRI studies this can be achieved by matching tasks between studies [146]. In other words, if a study found that the Neurosynth-derived "pain" map is activated when watching the prime minister give a press conference, then we cannot conclude that watching press conferences causes the spectator pain. However, if we found this set of brain regions to be active in a volunteer who is exposed to a noxious stimulus, then we can reasonably infer that this activation is related to pain, and not to another cognitive process. When reverse inference is used correctly, it can thus be a powerful tool to generate hypotheses [251, 146]. This framework can be applied to make inferences about the infant experience based on studies with matched experimental paradigms in adults [80].

Williams and colleagues [349] were the first to show that it was feasible to collect noxious-evoked fMRI data in infants. As clinical procedures cannot be conducted in the scanner environment, the authors applied an experimental noxious pinprick stimulus to the foot of a term infant. This stimulus activates Aδ fibers [203] without piercing the skin or causing behavioural distress [132]. This landmark study showed widespread brain activation, including the primary and secondary somatosensory cortex [349]. Goksan and colleagues recorded fMRI activity in a group of 10 term infants in response to pinprick stimuli applied to the foot and demonstrated that the spatial pattern of activation was remarkably similar to that of adults [116]. These findings were corroborated by two other studies. In one comparative study of adults and infants, the adult NPS signature was expressed in the infant brain and was graded with stimulus intensity [80]. In a different study, group BOLD responses to the pinprick stimulus in infants were correlated with a spatial pattern that was derived from "pain" Neurosynth (adult) meta-analysis maps, but not with spatial patterns associated with the terms "visual","arousal" and "attention" [21]. These results tentatively suggest that infants may also be capable of having a multidimensional subjective experience - although we will never be able to confirm this in the absence of verbal report.

1.9.3 Summary

EEG and fMRI can be used to quantify noxious-evoked brain activity in infants. The main benefits of EEG are its high temporal resolution and its practicality (i.e. it is low-cost and portable). The main disadvantage is its low spatial resolution. In contrast, fMRI has high spatial resolution which facilitates reverse inference through e.g. comparisons between adults and infants. Drawbacks are that it is dependent on brain haemodynamics, which can be influenced by non-neuronal factors, its cost, and that it cannot be used at the cot-side or during clinical procedures. As such, these techniques are complementary: EEG-derived measures appear to be most promising for clinical applications, whereas fMRI derived measures may be more important for the interpretability of noxious-evoked signals and their modulation.

1.10 Multimodal measurements

At the beginning of this chapter I outlined the concept of content validity and face validity. As pain is a multidimensional experience that is generated in the brain, pain assessment tools that only use behavioural and ANS measurements have lower face validity compared to measures derived from the brain. Still, without verbal report, it is impossible to determine whether even these measures are truly reflective of the experience of pain. This does not mean we should give up, or settle for pain assessment tools with low face validity. Instead, pain assessment should be based on measurements from multiple potentially relevant modalities (behavioural, ANS and cerebral) [232] and continuously improved based on enhanced understanding of the construct of pain in infants, particularly in those who are born preterm.

1.11 Thesis overview

This introduction highlighted that pain assessment in neonates is critical, but that many existing pain scales lack content validity. Brain-derived and multimodal measurements have higher face validity than behavioural or physiological measurements alone, but these are not yet fully understood. The overarching goal of this thesis is therefore to characterise infants' multimodal responses to noxious events, with a particular emphasis on cerebral responses, to improve multimodal neonatal pain assessment.

The research chapters in this thesis are divided over two parts, which each focus on one important research gap. First, while a template of noxious-evoked EEG activity was developed for use in neonates aged 34 weeks PMA and over, there is no equivalent measure for infants below 34 weeks. Characterising noxious-evoked EEG responses as well as ANS and behavioural responses in preterm infants is therefore vital. This is addressed in Part II. In Chapter 3, I establish whether premature infants' responses to noxious and non-noxious stimuli can be discriminated based on multimodal measurements, and in Chapter 4 I investigate how these responses develop with age. Second, the neuronal activity that underlies infant noxious-evoked

ERPs is poorly understood. Further understanding the spatial neural correlates of variability in infant ERP responses using fMRI will be key to interpreting changes in ERP magnitude related to analgesic modulation. In Part III I present a pilot study that relates noxious-evoked ERPs to noxious-evoked fMRI responses in term infants.

2
General methods

This chapter provides an overview of the datasets, clinical research procedures, recording methods, stimuli and basic principles of data analysis used in this thesis. More detailed procedures can be found in the corresponding results Chapters.

2.1 Retrospective datasets

In Chapter 3 and Chapter 4, three datasets were used: the Oxford Heel Lance Dataset, the UCL Dataset and the Immunisation dataset. Parts of the descriptions of the Oxford Heel Lance Dataset and the UCL Dataset in the following section were published in a scientific article, of which I am the first author, under a CC-BY license in *Cerebral Cortex* [321] and deposited as a preprint under a CC-BY license on *medRxiv* [322]. The published work was restructured to comply with the format of a thesis.

2.1.1 Oxford Heel Lance Dataset

The Oxford Heel Lance Dataset consists of data from 70 infants, selected from a local database which contains data collected by the Paediatric Neuroimaging Research Group at the John Radcliffe Hospital, Oxford University Hospitals (OUH) NHS Trust, UK, in research programmes investigating neural responses to

pain and other sensory stimuli in the developing infant brain. Ethical approval for these programmes was obtained from the National Research Ethics Service (REC references: 11/LO/350, 12/SC/0447 and 19/LO/1085) and written informed parental consent was obtained prior to infant enrolment. Studies conformed to the standards of the Declaration of Helsinki and Good Clinical Practice guidelines. Parts of this dataset are included in other publications ([133, 125, 120, 320, 46, 275, 45]).

2.1.1.1 Data inclusion criteria

For the current study, data inclusion criteria were: infants aged between 28 (inclusive) and 40 (non-inclusive) weeks PMA for whom 1.55 seconds of time-locked artefact-free EEG was recorded at the Cz electrode during a non-noxious control heel lance and a noxious heel lance (see Section 2.1.1.2 below). Exclusion criteria were: postnatal age (PNA) of 8 weeks or more, intraventricular haemorrhage (IVH) grade 3 or 4, hypoxic-ischaemic encephalopathy (HIE), prior use of the data in the development (but not validation) of the previously described standardised template of noxious-evoked EEG activity ([131], see Section 1.9.1.3), and if infants were part of the intervention group of a sub-study that specifically investigated the effect of a pain-relieving intervention (e.g., gentle touch or kangaroo care). Data for inclusion were identified as follows:

1. First, a database search on previous participants was performed to identify infants who had EEG recorded during a heel lance and control heel lance and who fit the demographic inclusion criteria and exclusion criteria. As some infants participated in multiple test occasions, only first test occasions were considered in this search. I only included data from the database where all required fields had been populated with data.

2. Next, metadata associated with the EEG recordings were reviewed to ensure the stimuli had been time-locked to the EEG, and that the Cz electrode was available.

3. For infants with time-locked EEG data at the Cz electrode, EEG data were filtered and epoched and artefacts were visually identified. As per the inclusion

criteria, only infants with at least 0.5 seconds artefact-free data before and at least 1.05 second artefact-free data after both the control heel lance and the heel lance stimulus were included. The following criteria were used to visually identify artefact-free trials: a stable baseline period (absence of voltage changes of more than 100 μV across the 0.5 seconds before the stimulus), no voltage over 200 μV at any point, and no substantial electrical artefact after filtering. The preterm EEG is characterised by periods of high-amplitude bursting which can exceed 100–200 μV and which complicates artefact recognition [7, 340]. To avoid the spurious rejection of traces from preterm infants, traces in which the above amplitude thresholds were exceeded were only rejected if visual inspection of the data suggested they were likely to contain substantial movement artefact. These criteria were chosen to ensure the exclusion of traces with severe artefact, while keeping data attrition as low as possible. A total of 168 control heel lance and heel lance traces from 84 eligible infants were reviewed, and 11 of these traces ($\approx$ 7% of traces, $\approx$ 13% of infants) contained artefact. Therefore these 11 infants were excluded. A rejection rate of 13% of infants lies between the artefact rejection rates for prospectively collected data in highly regulated clinical trials (12% of infants in [285] and 3% of infants in [134]) and previous exploratory studies (20% of heel lance trials in [125], 16% of heel lance trials in [275]) which was expected due to the nature of the dataset, which included data from exploratory studies.

4. The database contained a relatively high proportion of infants aged 36–36+6 weeks PMA compared to the other age groups. To prevent this age group from dominating the analysis, only 10 infants aged 36–36+6 weeks PMA were included. First, priority was given to data which were recorded using an automated detection interface [356] instead of a microphone (n = 4) to time-lock the stimuli to the EEG and then 10 study numbers were selected pseudo-randomly from the remaining pool of 13 study numbers before formal data analysis.

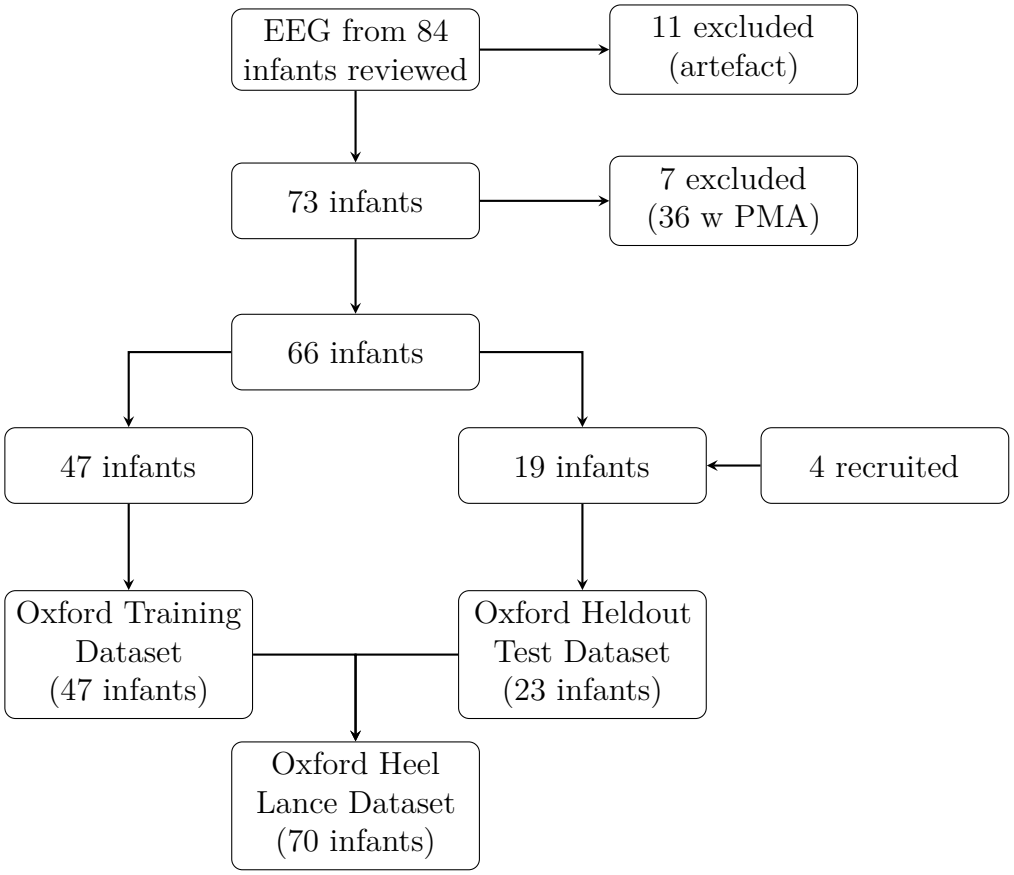


Figure 2.1: Overview of data included in the Oxford Heel Lance Dataset. To compile the Oxford Heel Lance Dataset, EEG data were reviewed for artefacts for infants who fit the demographic inclusion and exclusion criteria. This flowchart shows the number of infants for whom EEG data was reviewed, and the number of infants included in the Oxford Training Dataset and the Oxford Heldout Test Dataset, which make up the Oxford Heel Lance Dataset.

5. The Oxford Training Dataset (n = 47) was sampled from the Oxford Heel Lance Dataset for the purpose of training the classification model in Chapter 3. While I was analysing the Oxford Training Dataset, four infants fitting the inclusion criteria for the Oxford Heel Lance Dataset were recruited. These infants were therefore included in the Oxford Heldout Test Set (see Chapter 3 and Figure 2.1).

Figure 2.1 depicts the number of EEG traces reviewed and excluded.

2.1.1.2 Noxious and non-noxious stimuli

Two stimuli were used in the Oxford Heel Lance Dataset.

Heel lance (noxious) Heel lances are used clinically to obtain blood samples from neonates, for instance to monitor infection parameters or bilirubin levels. The lancet is placed against an infant's heel and upon activation a small blade is released that pierces the skin and allows for a capillary blood sample to be collected. In the studies described in this thesis, the heel lances were performed by a clinical member of the research team with a BD Microtainer Quikheel Infant Lancet (Becton, Dickinson and Company). After the heel lance, the foot was not squeezed for 30 seconds to allow the observation of facial expressions related to the skin-breaking aspect. All heel lances were clinically required; no heel lances were performed solely for research purposes. Comfort measures, such as swaddling or non-nutritive sucking, were used as clinically appropriate.

Control heel lance (non-noxious) When released, the heel lance produces vibrations and a clicking sound. Therefore, an experimental control heel lance procedure was implemented as the non-noxious control. During the control heel lance, the lancet is placed against the infant foot in a rotated position, so that the blade fires in the air and does not break the skin, while the vibratory and auditory components are matched to the heel lance.

Time-locking The control heel lance and heel lance were time-locked to the electrophysiological (EEG, EMG, and ECG) recordings with an accelerometer and an automated detection interface [356]. In two infants the stimuli were time-locked using a microphone that recorded the click produced by the lancet and was recorded along with the electrophysiological recordings.

2.1.1.3 Recording techniques

Electroencephalography (EEG) EEG activity was recorded with a SynAmps RT 64-channel headbox and amplifiers (Compumedics Neuroscan) in CURRYscan7 neuroimaging suite (Compumedics Neuroscan) using a sampling frequency of 2000 Hz. Activity was recorded from electrode position Cz and 3 to 24 other recording electrodes (Ambu Neuroline disposable Ag/AgCl cup electrodes) placed according

to a modified 10-20 system. The reference electrode was located at Fz and the ground electrode at Fpz / the forehead. In two infants, data were recorded with a reference electrode at Fpz and re-referenced to Fz offline, and in two different infants the sampling frequency was 1000 Hz and data were upsampled to 2000 Hz offline to conform with the rest of the dataset and the template of noxious-evoked EEG activity [131]. Before the electrodes were placed using EEG paste (Elefix EEG paste, Nihon Kohden), the skin was gently rubbed with preparation gel (Nuprep gel, D.O. Weaver and Co.). Electrocardiogram (ECG) was recorded with an electrode (Ambu Neuroline) below the left clavicle which was referenced to the EEG reference electrode.

Electromyography (EMG) EMG was recorded with surface bipolar electrodes (Ambu Neuroline) placed on the biceps femoris of both legs.

Behavioural measurements Infants were filmed for 15 seconds before a stimulus and 30 seconds after and the duration of brow bulge in the 30 seconds after the stimuli (see Section 1.4) was assessed offline by researchers trained in PIPP-R scoring. The moment of the stimulus was indicated by a LED-light that was manually activated at the time of the stimulus by one of the attending researchers.

2.1.2 Oxford Immunisation Dataset

The Immunisation Dataset was derived from a dataset previously published in [46] and contains cerebral, autonomic and behavioural data from preterm-born infants receiving routine vaccinations in the neonatal unit. These data were collected between 2017 and 2020 as part of the "Pain in the developing brain" research programme in the Paediatric Neuroimaging Group. The study was approved by the National Research Ethics Service (reference: 12/SC/0447) and informed written parental consent was obtained before each study. Studies conformed to the standards of the Declaration of Helsinki and Good Clinical Practice guidelines.

The Oxford Immunisation Dataset was collated by searching the local research database for all previous participants aged between 28 and 40 weeks PMA for whom

time-locked EEG was recorded during routine immunisations in the neonatal unit (n = 27). Only first test occasions were considered. EEG data were visualised to identify artefacts (see Section 2.1.1.1) and only infants with at least 1.6 seconds of time-locked artefact-free EEG data for at least one immunisation at the Cz electrode were included (n = 26). Some infants in the immunisation dataset were also studied on a separate test occasion that formed part of the Oxford Heel Lance Dataset.

EEG, ECG and behavioural data were collected in the same way as for the Oxford Heel Lance Dataset. The skin-breaking immunisation was time-locked to the EEG using a high-speed video camera [46].

2.1.3 UCL Dataset

The UCL Dataset was extracted from a UK Data Service repository [166, 169]. This repository was compiled by Jones (University College London (UCL)) and colleagues as part of a study investigating neonatal pain processing, and is made available upon reasonable request to any researcher or health professional. The UCL study was approved by the NHS Health Research Authority and conformed to the declaration of Helsinki [169]. Written parental consent was obtained before each test occasion [169]. Further details on research governance can be found in the publication associated with the data repository [169].

The repository contains demographical information as well as behavioural, EEG and ECG data for 112 infants during a heel lance, control heel lance and auditory stimulus. These data were collected by a different group of researchers at a different research site, but using the same techniques as the Oxford dataset. A total of 74 infants in the repository had EEG recordings at Cz during both a control heel lance and a heel lance and were aged between 28 and 40 weeks PMA, and less than 8 weeks PNA. EEG traces were reviewed, and no substantive artefacts were identified. Therefore these 74 infants were included in the analyses in this thesis as the "UCL Dataset". Of these infants, 9 had IVH of unknown severity. Comfort methods such as swaddling and kangaroo care were used for each test occasion.

2.2 Prospective datasets

For Chapter 6 and Chapter 7, retrospective EEG and fMRI data from three infants was combined with prospectively collected EEG and fMRI data in four infants and six adults. Recruitment of adults and infants to the studies was stopped due to the COVID-19 pandemic in March 2020.

2.2.1 Experimental noxious stimuli

As clinical procedures cannot be performed in the MRI environment, experimental noxious pinprick stimuli were used in the infant and adult MRI studies. The experimental noxious stimuli used for the studies in this thesis were PinPrick stimulators (MRC Systems GmbH), which are punctate probes with a flat tip that activate nociceptive Aδ fibers [147, 203] without piercing the skin. The PinPrick stimulators are designed to deliver a consistent force. In term neonates, stimulation with the 128 mN pinprick does not cause behavioural distress or tissue damage [132]. As such, the 128 mN PinPrick stimulator provides a safe and ethical way to probe the infant nociceptive system. In the adult experiments in this thesis, forces of 32 to 512 mN were applied, while in infants only the 128 mN PinPrick was used.

2.2.2 Infant EEG-MRI study

Prior to me joining the research group, three infants had participated in both an EEG and an fMRI test occasion as part of the "Pain in the Developing Brain" research programme in the Paediatric Neuroimaging Group. Together with other members in the research team, I prospectively collected EEG and fMRI data from a further four infants (of these, one EEG test occasion was performed by other team members as part of a different sub-study). The procedures below refer to the collection of the prospective dataset, but were also applied in previous studies, unless otherwise indicated. Ethical approval for these studies was obtained from the National Research Ethics Service (references: 12/SC/0447) and informed written parental consent was obtained before each study. Studies conformed to the standards of the Declaration of Helsinki and Good Clinical Practice guidelines.

2.2.2.1 Recruitment

Infants were recruited between 2019 and March 2020 from the neonatal unit and the postnatal wards of the John Radcliffe Hospital, Oxford University Hospitals NHS Trust, UK. Inclusion criteria were infants born to a mother over 16 years of age, with a PMA of 36–44 weeks at study, who were self-ventilating in air and needed a heel lance for their clinical care [1]. Exclusion criteria were clinical instability, contra-indications for MRI (including treatment that was incompatible with MRI, e.g. phototherapy or feeding through a nasogastric tube), neurological anomalies, IUGR (< 3rd percentile), mother regular user of recreational drugs or opioids and any social, psychological or other factors that would make it inappropriate to approach parents as judged by the midwife caring for the baby.

The studies were planned for days that the MRI scanner was bookable and that trained staff (a clinical member of the research team with Neonatal Life Support training, two experienced researchers and a radiographer) were available. Eligible infants were identified in the days leading up to the potential MRI study day. This required some coordination, as it was also necessary for a dedicated team (a clinical member of the research team who could perform the blood test and two specially trained researchers) to be available at the time of the clinically required blood test.

After obtaining confirmation from the clinical care team that it was appropriate to approach, parents of newborn infants were approached and asked if they would be interested in hearing about a research study. The study procedures and stimuli were explained, and where applicable demonstrated, and parents were given an opportunity to ask questions. If parents expressed interest, they were given information leaflets detailing the EEG and MRI part of the study (see Appendix A) and given an appropriate amount of time to decide upon participation. If they agreed to take part, parental written informed consent was obtained prior to any study procedures. During the consent process, parents were informed that they

[1] Initially, I followed a recruitment strategy where parents were approached for an ongoing fMRI study, with the plan to re-approach (with parental consent) for an additional EEG study if their child subsequently needed a blood test for clinical purposes. However, it became apparent that this was not feasible in practice. Therefore, clinically requiring a heel lance was added as an inclusion criterion.

could withdraw at any point and that this would not affect their baby's clinical care. Prior to each session, the attending parent was verbally asked if they continued to be happy to take part. The sessions were planned so that they did not interfere with the baby's clinical care.

The three infants who were recruited prior to the start of my DPhil fit the inclusion and exclusion criteria detailed above.

Impact of the COVID-19 pandemic Data collection for this project was severely impacted by the COVID-19 pandemic. Recruitment for the overarching "Pain in the Developing Brain" research programme was paused between March 2020 and September 2020 and again in January 2021 as a necessary precaution to protect patients and staff. As the infant MRI studies involved additional complexities, such as the transport of infants from the postnatal ward to the MRI scanner, as well as the necessity of specialised clinical and research staff, recruitment for infant MRI studies in my research group only resumed at the end of 2021 (with the first scan conducted in April 2022), and I was not able to collect more data for my thesis.

2.2.2.2 EEG session

During the EEG session, EEG, EMG, ECG and behavioural responses were recorded around a heel lance and a control heel lance as described above for the Oxford Heel Lance Dataset. Parents were invited to attend the session.

2.2.2.3 MRI session

A process for MRI scanning of newborn infants was developed by the Paediatric Neuroimaging Group prior to me joining [116], and the same processes were followed in the current study.

Infant preparation and safety Only infants who were clinically stable were recruited to MRI studies. Scanning was carried out in the presence of at least one radiographer, one clinical member of staff and one experienced researcher. On the day of study, infants were transported from the postnatal wards to the Wellcome

Centre for Integrative Neuroimaging (WIN) in a pram by a clinical member of the research team trained in Neonatal Life Support, and a specially trained researcher. The research team carried an oxygen cylinder and neonatal resuscitation equipment at all times. Before entering the scanning suite, infants were screened for metal (including metal poppers on clothing that were in direct contact with skin or near the head) and other contra-indications for MRI by the clinical member of staff. The radiographer carried out safety screening with parents. Infants were fed and swaddled in blankets to encourage natural sleep. No sedatives were used at any point. As the scanner makes a loud repetitive noise, three levels of ear protection were used: mouldable ear putty, ear muffs (MiniMuffs, Natus Medical) and MRI-safe ear defenders (Baby Ear Defenders, Em's 4 kids). In the scanner, infants were placed on a vacuum-position mattress positioned within a foam cradle to restrict motion. If necessary, extra padding was applied around the head to ensure a comfortable but good seal between the ear defenders and the head and to further restrict head motion. An MRI-compatible neonatal oxygen saturation probe was placed around the right foot. Scanning only commenced if vital signs were normal and the infant was settled. During the scan, heart rate and oxygen saturation were continuously monitored by the clinician. If infants became unsettled during the study, or if vital signs were outside normal bounds, scanning was paused to attend to the infant's needs. Scanning could then be resumed if deemed appropriate by the clinician, research team and parents, or terminated. At the end of the study, the infant was transported back to the ward by the clinical member of staff. Parents were invited to accompany their child to WIN, and one parent was invited to sit in the scanning suite during the scan. If they chose to do so, they were screened for MRI safety by a radiographer before entering the scanner room and were required to wear hearing protection for the duration of the scan. If they did not wish to enter the scanner room, they resided in a separate area close to the control room.

Experimental design MRI data were acquired on a Siemens Prisma 3T with a 32-channel adult head coil. Acquisitions included a structural T2-weighted scan, a

fieldmap scan and a stimulus response scan. During the stimulus-response scans, a train of ten 128 mN pinprick stimuli were applied to the dorsum of the left foot. The stimuli were applied when the infant was naturally still with an inter-stimulus interval of at least 25 seconds. If time permitted, ten minutes of resting state fMRI and diffusion-weighted MRI data were also collected. These data were not analysed in this thesis, but were collected for future analyses.

Acquisition parameters Infant MRI data were collected with the following parameters:

Structural T2 weighted, TSE factor 11, TE 89 ms, TR 14740 ms, 1 mm isotropic voxels, 150 degree flip angle, parallel imaging GRAPPA 3;

Functional T2* BOLD-weighted, GRE, EPI readout, TE 50 ms [117], TR 1300 ms, 2 mm isotropic voxels, multiband 4, 70 degree flip angle;

Fieldmap GRE, 2DFT readout, dual echo TE1/TE2 4.92/7.38 ms, TR 550 ms, 2 mm isotropic voxels, 46 degree flip angle.

2.2.3 Adult EEG-MRI study

This section outlines the technical details of the adult EEG–MRI study. Experimental design is specified in Chapter 6. The study was approved by the Oxford Central University Research Ethics Committee (CUREC) and written informed consent was obtained from participants before any study procedures (see Appendix B for the information leaflet and consent form used in the study). Inclusion criteria were healthy adults aged 18 to 60 years of age who were fluent in English. Exclusion criteria were any MRI contraindications, chronic pain or use of analgesics in the 6 hours leading up to the study. Adults were recruited through word of mouth and approved advertisement materials.

2.2.3.1 EEG session

EEG data were recorded from 21 electrodes (Cz, FCz, Pz, Oz, Fp1, Fp2, F3, F4, F7, F8, A1, A2, T3, T4, T5, T6, C3, C4, P3, P4, O1, O2) placed according to the 10-20 system with a reference at Fz and a ground electrode on the forehead. Disposable Ag/AgCl cup electrodes (Ambu Neuroline) were used and to ensure good contact with the skin, the scalp was gently cleaned with preparation gel (Nuprep gel, D.O. Weaver and Co.) before application with conductive paste (Elefix EEG paste, 8 Nihon Kohden). ECG was recorded alongside the scalp electrodes using an electrode above the left clavicle. During the recording, pinprick stimuli of varying intensities were applied to the dorsum of the left foot (see Chapter 6 for details).

2.2.3.2 MRI session

Adult MRI data were collected with the following parameters:

Structural T1 weighted, TE 3.97 ms, TR 1900 ms, 8 degree flip angle, 1 mm isotropic voxels;

Functional T2* BOLD-weighted, multiband 8, TE 39 ms, TR 735 ms, 52 degree flip angle, 2.4 mm isotropic voxels;

Fieldmap GRE, dual echo TE1/TE2 4.92/7.38 ms, TR 482 ms, 46 degree flip angle, 2 mm isotropic voxels.

Resting state fMRI and diffusion-weighted MRI data were also collected for future analyses, but were not analysed in this thesis.

2.3 Data analysis principles

Several preprocessing steps are required to extract noxious-evoked responses from raw EEG and fMRI data. This section gives a general overview of preprocessing steps – the methods sections in Chapter 3, Chapter 6, and Chapter 7 provide more detailed information.

2.3.1 EEG

EEG recordings contain neuronal signals as well as noise. For example, slow drifts in the EEG arise through changes in skin hydration or electrode movement, while muscle activity (e.g. due to sucking in infants) and line noise cause high-frequency artefacts [198]. In the studies in this thesis, a high-pass filter, low-pass filter and 50 Hz notch filter were used to remove signal at these frequencies [199]. Independent Component Analysis (ICA) was used to remove further artefacts, such as eye blinks, in the adult recordings [198, 68], but not in the infant recordings, as only few electrodes were available and ICA is not as well developed for infant EEG data. For both ERP and ERSP analysis, the continuous data was split into epochs around each stimulus. For ERP analysis, epochs were baseline-corrected by subtracting the mean activity in the pre-stimulus period from the activity in the post-stimulus period to remove voltage offset from 0 [198]. Individual epochs were visually inspected and rejected if residual artefact was present. Fig. 2.2 shows the effects of filtering, epoching and baseline correction on an example trial.

2.3.2 fMRI

Adult and infant fMRI data require slightly different analysis pipelines which are described in Chapter 6 and Chapter 7, respectively. Generally, preprocessing of single-subject fMRI data included distortion correction to correct for geometric distortions caused by inhomogeneities in the magnetic field, motion correction to remove artefacts due to head motion, spatial smoothing to increase signal-to-noise ratio (SNR), temporal filtering to remove slow drifts caused by for instance heart rate, and grand mean scaling to account for differences in mean signal across participants [163]. In this thesis, spatial ICA was used to remove additional noise components related to for instance head motion or physiology [23, 121, 122, 272]. Functional data were registered to age-appropriate anatomical templates using participants' structural scans as intermediate warps [163]. For statistical analysis, the time course of the pinprick stimuli was first convolved with an age-appropriate haemodynamic response function (HRF) to model the expected BOLD signal in

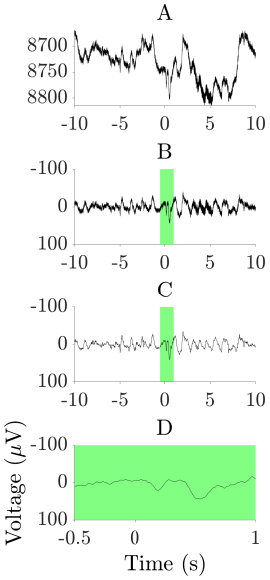


Figure 2.2: Infant EEG data were preprocessed in several steps. This example shows 20 seconds of EEG data from a single infant from the Oxford Dataset, where an event occurred at time = 0. A) Raw EEG data contains high-frequency artefact and slow drifts. B) The EEG data are high-pass filtered to remove slow drift. Green shading indicates the epoch that will be extracted for ERP analysis. C) A low-pass filter is applied to remove high-frequency artefact. Green shading indicates the epoch that will be extracted for ERP analysis. D) The filtered data is epoched around the event and the baseline of the epoch is subtracted.

each voxel [163]. A general linear model (GLM) was then fit to the observed BOLD time course data in each voxel using the expected BOLD signal as a regressor, and a contrast was set up to test in which voxels the GLM parameter estimates were positive (i.e., which voxels were activated in concordance with the pinprick stimuli, relative to baseline) [161]. The result of this statistical analysis is a spatial map of GLM contrast parameter estimates (COPE), which contains a GLM parameter estimate for each voxel and can be interpreted as a spatial amplitude response map [161]. The COPEs from individual participants were used to set up group-level analyses, which are described in Chapter 6 and Chapter 7.

2.4 Software

Software tools used for analysis will be reported as appropriate throughout this thesis. All data plots, except the fMRI figures in Chapter 6 and Chapter 7, were

generated in MATLAB (Mathworks, version 2019a). Colour schemes for these plots were generated using the function *brewermap* [295], shaded error bars on plots were created with the function *shadedErrorBar* [36] and the "raincloud plots" were created with the function *raincloud plot* [2].

A short history of the Data Task Force

The Paediatric Neuroimaging Group has collected neuroimaging, behavioural and physiological data from over 600 infants across tens of sub-studies in the past ten years. Each participant is assigned a unique study number upon enrolment in the study, which is used to label all collected data. Demographic information, such as infant age and sex, and clinical pain scores were collected by using bespoke data collection forms. While these forms were standardised, they were completed on paper, which meant that retrospective data searches were time-consuming and had to be performed on-site. All electronic EEG and physiological data were kept on a departmental drive using a standardised naming system for files and folders, however small variations were present in file names and event labels linked to the EEG data. For instance, annotations on the EEG indicating the time point at which a heel lance occurred could be written as *Heel Lance*, *heellance* or any variation thereof. While these annotations are easily interpretable by humans, they are not by machines, rendering the process of pooling data highly time-consuming. In 2019, a secure electronic database was introduced to securely store demographic information. All prospectively obtained demographic information was entered directly into the database from thereon, and in a collective effort, all current team members participated in entering data from paper forms to the database retrospectively. This tremendously increased the feasibility of performing searches on available data, and proved invaluable when on-site working was prohibited during the COVID-19 lockdown in England between March and September 2020. The COVID-19 lockdown was an opportunity to further improve data management practices. In March 2020, several colleagues and I spearheaded the Data Task Force: a team that would lead the standardisation of all our electronic data in order to improve the process for retrospective data analysis and data pooling. We designed a standardised naming system for all electronic files and for all labels appended to the EEG and phsyiological data. In brief, events were labelled by the type of stimulus (e.g. *heellance* or *immunisation*), the intensity and location of the stimulus (e.g. *thigh*), if applicable, and any relevant factors that could affect the response (e.g. *paracetamol*). Over the course of several weeks, the Data Task Force outlined this plan to the rest of

the group, and the entire research team, from students to clinical members to senior academics, participated in the effort to standardise all file names and labels for every recording stored on our group's drive. In addition, guides were produced to describe in detail the standardised naming system for future data collection. The standardised naming system combined with the electronic database containing demographic information and clinical pain scores was key to being able to compile and extract the Oxford Heel Lance Dataset and the Oxford Immunisation Dataset for this thesis. All the while, myself and others had started working on another problem that had arisen from the COVID-19 lockdown. The software that was previously used to load and epoch raw EEG data files from our acquisition computer was only accessible from a central computer located on-site. I explored alternatives and, using open-source software (Brainstorm [300] and EEGlab [68]), devised a pipeline which could import our raw EEG data into MATLAB for further processing. I then explored ways of integrating the research group's existing EEG processing scripts (such as the scripts devised and shared by Hartley and colleagues [131] to project the template of noxious-evoked EEG activity to EEG traces) with the new data format to generate the EEG, ECG and EMG analysis pipelines used in this thesis. My colleague Caroline Hartley went through the hard work of adapting and expanding my scripts to make them generalisable to the wider lab, and these scripts are now being used for new projects, including a clinical trial [44]. Upon completion of all retrospective data standardisation, the Data Task Force was informally disbanded, although it remains in hibernation, ready to spring back into action when the next challenge arises.

Part II

Investigating multimodal noxious-evoked responses in premature infants

3

Discrimination of responses to noxious and non-noxious stimuli

Part of the work in this chapter and Chapter 4 was published under a CC-BY license in *Cerebral Cortex* in a paper on which I am the first author [321]. A preprint was also deposited under a CC-BY license on *medRxiv* [322]. The published work was restructured to comply with the format of a thesis and expansions were made throughout the introduction, methods, results and discussion sections.

3.1 Introduction

As described in Chapter 1, term infants show a wide range of responses to noxious stimulation. At the behavioural level, this includes facial grimacing [123, 73] and withdrawal of both the affected and unaffected limb [132, 53]; at the level of the autonomic nervous system, a rise in heart rate [343] and an increase in skin conduction [86] are observed; and at the neural level, a distinct noxious-evoked ERP is present after noxious stimulation, but not after other sensory events [290, 131]. Quantification of these responses forms the cornerstone of infant pain assessment [232]. For example, the commonly used PIPP-revised (PIPP-R) score utilises changes in pain-related facial expressions and physiological parameters to generate an indicator of an infant's pain state [296], and the derivation of the

template of noxious-evoked EEG activity [131] (also see Section 1.9.1.3) facilitates the measurement of noxious-related neural activity in clinical trials in (near-)term infants [285, 131, 44, 134].

In preterm infants, who are faced with the largest burden of noxious clinical procedures [38], several lines of evidence suggest that noxious-evoked responses across multiple modalities are less robustly expressed or are less specific to nociception than in term infants [280, 260], which poses a challenge for pain assessment. For example, in very preterm infants, noxious-evoked facial expressions [56, 164, 120, 112, 113] and cardiovascular responses [56, 253] are dampened and facial grimacing [120] and withdrawal reflexes [53] can be evoked by noxious and tactile stimuli alike. The likelihood of observing the characteristic noxious-evoked ERP response in EEG recordings is inversely related to infant age [89] and infants below 34 weeks PMA are more likely to display delta brush responses to both noxious and non-noxious stimuli [89] (also see Section 1.9.1.3). Developmental studies in rodents provide converging evidence. For instance, in rat pups, cutaneous reflexes to low-intensity stimulation are large and disorganised [97]. In addition, differential neural oscillations in the somatosensory cortex to C-fiber and A-fiber stimulation does not occur in rat pups at birth, but develops across the first weeks of life [42].

It can be tempting to infer from these observations that the preterm infant nervous system simply is not mature enough to generate discriminable responses between noxious and non-noxious events. This conclusion is too simplistic, both from a statistical and a biological point of view. Statistically this claim cannot be verified, as it would require us to measure *all possible* noxious-evoked nervous system responses. In reality, we will never be able to get even close to achieving this. Especially in comparison to pain studies in adults, which employ all modalities under investigation in infants in addition to more complex and invasive techniques, such as microneurography, intra-cranial EEG and positron emission tomography (PET) imaging [309], and targeted experimental designs with large sample sizes which allow more sophisticated analyses, relatively few pain-related responses can (and should) be ethically investigated in infants. Therefore, it cannot be excluded that preterm

infants display discriminable responses, but that we simply do not have the right techniques to measure them. At the same time, the state of the preterm nociceptive nervous system does not seem to preclude the ability of infants to differentially respond to noxious and non-noxious events and some findings even carefully hint at this ability. As touched upon in Chapter 1, at least part of the anatomical circuitry necessary for nociception is already present in early development [329]. For instance, A-fibers are functional in newborn rodents [97], and preliminary thalamo-cortical connections start to form as early as 26 weeks gestation in human fetuses [180]. Functionally, precursors of adult resting networks can be identified in the preterm infant brain [77], which is relevant because the strength of these networks predicts noxious-evoked fMRI responses in term infants [21]. In addition, while delta brush responses in preterm infants can be evoked by non-noxious stimuli, their occurrence is more frequent to noxious stimuli [89], which suggests that the capability of intensity-encoding is already present at an early stage of development [329].

Thus, the reported lack of preterm infant responsiveness to noxious stimuli may well be the result of limitations in our measurement techniques and new methods of pain assessment should be explored. A multimodal approach, which allows the investigation of multivariate patterns of noxious-evoked activity, might improve pain assessment. As a hypothetical example; an isolated moderate increase in heart rate might not be specific to a noxious stimulus, whereas the interaction of a heart rate increase and the presence of noxious-evoked EEG activity could be. Conceptually a multimodal approach to pain assessment has face validity because it fits with the notion that pain is a multi-sensory experience [258] (also see Chapter 1). In addition, each modality described above has different characteristics and measurement errors, and therefore a combination of measurements may be more informative than any of the modalities alone [232]. For instance, EEG recordings may suffer from movement artefacts [139] and low signal-to-noise ratio when using single trials [30], while behaviour and heart rate may be modulated by sleep state [260]. In practice, multimodal approaches improved the discrimination of noxious and non-noxious responses in infants from 34 weeks PMA [320] and

the discrimination between high and low pain scores in a group of infants with an average gestational age of 36 weeks [360].

The use of multiple modalities requires decisions on the integration of the individual components into a final score. This may be achieved by either *decision-level fusion*, where each modality is evaluated independently and the final score is determined by the merger of each model's prediction, or by *feature-level fusion*, where modalities are combined within a single model [359]. The advantage of the latter is that the different modalities are automatically weighted in a data-driven manner. In addition, the model will be sensitive to interactions between modalities and PMA can be automatically incorporated in the model to derive a developmentally-sensitive tool. This would be an improvement over existing pain scales, which are often either developed for a specific age group [51] or which account for age-related differences in responses by inflating scores for younger infants [297].

In this chapter I developed a classification model that predicts exposure to a noxious heel lance or a non-noxious control stimulus across a group of preterm and term infants, with the aim to:

1. Investigate whether infants aged 28–40 weeks PMA display a discriminable multimodal pattern of responses to noxious and non-noxious stimulation;

2. Explore future applications of such a multimodal classification model in clinical trials.

3.2 Methods

3.2.1 Study design

In the Oxford Training Dataset, features were extracted from behavioural, autonomic and cerebral recordings from a group of infants aged 28–40 weeks PMA undergoing a noxious and non-noxious stimulus. These were used to train a classification model which discriminates responses to the noxious stimulus from responses to the non-noxious stimulus. The generalisability of the resulting model was then tested in two independent datasets: the Oxford Heldout Test Dataset (collected in Oxford, used

to test generalisability across datasets within site) and the UCL Dataset (collected at a different research site, used to test generalisability across sites). Finally, the Immunisation Dataset (collected in Oxford) was used to investigate the model's generalisability to a different stimulus, and to investigate whether model scores were sensitive to modulation by paracetamol.

A graphical representation of the study design can be found in Fig. 3.1.

3.2.2 Datasets

For the primary model development and assessment, infants aged 28–40 weeks PMA were included in the following three datasets:

1. Oxford Training Dataset (n = 47): 67 % of infants from the Oxford Heel Lance Dataset (see Chapter 2, Section 2.1.1 for details). This subgroup was drawn at random using a computer-based strategy that allowed for stratification of PMA.

2. Oxford Held-out Test Set (n = 23): the remaining 33% of infants from the Oxford Heel Lance Dataset.

3. UCL Dataset (n = 74): a dataset collected at a different research site ([169, 166], see Chapter 2, Section 2.1.3 for details).

Fig. 3.2 and Table 3.1 summarise the demographic information for infants in the Oxford and UCL datasets.

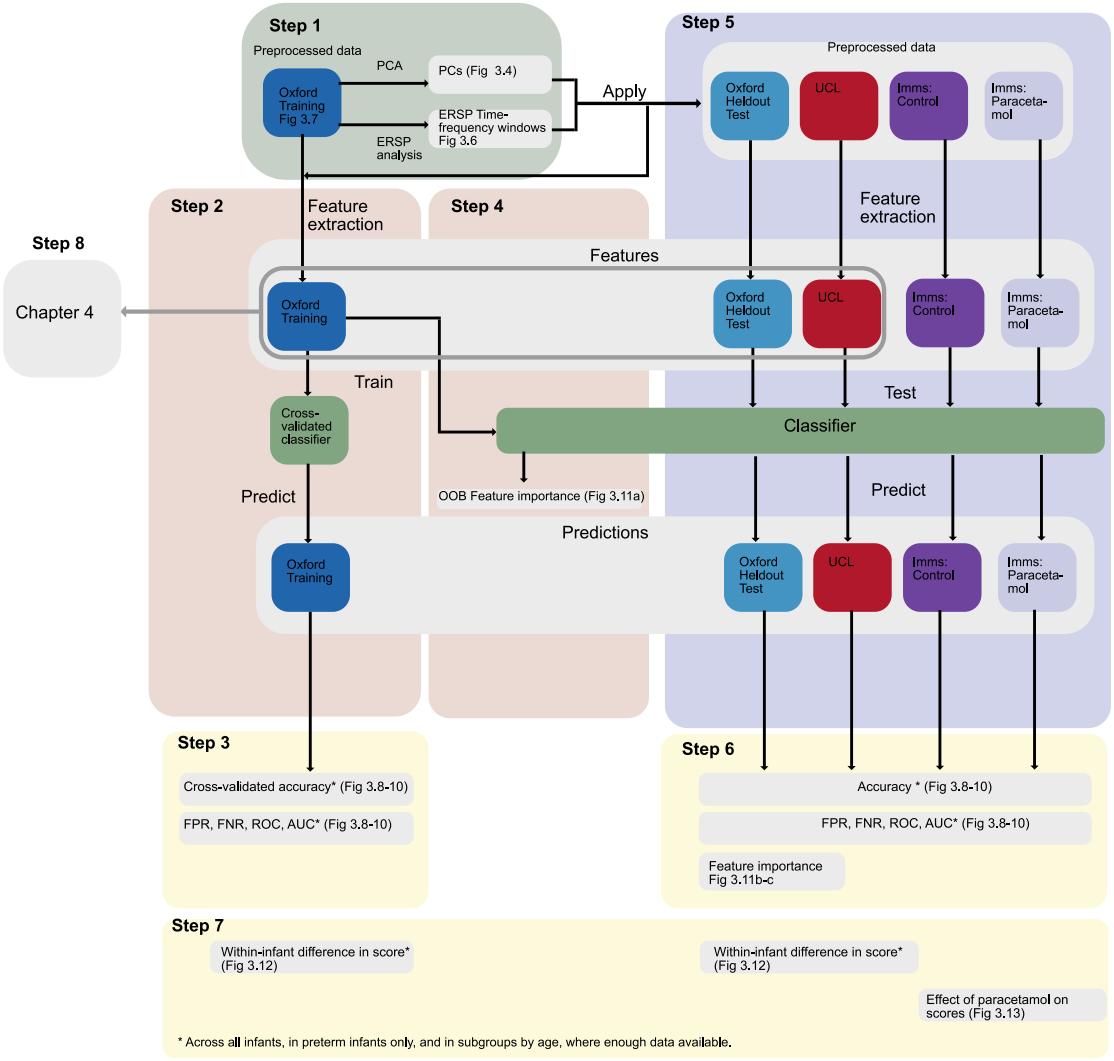


Figure 3.1: Graphical representation of Chapter 3. Step 1: EEG feature design. The Oxford Training Dataset (Fig. 3.7, Section 3.3.1) was used to identify (1) the PCs in the EEG data (Fig. 3.4) using PCA; and (2) ERSP time-frequency windows of interest (Fig. 3.6). Step 2: Classifier development. Cerebral, autonomic and behavioural features were extracted from the Oxford Training Dataset (see Table 3.2), using the PCs and ERSP time-frequency windows identified in Step 1. These features were used to train a cross-validated classification model. Step 3: Classifier performance in the Oxford Training Dataset. Cross-validated predicted scores and classes were used to estimate the accuracy, FPR, FNR, and AUC of the ROC curve in the Oxford Training Dataset, across all the data (Section 3.3.2, Fig. 3.8), and in sub-groups (Section 3.3.4, Fig. 3.9, Fig. 3.10). Step 4: The model was retrained on all the data in the Oxford Training Dataset and feature importance in the Oxford Training Dataset was estimated (Section 3.3.5, Fig. 3.11A). Step 5: Testing the classifier. Features were extracted from the Oxford Held-out Test Dataset, UCL dataset and Immunisation dataset. These were input to the classifier to generate predicted scores and classes for all observations in these datasets. Step 6: Classifier performance in independent datasets. The predicted scores and classes were used to calculate accuracy and other performance metrics in each of the datasets (Section 3.3.3, Section 3.3.6.2, Fig. 3.8, Fig. 3.9, Fig. 3.10). Feature importance was also calculated for the Oxford Heldout Dataset (Fig. 3.11B-C). Step 7: Examples of classifier applications. The predictions were also used to explore within-infant discrimination (Section 3.3.6.1, Fig. 3.12) and the effects of a pharmacological intervention (Section 3.3.6.2, Fig. 3.13). Step 8: The features from the Oxford Training Dataset, Oxford Heldout Test Set and UCL Dataset were carried over to Chapter 4. Abbreviations: ERSP = event-related spectral perturbations, PC(A) = principal component (analysis), FPR = false positive rate, FNR = false negative rate, AUC = area under the curve, ROC = receiver operator characteristic, Imms = Immunisation Dataset.

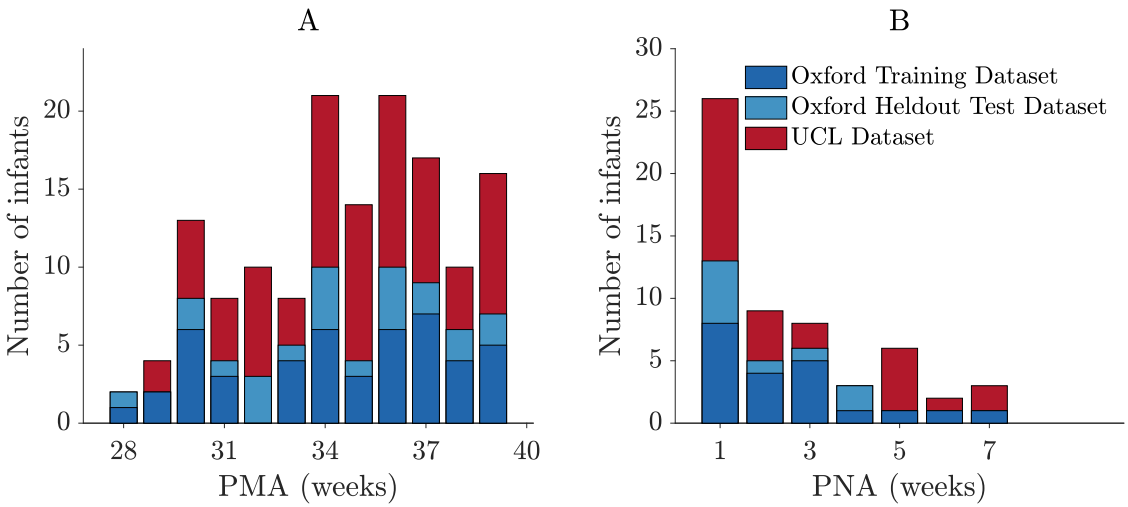


Figure 3.2: PMA and PNA distributions in the three main datasets. A) Distribution of PMA in the Oxford Training Dataset, Oxford Held-out Test Dataset and the UCL Dataset. The datasets contain data from infants between 28 and 40 weeks PMA, and the age distribution is similar across the three datasets. This figure is my own work and was published in an edited format in the Supplementary Materials of [321]. B) Distribution of PNA in the Oxford Training Dataset, Oxford Held-out Test Dataset and the UCL Dataset.

	Oxford Dataset		**UCL Dataset**
	Training	Heldout Test	
Infants	47	23	74
PMA (weeks)	35.3 (32.2–37.6)	34.9 (32.4–37.4)	35.6 (33.0–37.1)
GA (weeks)	34.3 (29.2–37.2)	34.1 (30.2–36.7)	34.3 (31.0–36.3)
PNA (weeks)	0.7 (0.3–2)	0.7 (0.2–1.5)	0.7 (0.6–1.6)
Females	21 (45)	10 (43)	38 (51)
Vaginal delivery	25 (53)	14 (61)	34 (46)
Prior blood tests (number)	9 (3–14.8)	6 (3–13.3)*	13 (7–20)

Table 3.1: Demographic variables for the participants in Oxford Training Dataset, the Oxford Held-out Test Dataset and the UCL Dataset. Values are either median (25th percentile, 75th percentile) or number (percentage). *Number of prior blood tests is missing for one infant.

The model was also tested in the Immunisation Dataset ([46], see Section 2.1.2 for details), which contains cerebral, autonomic and behavioural data from preterm-born infants receiving routine childhood vaccinations in the neonatal unit at OUH NHS Trust. Until October 2018, infants were given paracetamol immediately after the administration of meningitis B immunisations to prevent fever. After this time a local guideline change occurred and from then on paracetamol was given an hour before immunisations. As such, the immunisation dataset contains data from

a group of infants receiving paracetamol post-immunisation (n = 13, 5 females, median PMA = 37.3 weeks, median PNA = 9.1 weeks, "Control") and a group of infants receiving paracetamol pre-immunisation (n = 13, 5 females, median PMA = 37 weeks, median PNA = 9.0 weeks, "Paracetamol"). Infants received up to three immunisations as per the clinical protocol, and only the first immunisation with 1.6 seconds of artefact-free EEG recording was considered for each infant.

3.2.3 Features of interest

Before a classification model could be trained, features (predictors) needed to be extracted from data recordings. Literature was reviewed to identify modalities and corresponding features that were available in the Oxford Training Dataset that could possibly achieve discrimination between noxious and non-noxious stimuli in infants aged less than 40 weeks PMA. The following paragraph and Table 3.2 provide a brief overview of the included features and modalities. A more detailed summary of each modality can be found in Chapter 1. The behavioural domain included facial expressions and reflexive limb withdrawal. Facial expressions were included because they are widely utilised in infant pain assessment and are more responsive to noxious compared to non-noxious stimulation in term and late-preterm infants [123, 120]. Limb withdrawal responses were included because they encode stimulus intensity in term infants [132], and are evoked by noxious stimuli in infants of a wide age range [133, 53], although it is important to note that PMA is likely to influence their discriminative potential [53]. An increase in heart rate has been preferentially linked to noxious compared to non-noxious stimulation in both (near-)term infants [16, 320] and preterm infants [165, 141, 56, 164, 337]. Finally, neural responses were considered. The magnitude of the infant noxious-evoked ERP as measured by the template of noxious-evoked EEG activity [131] was included as this template has been validated in infants from 34 weeks PMA. In addition, PCA was used to extract new features from the noxious and non-noxious time-locked EEG responses (see Section 3.2.5.3). Features were also extracted from noxious-evoked spectral EEG changes, as a study in term infants showed that noxious stimulation

evoked spectral changes that were not present in response to a non-noxious control procedure [90]. PMA was included so that the model would be developmentally-sensitive. This list of features was designed to cover multiple response domains while using modalities with the strongest evidence from previous literature. However, it was restricted by data availability and is by no means an exhaustive list of all possible infant noxious-evoked responses.

Response domain	Response modality	Measurement technique	Extracted feature	Feature name
Neural	Cerebral	EEG	Magnitude of previously derived template of noxious-evoked EEG activity [131]	Template
			ERPs summarised as magnitude of first 3 PCs derived from noxious, non-noxious and combined noxious and non-noxious EEG data.	PC 1-3 noxious, PC 1-3 non-noxious, PC 1-3 all data
			ERSP: mean log power in 5 time-frequency windows.	Early delta, early alpha, late delta, late alpha, late beta
Autonomic	Cardiovascular	ECG	Heart rate increase from mean in 15 seconds pre-stimulus to 15 seconds post-stimulus (14 seconds in UCL dataset)	Heart rate
Behavioural	Limb withdrawal	EMG	Root mean square (RMS) of EMG activity recorded from the ipsilateral and contralateral biceps femoris in the first second post-stimulus.	Ipsilateral reflex, contralateral reflex
	Facial expressions	Video	Duration of brow bulge in seconds in the 30 seconds post-stimulus.	Brow bulge duration
Age	N/A	Best obstetric estimate	PMA (Gestational age at birth + postnatal age at study)	PMA

Table 3.2: Response modalities and features used in this study, organised by response domain. This table was published in an edited version in [321] under the terms of a CC-BY license.

3.2.4 Data acquisition

Recruitment and data acquisition for the Oxford Heel Lance and Immunisation datasets is described in detail in Chapter 2. In brief, EEG, ECG and EMG (Oxford Heel Lance Dataset only) data were recorded in infants during a noxious heel lance and a non-noxious control heel lance or during routine immunisations to

capture cerebral, cardiovascular and limb withdrawal responses. EEG data were recorded from Cz and 3 to 20 other electrodes, ECG data were recorded with the EEG using an electrode on the left clavicle referenced to the EEG reference electrode and EMG (Oxford Heel Lance Dataset only) was recorded from the biceps femoris from both legs. The control heel lance and heel lance were time-locked to the electrophysiological recordings using an automated user detection interface [356], while the immunisations were time-locked using a high-speed video camera. Facial expressions were recorded using a video camera from 15 seconds before the stimulus until 30 seconds afterwards.

The UCL Dataset was obtained from the original authors in UCL through the UK Data Service (see Section 2.1.3 and [169, 166]) and contains EEG and ECG recordings during a heel lance and control heel lance, as well as facial expression scores.

3.2.5 Feature extraction

Features were derived in the Oxford Training Dataset and then extracted from the Oxford Held-out Test Dataset, UCL dataset and Immunisation dataset in the same manner, unless indicated otherwise. Not all features were available for all infants because of artefacts or equipment failure. A summary of missing data is given in Table 3.3.

3.2.5.1 EEG preprocessing

The general steps in EEG preprocessing can be found in Chapter 2. EEG data at the Cz electrode were imported into MATLAB (Mathworks, version 2019a) using Brainstorm [300], then high-pass filtered at 1 Hz (lower pass-band edge) and low-pass filtered at 30 Hz (upper pass-band edge) with a notch filter at 50 Hz using a Hamming windowed sinc FIR filter implemented in EEGlab [68]. These filter cut-offs have been used in studies investigating infant heel lance responses in previous publications [167, 169, 328]. Epochs were extracted around the control heel lance, heel lance and immunisation events. Epochs were 1.55 seconds (1.6 seconds for the immunisation data) with 0.5 seconds pre-stimulus for PCA and

	Oxford Training Dataset	**Oxford Heldout Test Dataset**	**UCL Dataset**	**Immunisation Dataset**
Template and PCs	0/47 infants.	0/23 infants.	0/74 infants.	0/13 control group, 0/13 paracetamol group.
ERSP	3/47 heel lances (artefact).	1/23 heel lances (artefact).	N/A: only 2 seconds of post-stimulus EEG data available.	4/13 control group (artefact), 2/13 paracetamol group (artefact).
Heart rate	4/47 infants (artefact).	1/23 infants (ectopic beats).	19/74 control heel lances (4 artefact, 15 less than 14 seconds of ECG) and 9/74 heel lances (7 artefact, 1 ECG less than 14 seconds and 1 ECG not available).	0 /13 control group, 1/13 paracetamol group (artefact).
Ipsilateral reflex	2/47 infants (artefact).	0/23 infants	N/A: no EMG available.	N/A: no EMG recorded.
Contralateral reflex	3/47 infants (1 artefact, 2 contralateral EMG not recorded).	0/23 infants.	N/A: no EMG available.	N/A: no EMG recorded.
Brow bulge duration	3/47 infants (failed recording).	2/23 infants (failed recording).	4/74 heel lances and 4/74 control heel lances (no brow bulge score available, data from 5 infants).	0/16 control group, 2/16 paracetamol group (failed recording).

Table 3.3: Overview of missing data for each feature, split by dataset. Numbers refer to the number of infants excluded (not the number of observations), unless this is explicitly stated otherwise. This table was published in an edited version in [321] under the terms of a CC-BY license.

projection of the template and 6 seconds with 2 seconds pre-stimulus for the ERSP analysis. Epochs for the ERSP analysis were visually inspected and epochs with artefact were rejected (Table 3.3).

3.2.5.2 Template of noxious-evoked EEG activity

The pre-stimulus mean was subtracted from each epoch (baseline correction) and then the template of noxious-evoked EEG activity [131] was projected to each epoch as previously described in [131] (also see Section 3.2.5.3). To account for latency differences between infants, each trial was first Woody-filtered in the 400–700 ms post-stimulus with a maximum jitter of +- 50 ms. In brief, this involves finding the number of samples that the data should be shifted by to obtain the highest

cross-correlation between the template and the data, while limiting the maximum shift in positive and negative direction to 50 ms [353]. In the Immunisation Dataset the responses were time-locked to the EEG using a high-speed video camera, instead of the automated system that was used for the other datasets. To account for inaccuracies in time-locking, immunisation data were allowed to jitter by 100 ms instead of 50 ms. Fig. 3.5 shows example traces for two infants, before and after Woody-filtering and projection of the template of noxious-evoked EEG activity.

3.2.5.3 Derivation of age-dependent PCs

While the template of noxious-evoked activity is validated in term infants, it is unlikely to be present in preterm infants below 34 weeks [89]. Therefore I used a data-driven approach to find additional waveforms that explained across-trial variance in the shape of the ERPs at the Cz electrode by using PCA. PCA was carried out in the Oxford Training Dataset only, and the resulting PCs were projected to the Oxford Training Dataset, Oxford Held-out Test Dataset, UCL Dataset and Immunisation Dataset.

PCA decomposes a data matrix (a set of observations for a set of variables) into principal components (PC), which are uncorrelated linear combinations of the original variables [314]. In this analysis, I used PCA to decompose epoched EEG data, as previously described for infant noxious-evoked EEG data [290]. Trials (epochs) were considered as variables and timepoints as observations, and the resulting principal components are thus waveforms over time, which explain across-trial variability in ERP amplitude [285, 290]. Given a new epoch of EEG data, the presence of a given PC (waveform) in that epoch can be quantified by projecting the epoched data onto the PC to find the PC "coefficient" or "magnitude" (scaling factor). Vice versa, the PC magnitude is the scaling factor that the PC is multiplied with to best fit the data. In the rest of this thesis, this process will be described as "projecting the PC to the data", for consistency with previous literature [290, 131]. Positive PC magnitudes signal that the waveform is present in the data, with larger

values representing larger presence, whereas negative PC magnitudes signify that the inverse of the waveform is present in the data.

I aimed to identify waveforms that explained variance in the noxious data, the non-noxious data and waveforms that were common to both the noxious and the non-noxious data. Therefore, I conducted PCA in three subsets of trials within the Oxford Training Dataset: the noxious heel lance data, the non-noxious control heel lance data and the combined noxious heel lance and non-noxious control heel lance data. Each of the steps below was taken for each of the three subsets. As it was expected that ERP shape and latency would gradually change with age [89, 286], it was undesirable to limit the analysis to arbitrary age brackets and data from infants across the full age range of 28–40 weeks PMA were combined. To account for age-related differences in ERPs with potentially differing latencies, each trial was first aligned with an age-weighted average that was specific for that trial, allowing for a +- 50 ms jitter. This average was constructed by assigning weights between 0 and 1 to all the data, where data from infants who were close in age to the infant of interest were upweighted, and others were downweighted. The data of the infant itself, and of other infants of exactly the same age in days was assigned a weight of 1. For data from infants who were more than 28 days older or younger than the infant of interest, weights were set to 0. Data from infants whose PMA lay within +− 28 days of the PMA of the infant of interest were assigned a weight derived from a gaussian window (Fig. 3.3). The MATLAB function *gausswin* was used to derive weights w for each day of difference n. For a given infant, weights w for all trials within the window were computed for each day of difference n as follows:

$$w(n) = e^{-\frac{1}{2}\left(a\frac{n}{(L-1)/2}\right)^2}$$

a is the width factor which determines the width of the gaussian curve and L is the window length in days + 1. For this analysis $a = 2.5$, which is the default value in MATLAB's *gausswin*, $L = 2 * 28 + 1 = 57$ and $-28 \geq n \leq 28$.

The aligned trials were then concatenated and PCA was conducted across the first 1000 ms post-stimulus. For each set of trials, the first PCs that explained at least 75% of the cumulative variance were taken forward.

To obtain the magnitude of each PC for each infant, the identified PCs were projected to the EEG trials in all datasets. In the Oxford Training Dataset, the Oxford Heldout Test Dataset and the UCL Dataset, the data were Woody-filtered by +- 50 ms. In the Immunisation Dataset the responses were time-locked to the EEG using a video camera instead of an automated system. To account for inaccuracies in time-locking, data were Woody-filtered by 100 ms instead of 50 ms in this dataset. The PC magnitudes were brought forward as features in the model. Fig. 3.4 shows the nine PCs that were identified and carried forward to the model. Fig. 3.5 shows example traces for two infants with PC 2 noxious overlaid in blue.

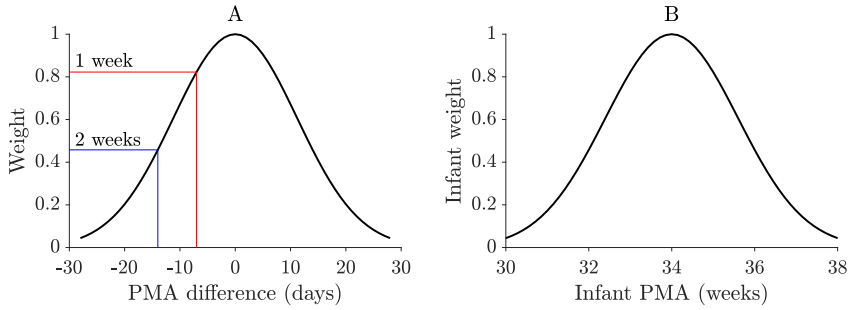


Figure 3.3: To construct age-weighted averages to align each trial to, a gaussian window was used to assign weights to the trial itself and the other trials in the dataset. A) The gaussian used to assign weights had a window size of 8 weeks with 28 days on either side of the PMA of the infant of interest. The weight corresponding to a 1-week difference from the centre PMA is indicated in red, the weight corresponding to a 2-week difference is indicated in blue. Trials with a PMA outside the window were assigned a weight of 0. B) Example: for an infant aged 34 weeks PMA, the age-weighted average includes data from infants aged 30+0 weeks PMA up to and including 38+0 weeks PMA.

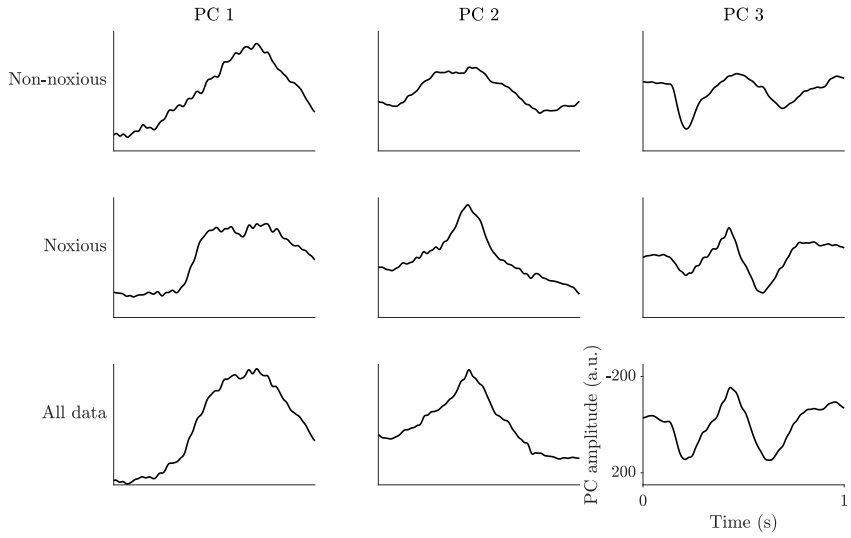


Figure 3.4: Principal component analysis was used to decompose the non-noxious, noxious and combined non-noxious and noxious EEG data into principal components (PC) in the Oxford Training Dataset. The upper row shows the three PCs derived from the non-noxious data, the middle row shows the three PCs derived from the noxious data and the lower row shows the three PCs derived from the combined non-noxious and noxious data. The nine PCs were projected to all EEG traces in all datasets. a.u. = arbitrary units.

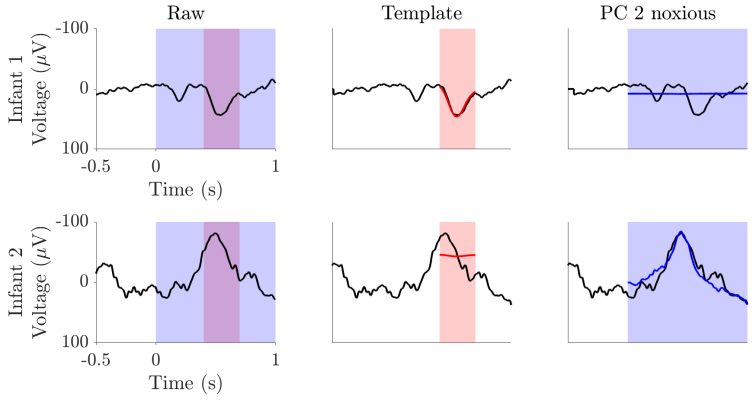


Figure 3.5: The template of noxious-evoked EEG activity and the nine PCs identified in the principal component analysis were projected to individual EEG traces to obtain the template or PC magnitude. This figure shows example EEG traces from two individual infants (infant 1 and infant 2, Raw). The traces were Woody-filtered and the template of noxious-evoked EEG activity (Template, in red) and PC 2 noxious (PC 2 noxious, in blue) were projected onto them and scaled. Shaded areas represent the time period in which the trace was aligned to the template (red) or PC 2 noxious (blue). For infant 1, the template of noxious-evoked activity has a high magnitude, and PC 2 noxious magnitude is low. For infant 2, this is reversed.

3.2.5.4 ERSP analysis

Data were decomposed into the event-related spectrum (ERS) consisting of $200 * 59$ time-frequency pixels using standard EEGlab functions [68, 118]. To obtain the baseline corrected ERSP data, the ERS value of each of the time-frequency pixels was divided by the mean value of the corresponding frequency in the 500 ms preceding the stimulus in each trial separately [118]. The single trials were then log-transformed to obtain the $ERSP_{log}$. The single trial $ERSP_{log}$ results were averaged within the noxious data across infants in the Oxford Training Dataset and time-frequency windows in which the average noxious-evoked $ERSP_{log}$ showed prominent derivations from 0 were visually identified (see Fig. 3.6 and Table 3.4). For each trial, the mean $ERSP_{log}$ in each of the time windows of interest was a feature in the model. Log transformation was applied prior to averaging as the log-transformed values would be features in the model.

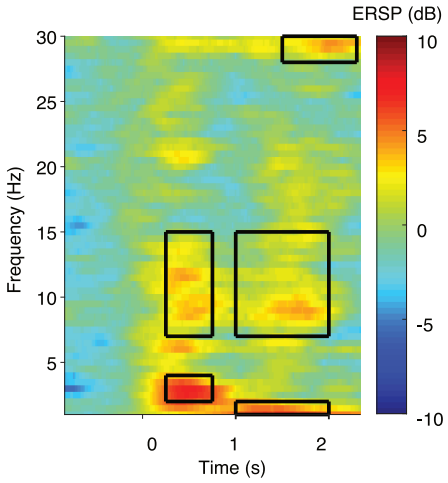


Figure 3.6: Average $ERSP_{log}$ plot at Cz in the noxious data in the Oxford Training Dataset. The black boxes indicate the windows of interest that were taken forward as features which were visually identified in this average plot — see Table 3.4. his figure is my own work and was reproduced from [321] with permission under the terms of a CC-BY license.

Name	Time window	Frequency
Early delta	250–750	2–4 Hz
Early alpha	250–750	7–15 Hz
Late delta	1000–2000	1–2 Hz
Late alpha	1000–2000	7–15 Hz
Late beta	1500–2300	28–30 Hz

Table 3.4: The five time-frequency windows of interest that were visually identified in the average $ERSP_{log}$ plot in the Oxford Training Dataset (black boxes in Fig. 3.6). This table was published in an edited version in [321] under the terms of a CC-BY license.

3.2.5.5 EMG analysis

EMG data were rectified and filtered between 10 and 500 Hz (using separate high-pass and low-pass filters, values denote lower and upper pass-band edge, respectively) with a notch filter at 50 Hz and harmonics at 100 and 150 Hz using a Hamming windowed sinc FIR filter implemented in EEGlab [68]. Epochs were created with 5 seconds before the stimulus and 15 seconds after the stimulus. Data were visually inspected and rejected if they contained artefact (see Table 3.3). The root mean square (RMS) of the signal was calculated in 250 ms bins. As inter-individual differences were present in the baseline EMG signal, the post-stimulus RMS data were divided by the mean RMS in the one second pre-stimulus. The mean baseline-corrected RMS in the first 1000 milliseconds post-stimulus for the ipsilateral and contralateral EMG were features in the model.

3.2.5.6 ECG analysis

As the ECG data were recorded with reference to the EEG reference electrode, data were filtered between 12 and 40 Hz (using separate high-pass and low-pass filters, values denote lower and upper pass-band edge, respectively) in EEGlab [68] and epoched with 30 seconds before and after the stimulus for the Oxford Dataset and 14 seconds before and after the stimulus for the UCL Dataset, as this length of data was available for most infants in the UCL Dataset. Next, RR peaks were automatically identified using a custom-written MATLAB script that employed the function *findpeaks.m*, which identifies local maxima. ECG and RR traces were visually examined, and where necessary, RR peaks that were mislabeled (e.g., the algorithm

sporadically missed low amplitude peaks or picked up artefactual peaks) were manually corrected. If the artefact was extensive, traces were rejected (Table 3.3). From the RR trace, heart rate was calculated at each second using the inverse of the RR rate in 3-second sliding windows. In [320] I showed that the change in heart rate from baseline to maximum in the 15 seconds pre- and post-stimulus had good discriminative accuracy in infants aged 34 weeks and over. Therefore, this metric was used as a feature in the model. In the UCL data, 14 seconds pre- and post-stimulus were used instead of 15, as this length of data was available for most infants.

3.2.5.7 Facial expression analysis

For the Oxford Dataset, the duration of brow bulge in seconds was assessed in video recordings of the control heel lance and the heel lance by trained investigators who were blinded to the stimulus. In the UCL Dataset, facial expression durations were provided with the dataset. Only one facial expression was included in the model, because the three facial expressions in the PIPP-R score are highly correlated [320]. Brow bulge was chosen because this involves the upper half of the face, making it observable even in infants who are on mechanical ventilation.

3.2.6 Classification model

3.2.6.1 Theory

The goal of this study was to investigate whether it is possible to distinguish infant responses to a noxious stimulus from infant responses to a non-noxious stimulus using a set of features derived from cerebral, autonomic, limb withdrawal and facial expression modalities. I implemented a bagged decision trees classification model [33, 32] which classified observations (infant responses) into one of two classes, noxious (heel lance) and non-noxious (control heel lance), based on 20 features (see Table 3.2). A bagged decision tree ensemble was chosen because this type of model generates robust predictions and can handle missing data [105]. A bagged decision trees model consists of a set of simple decision trees which are each trained on a bootstrap sample of the training data [32]. For each tree, the bootstrap

sample is referred to as the "in-bag" sample for that tree, whereas the rest of the data is referred to as the "out-of-bag" data. Growing a classification tree starts with a single node which contains the training observations of both classes. The algorithm selects the feature and threshold value for that feature that obtains the best split between the two classes. Observations are then assigned to one of the two resulting leaves and the process is repeated until a pre-determined leaf size is reached or further splitting does not improve the split between the two classes [33, 32]. Any new observation travels through the tree until it reaches its final leaf. The observation is then assigned a score for the noxious class, which is simply the fraction of observations in that leaf that belongs to the noxious class. For instance, if a new observation arrives in a leaf where 75% of the training observations belongs to the class "noxious", the score for "noxious" is 0.75. Inversely, the score for "non-noxious" is 0.25. Generally a threshold of 50% is used to assign the class label. Thus, in this example, as "score noxious" is higher than 50%, the predicted label is "noxious". Single decision trees tend to have low classification accuracy. In a bagged decision tree approach, hundreds of decision trees are grown on slightly different subsets of the data, obtained by bootstrapping, and the final prediction is made by aggregating the predictions of all the trees in the ensemble, which reduces overfitting and improves accuracy [33, 32].

3.2.6.2 Implementation and evaluation

A bagged decision trees classifier was implemented using the function *TreeBagger.m* in MATLAB (Mathworks, version 2019a). The interaction-curvature test was set as a split criterion, and the minimum leaf size was set to 3. The interaction-curvature test allows for interactions between features and unbiased importance estimates. The model was trained using leave-one-subject-out cross-validation where in each fold both observations from the same infant (i.e., both the heel lance and the control heel lance trial) were left out. This was done to minimise leakage from the training data to the testing data. Cross-validation accuracy was then calculated across folds to obtain an estimate of classification performance.

To investigate the generalisability of the classifier, the model was retrained on all the data in the Oxford Training Dataset and then tested on two independent data sets: the Oxford Held-out Test Set and the UCL Dataset.

Classification accuracy was calculated for all infants in each dataset, in the preterm infants only, and in four non-overlapping 3-week age bins: 28–30+6 weeks PMA, 31–33+6 weeks PMA, 34–36+6 weeks PMA and 37–39+6 weeks PMA. For simplicity, these are referred to as 28–31 weeks, 31–34 weeks, 34–37 weeks and 37–40 weeks PMA. Wald binomial proportion 95% confidence intervals were computed for each accuracy estimate. As this was a two-class problem with the same number of observations for each class, the expected null accuracy was 50%.

False positive rates (FPR), false negative rates (FNR) and Receiver Operator Characteristic (ROC) curves and the Area Under the Curve (AUC) were calculated as additional metrics to evaluate model performance. These performance metrics provide information on the optimal threshold to separate the two classes and on the balance between false positive and false negative predictions. ROC curves, optimal operating points and AUCs were generated with the MATLAB function *perfcurve*. The optimal operating point of the ROC curve was found by moving a line with a slope of 1 from the upper right corner down and to the right until it intersects with the ROC curve. This point does not necessarily represent the clinically optimal point, however it was used as a reference to investigate whether there were any large variations between the different datasets or age groups.

3.2.6.3 Feature importance

Feature importance was assessed in both the Oxford Training Dataset and the Oxford Held-out Test Dataset using permutation testing. In the Training Dataset, the feature importance was calculated with a built-in MATLAB function during model training. In brief, for each tree and feature separately, the out-of-bag values for that tree and feature were permuted and the decrease in accuracy was calculated. For each feature, the final feature importance was obtained by normalising the accuracy loss for the full model by averaging across trees and dividing by the

standard deviation. As there was leakage between the in-bag and out-of-bag folds (the two observations from the same infant were treated independently), feature importance was also assessed in the completely separate Oxford Heldout Test Dataset. Observations in this dataset were permuted 50 times for each tree and each feature and accuracy loss was calculated and standardised as above to gain feature importance values for the final model. This was analogous to the out-of-bag feature importance analysis conducted using the built-in MATLAB function, so that the feature importance estimates in the out-of-bag samples in the Training Dataset and the Held-out Test Dataset could be compared. As a comparison, feature importance was also calculated for the full ensemble instead of for each tree separately. Each feature was permuted 50 times and for each permutation, the full model was used to generate predicted values, and permuted accuracy was calculated. Feature importance was defined as the mean accuracy loss compared to the model true accuracy across the set of permutations. The absolute values of the feature importance values from this analysis cannot be compared to the out-of-bag feature importance estimates. Feature importance was not assessed in the UCL dataset as many of the features were missing.

3.2.7 Examples of classifier applications

A potential application of the classification model is to use the classification score as an outcome measurement in clinical trials of analgesics. Therefore, I explored whether the classifier could discriminate between noxious and non-noxious events within-infant to allow paired analyses, and I explored whether model scores were modulated by the analgesic paracetamol.

3.2.7.1 Exploring within-infant discrimination: paired comparisons

For the main analyses, all observations were considered independently, i.e., the classifier was not forced to assign a noxious label to one observation and a non-noxious label to the second observation from the same infant, and accuracy, FPR and FNR were calculated across all observations. This was done purposefully,

because the goal was to derive a classifier that predicts exposure to a noxious stimulus based on a single observation in a new infant, without the need for paired observations. However, the classifier could possibly be used in a clinical trial of analgesics with a cross-over design where infants are studied under two experimental conditions. In such a situation, the classifier score should be sensitive to differences in noxious-evoked responses within infant. Therefore, the level of within-infant discrimination was also investigated by comparing the classification scores (which can loosely be interpreted as the probability of that observation belonging to the noxious class) for the noxious and the non-noxious stimulus within infant.

The forced choice classification accuracy was calculated as the proportion of infants who had a higher classification score for the noxious stimulus than for non-noxious stimulus. This was done in each dataset separately for all infants and for the preterm infants, and in the four 3-week age groups in the Oxford Training Dataset and the UCL Dataset. 95% binomial Wilson confidence intervals were calculated around each value. The Wilson method was used instead of the Wald method because of the low number of infants in some age groups, which renders the Wald method overly lenient. For simplicity, the Wilson method was applied to all forced choice accuracies, also to those with a larger number of infants. The forced choice accuracy is not an estimate of model performance in the same way as the accuracy across all observations, rather it is an exploration of whether the classifier may be used in a clinical trial with paired observations in the future.

Scores for the noxious and the non-noxious stimulus were also directly compared in the four age groups to further evaluate within-infant discriminability. If responses to noxious and non-noxious events would not be discriminable within infant (H0), the difference in score for the noxious and the non-noxious stimulus within infant is expected to be 0. This null hypothesis was tested using a two-tailed one-sample t-test in each of the four age groups described above, with p-values derived non-parametrically in FSL PALM using 10000 (or exhaustive, whichever was lower) permutations [350].

3.2.7.2 Testing the classification model in a pharmacological study

The Immunisation Dataset was used to investigate whether the classifier would generalise to a different noxious procedure, and whether paracetamol would modulate the prediction scores. Predictions were generated for all infants in the immunisation dataset. As infants had recordings for a variable number of immunisations (1-3) in the same session, for each infant only the first successfully recorded immunisation (defined as at least 1.60 seconds of time-locked artefact-free EEG available around the immunisation event) was considered. First, the classification accuracy of the model was assessed in the group of infants who received paracetamol after the immunisation ("Control") to investigate the generalisability of the classifier to this different stimulus and population of ex-premature infants. The 95% Wald binomial confidence interval was calculated around the accuracy. Next, classifier scores were compared between the control group and the paracetamol group to assess whether paracetamol affected the predictions. The accuracy of the classifier and the 95% Wald binomial confidence interval was calculated for the "Paracetamol" group and groups were directly compared using a two-sample (unpaired) t-test with p-values derived non-parametrically using 10000 permutations implemented in FSL PALM [350].

3.3 Results

3.3.1 Noxious stimulation evokes changes in cerebral, autonomic and behavioural modalities in infants 28–40 weeks

First, I confirmed that the heel lance evoked changes in cerebral activity, heart rate, limb withdrawal activity and facial expressions which could be distinguished from responses to the control heel lance at the group level in the Oxford Training Dataset (see Fig. 3.7). As described in previous work using (parts of) the Oxford Dataset, the heel lance evoked two negative-positive (N-P) complexes with latencies roughly between 100–300 ms and 400–700 ms, whereas the control heel lance only evoked the first N-P complex. In the frequency spectrum, an increase in power in the lower frequencies (2-15 Hz) was present in the first 1000 ms post-stimulus, and

an increase in power in higher frequencies (28–30 Hz) was apparent after 1500 ms. Less pronounced increases in EEG power were visible in response to the control heel lance. Reflex withdrawal of both limbs occurred in response to the heel lance, but not to the control heel lance. Similarly, there was a heart rate increase to the heel lance, but not to the control, and facial expressions occurred more often to the heel lance than to the control. These findings confirmed that it was appropriate to include these modalities in the classification model.

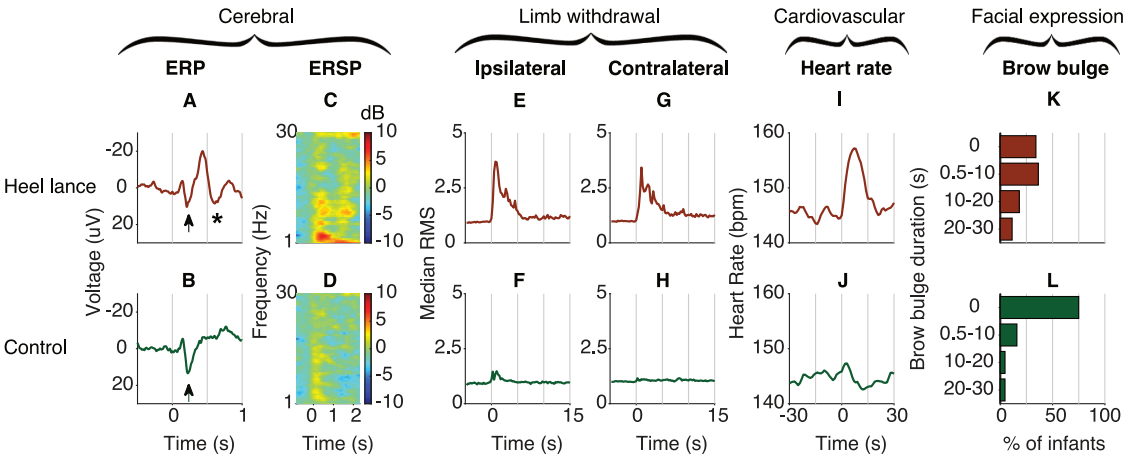


Figure 3.7: Average responses to the control heel lance and heel lance in the Oxford Training Dataset. The heel lance (noxious) and control heel lance (non-noxious) evoked changes in brain activity (at the Cz electrode), heart rate, limb withdrawal activity and facial expressions. Responses are averaged across the age range of 28–40 weeks. A-B) Both stimuli evoked an early N-P complex (arrow), but only the heel lance evoked a later N-P complex (asterisk). C-D) Event-related spectral perturbations were evoked by both stimuli, but the changes were visually larger for the heel lance stimulus. E-H) Both ipsilateral and contralateral withdrawal responses occurred after the heel lance, but less clearly after the control heel lance. I-J) Heart rate increased after the heel lance, but not after the control heel lance. K-L) A larger proportion of infants showed a brow bulge response to the heel lance compared to the control heel lance. This figure is my own work and was reproduced from [321] with permission under the terms of a CC-BY license.

3.3.2 Development of a classification model

A bagged decision trees classification model was trained using features from the cerebral, autonomic and behavioural modalities above and PMA as inputs. The classifier had an accuracy of 0.79 in the Oxford Training Dataset (95% CI 0.70–0.87, estimated by leave-one-subject-out cross-validation) with a FPR of 0.28, FNR of

0.15 and AUC of 0.84 (95% CI: Fig. 3.8,A, B). This combination of metrics indicated that the classifier was sensitive to picking out the heel lances, and had a higher tendency to misclassify control heel lances as heel lances than vice versa.

3.3.3 Generalisation of classification results

The classifier was tested in two independent datasets: the Oxford Heldout Test Dataset (to assess generalisability within the same research site) and the UCL Dataset (to assess generalisability across different research sites). The accuracy and ROC were similar to the Oxford Training Dataset in both datasets: the classifier had an accuracy of 0.76 (95% CI: 0.64–0.88) and AUC of 0.85 in the Oxford Heldout Test Dataset and an accuracy of 0.84 (95% CI: 0.78–0.90) and an AUC of 0.92 in the UCL Dataset (Fig. 3.8A, B). The 95% confidence for the classification accuracy did not include 0.5 in any of the three datasets, indicating that the classifier performed better than chance. In the UCL Dataset, the FPR was 0.11 and the FNR was 0.22, indicating a higher tendency to misclassify noxious responses as non-noxious than vice versa. In the Oxford Heldout Testset, the FPR was 0.22 and FNR was 0.26, indicating a relatively even balance between the risk of false positives and false negatives. The optimal operating point on the ROC curve in the UCL data, i.e. the point that provides the best trade-off between FPR and FNR in this dataset, was close to the operating point of the true model. This indicates that the threshold of 0.5 that was set in the Oxford Training Dataset was generalisable to this new dataset (Fig. 3.8B) .

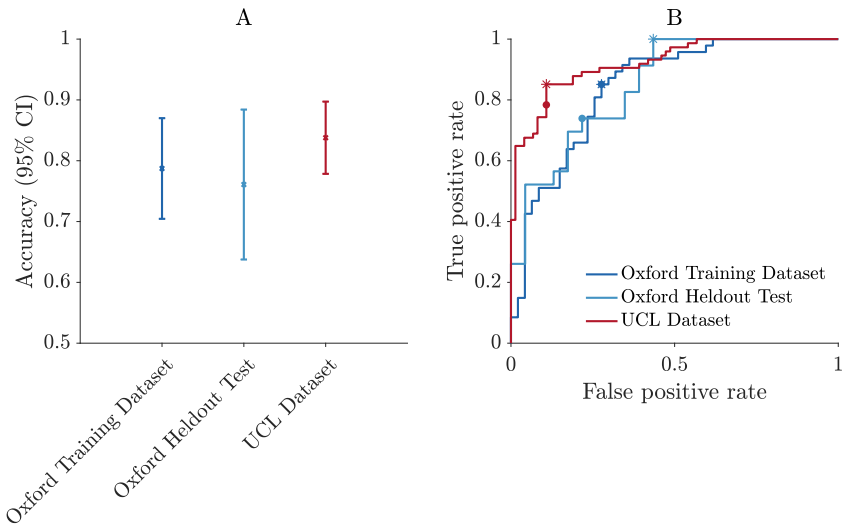


Figure 3.8: Performance of the classifier in the Oxford Training Dataset and the two independent datasets. A) Accuracy in classifying noxious from non-noxious responses in the three datasets. Error bars represent 95% binomial confidence intervals. In the Oxford Training Dataset, accuracy was calculated by using leave-one-out cross-validation. B) Receiver operator characteristic (ROC) curves showing the balance between true positives and false positives that can be achieved by varying the classification threshold. Solid dots indicate the operating point of the current model, whereas asterisks indicate the optimal ROC operating point.

3.3.4 Influence of PMA on classification results

) The influence of PMA on the classification results was investigated in subgroups of infants. First, performance was assessed across all preterm infants (PMA < 37 weeks) for the three datasets. The results in this group were similar to the overall classification results (Fig. 3.9A, B), with accuracy ranging from 0.74 in the Oxford Heldout Set to 0.83 in the UCL Dataset. As for the full dataset, the confidence intervals did not include 0.5. For each dataset, the ROC curves had similar shapes to those for the full dataset, indicating a similar trade-off between sensitivity and specificity. The model FPR and FNR were close to the optimal operating point for the Oxford Training Dataset and the UCL Dataset, which indicated that the threshold for classifying observations as noxious was appropriate.

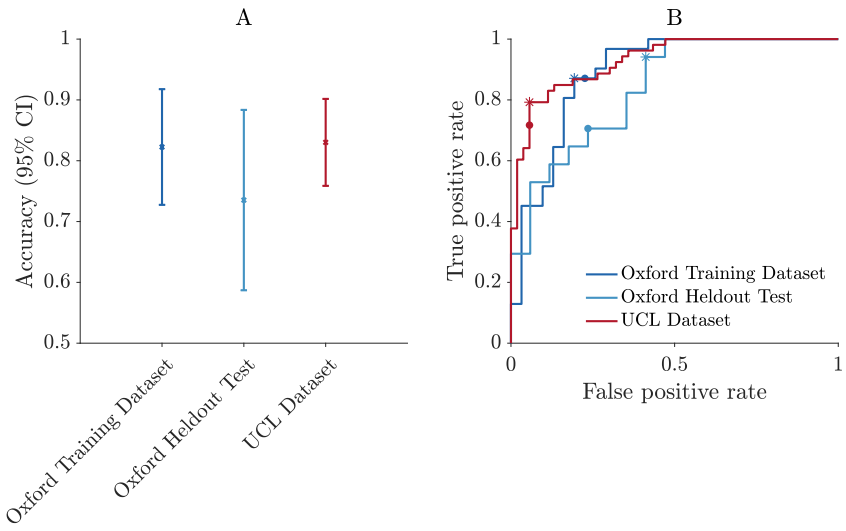


Figure 3.9: Performance of the classifier in the Oxford Training Dataset and two independent datasets, in the premature (postmenstrual age < 37 weeks) infants only. A) Accuracy in classifying noxious from non-noxious responses in premature infants in the three different datasets. Error bars represent 95% binomial confidence intervals. In the Oxford Training Dataset, accuracy was calculated by using leave-one-out cross-validation. B) Receiver operator characteristic (ROC) curves showing the balance between true positives and false positives that can be achieved by varying the classification threshold. Solid dots indicate the operating point of the current model, whereas asterisks indicate the optimal ROC operating point.

To further explore the effects of age on classification results, the Oxford Training Dataset and the UCL Dataset were split into non-overlapping 3-week age bins to assess whether accuracy changed across groups. The Oxford Heldout Dataset was not split any further, as the resulting subgroups would be too small (n = 3 in the youngest age group). For both datasets, classifier accuracy was balanced across the four age groups (Fig. 3.10). In the Oxford Training Dataset, accuracy was highest in the 31–34 week PMA group (accuracy = 0.86) and lowest in the highest age category (accuracy = 0.72), whereas in the UCL Dataset, the classifier had the highest performance in the two oldest groups, and the lowest in the 31–34 week PMA group (accuracy = 0.71), although confidence intervals of the different age groups overlapped. The effect of the small sample sizes especially in the youngest infants is evidenced by the large confidence intervals in these categories (Fig. 3.10). Notably, the accuracy for the 28–31 week old infants was 0.78 (95% CI 0.59–0.97) in the Oxford Training Dataset and 0.79 (0.57–1.00) in the UCL Dataset, showing

that even in infants of this age classification was possible above chance level. The optimal thresholds were variable for the different age groups. For the 37–40 week age group in the UCL Dataset, the optimal operating point on the ROC curve was relatively far removed from the position of the actual model on the ROC curve (Fig. 3.10). This suggested that for this age group and this dataset specifically, a different threshold could be of benefit.

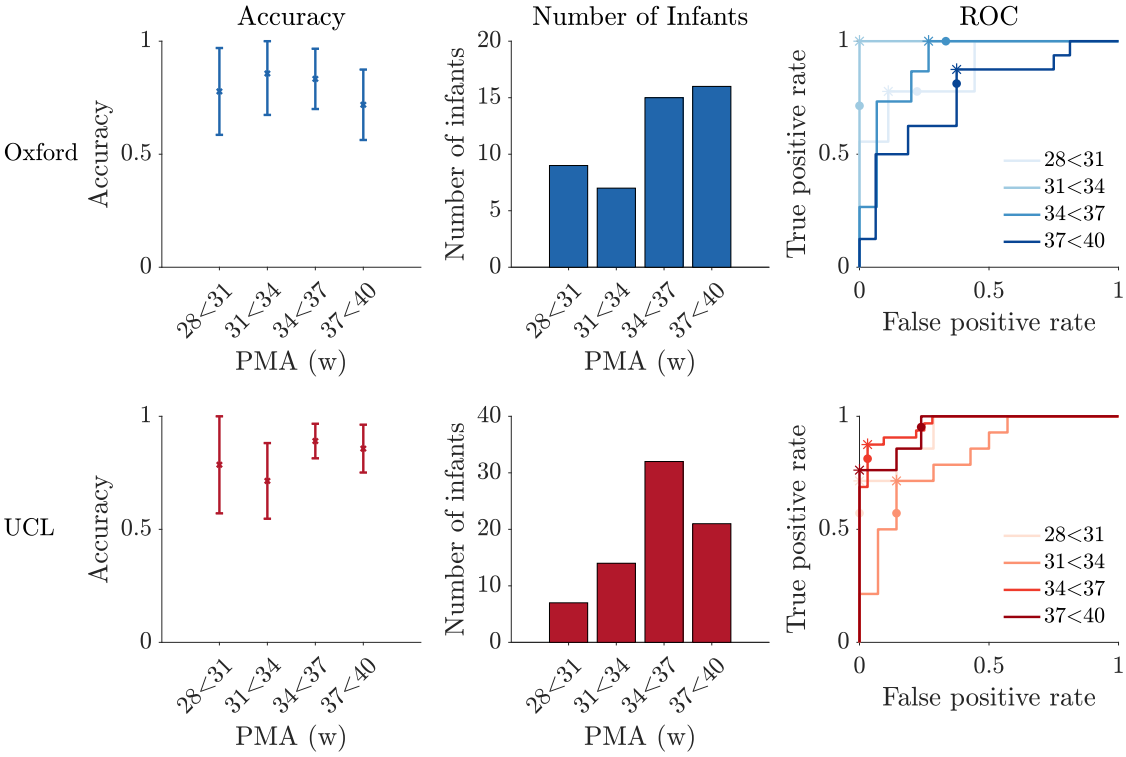


Figure 3.10: Classifier performance by postmenstrual age, displayed for four non-overlapping age categories: 28–31 weeks PMA, 31–34 weeks PMA, 34–37 weeks PMA and 37–40 weeks PMA. Upper row shows the classification accuracy with 95% binomial confidence intervals, number of infants in each age bin and receiver operator curves (ROC) in 3-week age groups for the Oxford Training Dataset. Lower row shows the same plots for the UCL Dataset. For the ROC curves, solid dots indicate the operating point of the present model, whereas asterisks indicate the optimal ROC operating point.

3.3.5 Feature importance

The contribution of individual features to model performance was assessed using permutation-based methods. The importance was assessed in the Oxford Training Dataset using the out-of-bag observations, and the Oxford Held-out Test set. For the Oxford Heldout Test Dataset, feature importance was assessed for the

individual trees in the model, and across the entire ensemble. Across the first two analyses, the four most important features were heart rate, late delta, ipsilateral limb withdrawal and the template of noxious-evoked EEG activity, which span autonomic, cerebral and behavioural modalities (Fig. 3.11A, B). For the full ensemble, the most important features were heart rate, followed by a range of cerebral features including, PC 2 noxious and PC 2 all data (Fig. 3.11C). Yet, in the Oxford Held-out Test Dataset (Fig. 3.11B, C), most features and all modalities contributed to classification accuracy. The difference between the feature importance calculated for individual trees and the full ensemble can potentially be explained because not all features are used in each individual tree. If features are correlated, removing these from the full ensemble might not yield a large accuracy loss, while removing these from the individual trees might do so.

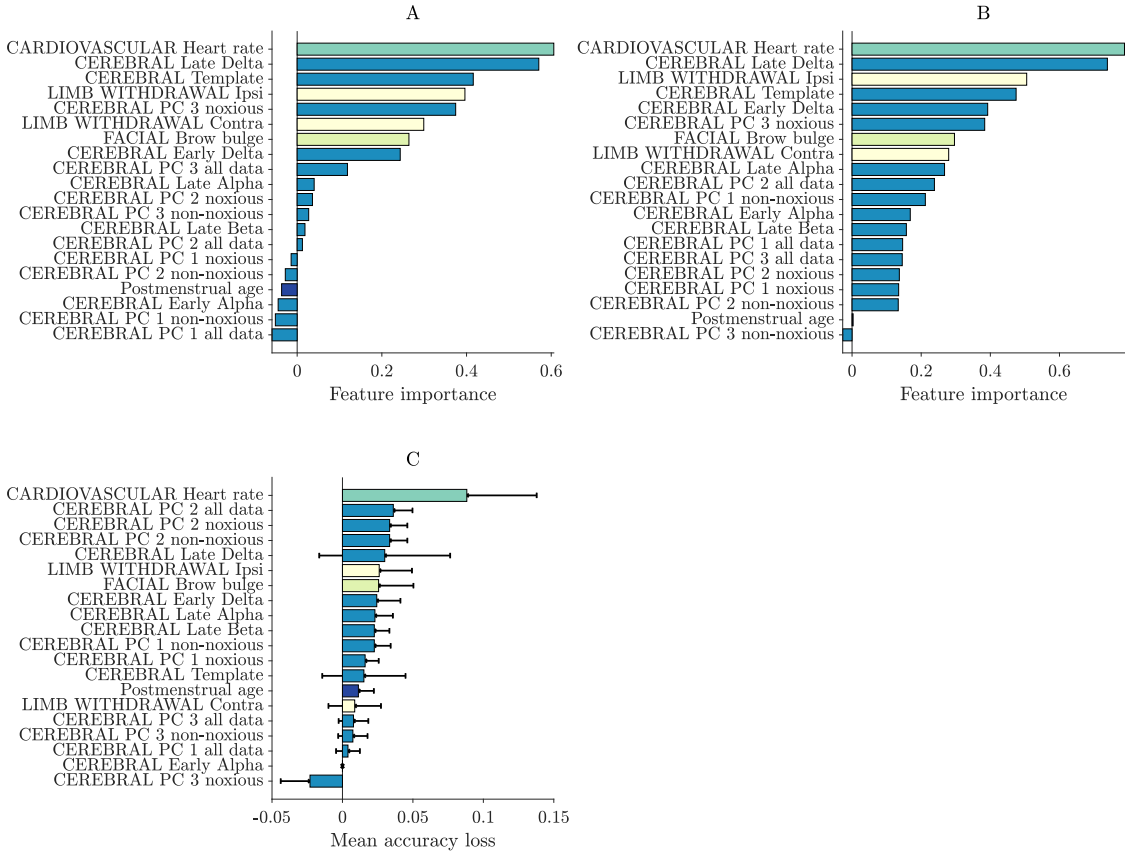


Figure 3.11: Feature importance estimates. Bars with the same colour belong to the same modality (cerebral, limb withdrawal, cardiovascular, facial expression). A) Feature importance in the Oxford Training Dataset, estimated on the out-of-bag samples, calculated individually for each tree and then standardised across all trees in the ensemble. B) Feature importance in the Oxford Heldout Dataset, calculated individually for each tree and then standardised across all trees in the ensemble. C) Feature importance in the Oxford Heldout Dataset, calculated across the full ensemble. Each bar represents the mean accuracy loss when that feature was permuted 50 times and error bars indicate the standard deviation of the accuracy loss across the 50 permutations.

3.3.6 Examples of classifier applications

Ultimately, a classifier such as the one presented here, could become a multimodal outcome measure in clinical trials of analgesics. This section provides preliminary examples of classifier applications in future clinical trials.

3.3.6.1 Exploring within-infant discrimination: paired comparisons

In the previous analyses, classification and discrimination was assessed across all observations independently, i.e. the classifier was not forced to assign a noxious

label to one observation and a non-noxious label to the second observation from the same infant. This was done purposefully, because the goal was to derive a classifier that predicts exposure to a noxious stimulus based on a single observation in a new infant. However, in some situations it is necessary to have an outcome measure which is sensitive to within-infant variation. An example is a cross-over trial, in which infants' noxious-evoked responses are assessed under two different conditions (e.g. a new analgesic versus standard care). Therefore the classifier's ability to detect within-infant differences was explored. This was done by calculating the forced choice accuracy of the classifier (the proportion of infants in whom scores were higher for the noxious stimulus compared to the non-noxious stimulus) and by evaluating within-infant differences in classifier score between the noxious and the non-noxious stimulus. It is important to note that neither of these analyses are a reflection of classifier performance, as they are only useful in specific situations in which paired observations are available. Rather, these results should be interpreted as an exploration of how the classifier scores could be used in a cross-over or longitudinal study of an analgesic.

For the three datasets overall, the forced choice accuracy was 0.85 (95% CI: 0.72–0.93) in the Oxford Training Dataset, 0.87 (95% CI: 0.68–0.95) in the Oxford Heldout Test Dataset and 0.91 (95% CI: 0.82–0.95) in the UCL Dataset. For the preterm infants, the forced choice accuracy was 0.87 (95% CI: 0.71–0.95), 0.82 (95% CI: 0.59–0.94) and 0.91 (95% CI: 0.80–0.96), respectively.

The within-infant difference in score between the noxious and the non-noxious stimulus was also investigated in smaller age groups. The forced choice accuracy values for each age group in the Oxford Training Dataset and the UCL Dataset are depicted in Table 3.5 and Table 3.6 and varied between 0.78 and 1. The confidence intervals around the forced choice accuracy did not include 50%, except in the 28–31 week group in the Oxford Training Dataset (forced choice accuracy = 0.78, 95% CI lower bound = 0.45). Fig. 3.12 shows the distribution of the score differences separately in the four age categories in the Oxford Training Dataset and the UCL Dataset. As reflected in the forced choice accuracy, the majority of infants had

higher scores for the heel lance compared to the heel lance (lower right quadrant in each subplot). In addition, the mean difference in score between the noxious and the non-noxious stimulus was significantly different from 0 in all age categories (Table 3.5, Table 3.6), indicating that within-infant discriminability was present on average in each age group. A minority of infants had lower scores for the heel lance than for the control heel lance (lower left quadrant in each subplot), indicating that in some individual infants, discrimination was not possible based on classifier scores.

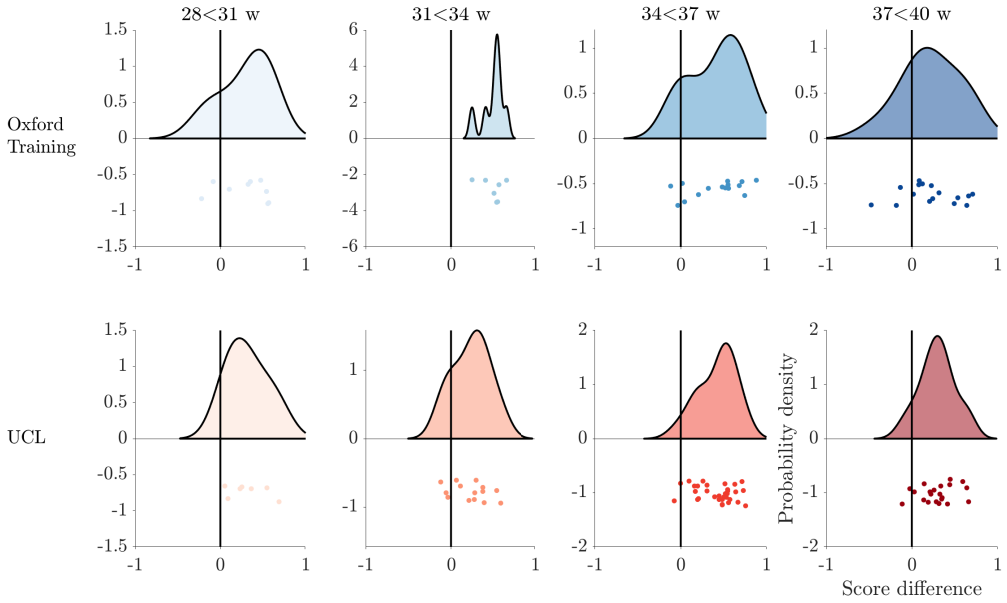


Figure 3.12: Probability distribution for the difference in classification score between the noxious and non-noxious stimulus, split by age group and dataset. Each dot represents one infant. Black lines indicate a difference of 0 between the noxious and non-noxious stimulus. Infants to the right of the black line have larger scores for the noxious procedure compared to the non-noxious procedure, indicating that responses were discriminable within-infant. Infants to the left of the black line have higher scores for the non-noxious procedure, indicating responses were not discriminable for these particular infants.

PMA group	Forced choice accuracy (95% CI)	Mean difference (std)	p
28–31	0.78 (0.45–0.94)	0.29 (0.29)	0.023
31–34	1.00 (0.65–1.00)	0.49 (0.13)	0.016
34–37	0.87 (0.62–0.96)	0.41 (0.32)	0.010
37–40	0.81 (0.57–0.93)	0.22 (0.33)	0.019
All			

Table 3.5: Forced choice accuracy with 95% binomial confidence intervals, mean difference in score for the heel lance and the control heel lance and p-value for the paired t-test comparing the scores in the Oxford Training Dataset. Values are displayed for four non-overlapping age categories: 28–31 weeks PMA, 31–34 weeks PMA, 34–37 weeks PMA and 37–40 weeks PMA.

PMA group	Forced choice accuracy (95% CI)	Mean difference (std)	p
28–31	1.00 (0.65–1.00)	0.32 (0.23)	0.016
31–34	0.79 (0.52–0.92)	0.23 (0.22)	0.0020
34–37	0.94 (0.80–0.98)	0.42 (0.22)	0.0001
37–40	0.90 (0.71–0.97)	0.30 (0.20)	0.0001

Table 3.6: Forced choice accuracy with 95% binomial confidence intervals, mean difference in score for the heel lance and the control heel lance and p-value for the paired t-test comparing the scores in the UCL Dataset. Values are displayed for four non-overlapping age categories: 28–31 weeks PMA, 31–34 weeks PMA, 34–37 weeks PMA and 37–40 weeks PMA.

3.3.6.2 Application of the classifier to a pharmacological study

The previous analyses were focused at discriminating two extreme conditions: a heel lance and a control heel lance. In clinical trials of analgesics, the change in noxious-evoked responses brought about by a an analgesic may be smaller. The sensitivity of the classifier scores to analgesic modulation was therefore investigated in pre-existing data from the Oxford Immunisation Dataset. This dataset contains recordings for 26 preterm-born infants (aged between 32 and 40 weeks PMA at study) receiving routine immunisations. Paracetamol is routinely given to infants at the time of their meningitis B immunisations for its anti-pyretic effects. Due to a change in clinical guidelines on the timing of paracetamol administration around vaccinations, 13 infants received paracetamol immediately after the immunisations ("Control") and 13 received paracetamol an hour before ("Paracetamol").

First, predictions were generated for the immunisation-evoked responses of the infants in the control group, to assess whether the classifier was able to detect

immunisation-evoked responses. The model correctly classified 11/13 (true positive rate = 0.85, 95% CI: 0.65 to 1.04) of the immunisation responses as "noxious", confirming the generalisability of the classifier to this different procedure. In the paracetamol group, the model only classified 7/13 of the immunisation observations as "noxious", which was not significantly different from chance level (true positive rate = 0.54, 95% CI: 0.27 to 0.81). The scores in the paracetamol group were significantly lower than in the control group (mean control group = 0.69, standard deviation control group = 0.22, mean paracetamol group = 0.54, standard deviation paracetamol group = 0.16, two-sample two-tailed t-test, t = 2.09, p = 0.044, Fig. 3.13), suggesting that a classifier like the one used here may be implemented to assess analgesic efficacy in future studies.

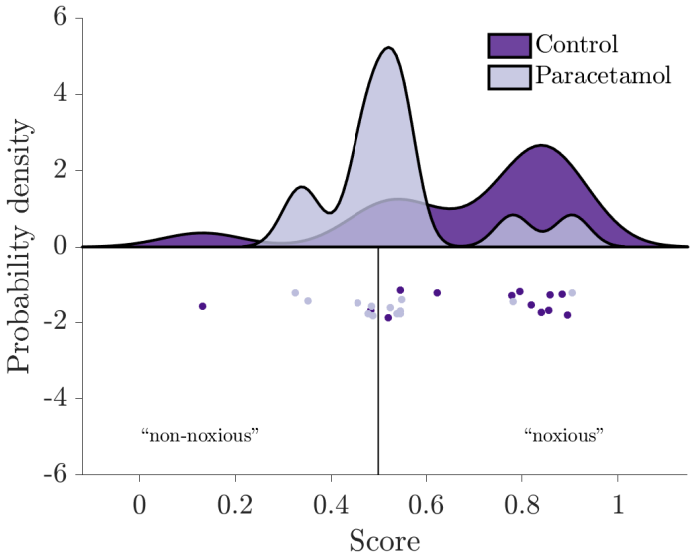


Figure 3.13: Scores and probability distribution for the control group (who received paracetamol after the immunisation) and the paracetamol group (who received paracetamol before the immunisation). Each dot represents one infant. The model uses a threshold of 0.5 (vertical black line) to assign the labels "non-noxious" or "noxious", therefore each infant to the right of the black line was labelled as "noxious". More infants in the paracetamol group were misclassified as "non-noxious" and classification scores in the paracetamol group were lower than scores in the non-paracetamol group.

3.4 Discussion

Pain assessment in preterm and early-term infants is challenging, as responsiveness to noxious stimuli changes with age. The goal of this study was to investigate whether discriminable multimodal activity patterns to noxious and non-noxious stimuli could be found in infants aged 28–40 weeks PMA, and to explore the feasibility of using such a multimodal approach in future clinical trials of analgesics.

3.4.1 Even very preterm infants display discriminable responses to noxious and non-noxious stimuli

A classification model was trained on data from infants aged 28–40 weeks PMA who underwent a noxious heel lance and a non-noxious control procedure. Overall, the model could distinguish noxious from non-noxious responses with an accuracy of 79% in the training data, estimated by leave-one-out cross-validation. Importantly, classification results generalised to a held-out set from the same site, a dataset collected at a different site by a different group of researchers and to a different noxious procedure (immunisations). The classification accuracy in the externally collected UCL dataset was 84%, which was higher than the accuracy estimates in the two local datasets. This is highly encouraging for the generalisability of these results to other centres. It also demonstrates the importance of sample size when estimating model performance, as a few outliers can affect accuracy estimates when the sample size is small.

Classification accuracy was investigated across the full age range, as well as in smaller subgroups. The classification accuracy was similar in the full age range compared to preterm infants only, and the classifier performed above chance in all age groups. Visual assessments of the ROC curves in the different age groups suggested that there would be little to gain in classification performance from varying the classifier threshold for different age groups, and thus that the model is suitable for use across the ages studied here. In the very youngest infants (28–31 weeks PMA), a classification accuracy of 78–79% was achieved, and on average, model prediction scores were higher for the noxious stimulus compared

to the non-noxious stimulus within infant. It is important to realise that the fact that the *classifier* discriminates noxious from non-noxious events, and the fact that we observe discriminable responses, does not imply that infants consciously discriminate between these two stimuli in the way that we as adults do [96]. It does suggest that at some level in the nervous system, these stimuli are differentially processed. Whether this is the result of a difference in stimulus intensity, or a difference in stimulus modality, remains yet to be discovered. Regardless of how noxious stimuli are perceived by infants, the notion that they are processed at different levels in the nervous system highlights the need for appropriate pain management strategies in infants of all ages.

The small number of infants in the youngest age category (28–31 weeks PMA, n = 9 in the Oxford Training Dataset and n = 7 in the UCL Dataset) was a limitation to this study and resulted in large confidence intervals around the accuracy estimates. Obtaining large sample sizes in studies of very preterm infants is challenging, as infants may be clinically unstable and require care which is incompatible with research studies. In addition, parents' openness to considering participation in research might be lower when coping with an unexpected preterm birth. The benefits of sharing data between researchers, of which the UCL Dataset [166] is a prime example, are increasingly acknowledged [145]. At least one other research group has published a shareable dataset of physiological and behavioural recordings around clinical procedures in preterm infants [271] and more multimodal neonatal studies are underway [266]. Further collaborative efforts are critical in obtaining larger sample sizes for more precise effect size estimates in the youngest infants.

3.4.2 Multimodal features contribute to classification performance

Feature importance analyses showed that overall most features, and all modalities, contributed to classification accuracy. This finding emphasises the importance of multimodal assessments. Heart rate was the most important feature in both the out-of-bag observations in the training data and in the Heldout Test Dataset, which

is in line with previous reports suggesting that noxious-evoked heart rate responses can be measured from an early stage of development [343]. Some features had relatively low importance scores, however, this does not necessarily indicate that these features are not related to the noxious procedure or unimportant for prediction. They may be correlated with other model components, resulting in a minimal loss of accuracy when the values of this feature are removed or permuted. This was illustrated by the high classification accuracy in the UCL Dataset, which did not include reflex withdrawal nor the spectral features, which had high importance in the Oxford Dataset. As infant EEG studies often deal with relatively high attrition rates [139], it is encouraging that even in the absence of several features with high importance, good classification accuracy can be achieved.

Besides measurements that are commonly used in infant pain assessment scales, such as heart rate and the duration of facial grimacing, several neural features were incorporated in the model as predictors. This included the magnitude of the previously defined template of noxious-evoked EEG activity [131], as well as the magnitude of several newly identified waveforms and features extracted from the time-frequency domain. The identified PCs and their interpretation are discussed in detail in Chapter 4. In the spectral domain, an increase in power was mostly present in the low-frequency ranges, which is in line with a previous study in term infants undergoing a heel lance [90]. This is in apparent contrast to adults, where activity in the gamma range is preferentially associated with pain reports [144]. This may be explained by the fact that the resting infant EEG is generally dominated by lower frequencies than the adult EEG [268].

3.4.3 Limitations of feature selection and extraction procedure

The modalities and features of interest were chosen a priori based on the strongest evidence in the literature and their availability in the Oxford Training Dataset. It is possible that other useful modalities or features may have been missed. For instance, observational reports have described a wide range of noxious-evoked

responses in infants which were not included here [37, 85], including changes in cry amplitude [269], oxygen saturation [267, 56] (although reports vary [320, 274]), skin conductance [245, 86, 335] and salivary cortisol [267]. In addition, features can be extracted from the chosen modalities in various ways. For instance, heart rate variability could provide different information compared to the heart rate increase used here [35], EEG activity could be analysed at multiple electrodes, and all features could be measured at various time points [270]. It would be of great interest to investigate these features and modalities further in larger prospective studies. Deep learning models seek to learn to features in the raw data as part of model training [193] and would thus be useful in the future to use as a data-driven approach to feature extraction. However, deep learning models require more data than was available in the modest-sized Oxford Training Dataset [91].

Feature extraction involved some subjective decisions around artefact rejection. As infant movement can induce artefact in the EEG spectrum [110], traces were visually observed and any with clear gross movement artefact were not included in the EEG analyses. As this process is subjective, and not all movement is easily distinguishable from high-amplitude ongoing EEG activity in infants, it is possible that some movement remained in the EEG traces and contributed to the model discrimination. However, the fact that the model achieved high accuracy in the UCL Dataset, which did not include any of the spectral EEG features, indicates that the results are robust and unlikely to be driven largely by movement. Additionally, even if results were influenced by infant movement, this would not be dismissive of the fact that infant responses are discriminable between noxious and non-noxious procedures. It would, however, impact the biological interpretation of the model predictions. It would also affect the suitability of the model in infants who are on sedatives or muscle relaxants, just as it would for behavioural pain assessment tools [72].

The spectral features and PCs were extracted using the average EEG data in the the Oxford Training Dataset, causing information leakage between cross-validation folds [252]. This is a problem in situations where the estimate of model performance is only based on cross-validation, as it may introduce artificially high classification

results [252], but this was not an issue here as results were confirmed in two datasets that were kept completely separate from the model training.

3.4.4 Exploring clinical applications

The goal of this study was primarily to investigate the feasibility of discriminating infant noxious and non-noxious evoked responses and not to develop a tool that can immediately be used as an endpoint in a clinical trial, yet, this chapter provides a framework for building on the current findings to develop such a tool. Several follow-up investigations would be of interest.

3.4.4.1 Establishing a trade-off between false positives and false negatives

Besides accuracy, FPR, FNR and ROC curves were evaluated, because these represent the trade-off between false positives (non-noxious responses classified as noxious) and false negatives (noxious responses classified as non-noxious). In the Oxford Dataset, the FPR was higher than the FNR, which indicates that the model was sensitive in detecting noxious responses, but less specific (more prone to classify non-noxious responses as noxious than vice versa). In the UCL Dataset, this was the other way around. This implies that some predictors were more consistent within dataset than across datasets. These similarities and differences between datasets is further evaluated in chapter Chapter 4. The optimal trade-off between FPR and FNR is dependent on the clinical context. In some settings, one might prioritise the identification of all noxious-evoked responses at the cost of exposing some infants to analgesics when this was not strictly necessary — however this may be undesirable for drugs such as morphine, which has profound side effects [134, 231]. At the other extreme, one could run the risk of under-treating infants who might have benefited from analgesics, and exposing infants to potential short-term clinical instability [222]. If one scenario prevails over the other, the threshold of the classifier can be varied to accommodate this. It would be of interest to combine data of the two centres in a single model and test these in new data from both sites, to assess whether the FPR and FNR become more consistent across datasets.

3.4.4.2 Evaluation of the classifier within the COSMIN framework

As described in Chapter 1, COSMIN guidelines emphasise that health measurement instruments should be valid, reliable and responsive [66, 64, 65].

Construct validity *Construct validity* is defined as "the degree to which the scores of a measurement instrument are consistent with hypotheses" [227]. The construct validity of the classifier predictions was assessed by testing the hypothesis that the predictions would be different for a non-noxious control heel lance compared to a noxious heel lance. While reasonable accuracy was achieved, it is likely that accuracy (and thus construct validity) can be further improved by performing systematic hyperparameter optimisation (e.g. investigating the effect of varying the minimum leaf size or number of trees in the ensemble), or by using alternative algorithms, such as deep neural networks, in larger datasets [193].

In this study, the classifier was trained on responses to two specific stimuli: heel lances and control heel lances. As infants undergo many different types of clinical procedures, future investigations should test the hypothesis that a model like this one generalises to other clinical procedures. A first step was already made in this study by applying the classifier to the Immunisation Dataset, and showing that immunisation responses were identified as "noxious" in the majority of cases. The control heel lance is an excellent negative control in a research setting, because it emulates all aspects of the heel lance, except the skin-break. However, it is not a procedure that infants would naturally encounter on the NICU, in contrast to many other more salient stimuli, such as sound and light [209]. Therefore it should also be tested whether noxious-evoked responses are discriminable from responses to other, non-noxious, but salient sensory stimuli, as is being done in adults [233]. Construct validity was also tested by comparing the prediction scores between infants receiving paracetamol before an immunisation and infants receiving paracetamol after an immunisation. Prediction scores were significantly lower in the group of infants who received paracetamol before the immunisation, suggesting that the classifier's scores are sensitive enough to detect differences in noxious-evoked responses brought upon

by analgesic treatment. The sample size in the current analgesic study was small, missing data was not completely balanced across the two treatment groups and the effect size was modest, therefore these findings should be corroborated in larger studies. Yet, this example illustrates that there is a place for the scores of multimodal models to be used as outcome measures in neonatal clinical trials of analgesics.

Reliability A measurement instrument's *reliability* addresses its measurement errors, for example across sessions or between raters [64]. Reliability for the classifier was assessed in the current study by using data from two different sites, and showing that the accuracy was comparable. Follow-up studies should include multiple observations from the same infant under the same conditions, to gain understanding of the variance in prediction scores within infant. Although the classifier generates automatic predictions, the data from which the features were extracted were assessed for quality by a human rater and rejected if quality was insufficient, which in theory can introduce variability in predictions across different raters or across time. In addition, the brow bulge feature was rated by human investigators. Thus, in the future, data should either be automatically quality assessed and extracted (without subjective human intervention) or features should be extracted by different raters and repeatedly by the same rater to assess the level of intra- and inter-rater reliability. The automated detection of artefacts in EEG recordings [340, 344] as well as noxious-related facial features from images [143, 270] are in full development and would be a valuable addition to the classifier described here.

Responsiveness Finally, *responsiveness* refers to an instrument's ability to detect change in the construct of interest over time [65], for instance in response to an analgesic. For many infant pain scales, responsiveness to analgesics has not been investigated, which calls the value of these measurement instruments into question [213]. Responsiveness to analgesics is measured by comparing scores before and after the administration of an analgesic within a single infant. This was not possible here due to the nature of the datasets. However, evaluating the within-infant difference between the heel lance prediction score and the control heel lance prediction score

(which likely reflects a change in the construct of pain) was a tentative first step in investigating classifier responsiveness.

Interpretability Besides validity, reliability and responsiveness, COSMIN emphasises the value of *interpretability* of measurement instruments [63]. This characteristic is defined as "the degree to which one can assign qualitative meaning – that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores" [227]. For the classification model developed here, a classifier score of 1 represents the response to a heel lance, while a score of 0 indicates the response to a control heel lance. It is not straightforward to attach meaning to any score in between, because scores are generated by automatically weighing various features in the model and are not a single direct measurement of a clinically relevant domain. However, because the individual predictors were chosen because they are clinically or biologically meaningful and may represent different aspects of the construct of pain (cerebral, physiological, behavioural), post-hoc tests can be conducted to investigate whether there were any features in the model that were preferentially modulated by the analgesic.

3.4.5 Summary and next steps

In this chapter I showed that responses to noxious and non-noxious stimuli are discriminable in infants aged 28–40 weeks PMA, and that even in very young infants it is possible to identify multimodal discriminative patterns of noxious-evoked responses. Further development is necessary to translate the findings from this study to a clinically usable model. The goal of this study was to identify discriminative features across a wide age range, however it is likely that the included features vary with age. The EEG PCs which were extracted from the noxious data are of particular interest in this respect, as these include not previously described waveforms. Therefore, in the next chapter, I investigate how multimodal noxious-evoked responses change with age.

Acknowledgements

I would like to thank Laura Jones and colleagues for sharing the UCL Dataset through the UK Data service.

4

Development of noxious-evoked responses

Part of the work in this chapter and Chapter 3 was published under a CC-BY license in *Cerebral Cortex* in a paper on which I am the first author [321]. A preprint was also deposited under a CC-BY license on *medRxiv* [322]. The published work was restructured to comply with the format of a thesis and expansions were made throughout the introduction, methods, results and discussion sections.

4.1 Introduction

The preterm period in infants is characterised by rapid maturational changes in the somatosensory nervous system (see Chapter 1). Top-down modulation of the spinal cord dorsal horn leads to a refinement of peripheral responses [95], while in the brain, connections are formed between the thalamus and cortical targets [127]. These structural changes are accompanied by changes in functional activity. The somatosensory resting state network emerges and strengthens [77], the infant background EEG changes from a pattern of high-amplitude, low-frequency intermittent bursting to continuous activity with lower amplitude and higher frequency content [240, 7, 236] and autonomic nervous system tone increases [235]. Previous studies have highlighted important developmental changes in noxious-evoked responses. Both facial reactivity [120, 164] and cardiovascular responses

[56, 253] to noxious stimuli are low in very preterm infants, and increase with age. In contrast, nociceptive reflexes in preterm infants are of longer duration and greater magnitude than in term infants [133, 53]. It it is important to understand how age influences noxious-evoked responses, as this impacts the design and interpretation of multimodal neonatal pain assessment tools. This was highlighted by observations that NIPS scores increase with infant age [348] and in a recent study, which showed that the clinical usefulness of the Bernese Pain Scale, which includes behavioural and physiological measurements, would be improved if different cut-off values for detecting "pain" were used in different age groups [274]. As heart rate and behavioural features contributed to the classification model derived in Chapter 3, it is highly relevant to understand their developmental trajectories.

While these changes concern response magnitudes, a striking morphological change has previously been described for noxious-evoked ERPs. A landmark study by Fabrizi and colleagues [89] showed that time-locked EEG activity undergoes a transition from non-specific delta brushes, which are evoked by tactile and noxious stimuli alike, to a clearly delineated noxious-evoked ERP [89]. This ERP consists of two N-P complexes, of which the first (100–300 ms post-stimulus) is present in response to noxious and non-noxious stimuli alike, while the second (400–700 ms post-stimulus) is preferentially evoked by noxious stimuli [131] and increases in magnitude with age [133, 275] (see Chapter 1). The shift from delta brush responses to ERPs is estimated to occur between 35–37 weeks PMA [89] and is related to the emergence of discriminative facial expressions [120].

The studies investigating the maturation of time-locked EEG responses in premature infants to date [89, 120, 275] have primarily focused on delta brushes and the characteristic noxious-evoked ERP. However, as described in Chapter 3, the state of the preterm infant nervous system does not preclude differential processing of noxious and non-noxious stimuli, and it is possible that that intermediary discriminatory ERP stages exist. Indeed, NIRS studies have shown differential patterns of oxygenated blood flow to noxious and non-noxious stimuli even in very preterm infants [288] and in Chapter 3 newly derived PCs were found to contribute

to the discrimination of noxious and non-noxious responses. Further characterising age-related morphological changes in noxious-evoked EEG activity will improve understanding of the underlying functional changes in the infant nervous system and could pave the way for the use of brain-derived measures of infant pain in multimodal pain assessment tools in very preterm infants, in the same way that the template of noxious-evoked EEG activity [131] is being implemented in clinical trials in infants from 34 weeks PMA [134, 44].

This chapter therefore investigates the development of multimodal noxious-evoked responses in infants aged 28–40 weeks. The development of each modality included in the classification model in Chapter 3 is characterised, with special emphasis placed on the development of noxious-evoked ERPs.

This goals of this chapter were to:

1. Characterise the developmental changes in multimodal noxious-evoked responses in infants aged 28–40 weeks PMA;

2. Characterise morphological changes in noxious-evoked ERPs, evaluate their discriminative ability and explore their relationship to other modalities.

4.2 Methods

4.2.1 Chapter overview

In the first part of this chapter, the developmental trajectories of the individual modalities included in the classification model in Chapter 3 are presented. The second part of the chapter focuses on the morphology of developmentally-distinct noxious-evoked ERPs, their discriminative ability and their relationship to behavioural and cardiovascular modalities. The third part of the chapter includes exploratory analyses into the reproducibility of the findings across datasets and the influence of PNA and other clinical factors on the primary analysis.

A graphical overview of the analyses in this chapter can be found in Fig. 4.1.

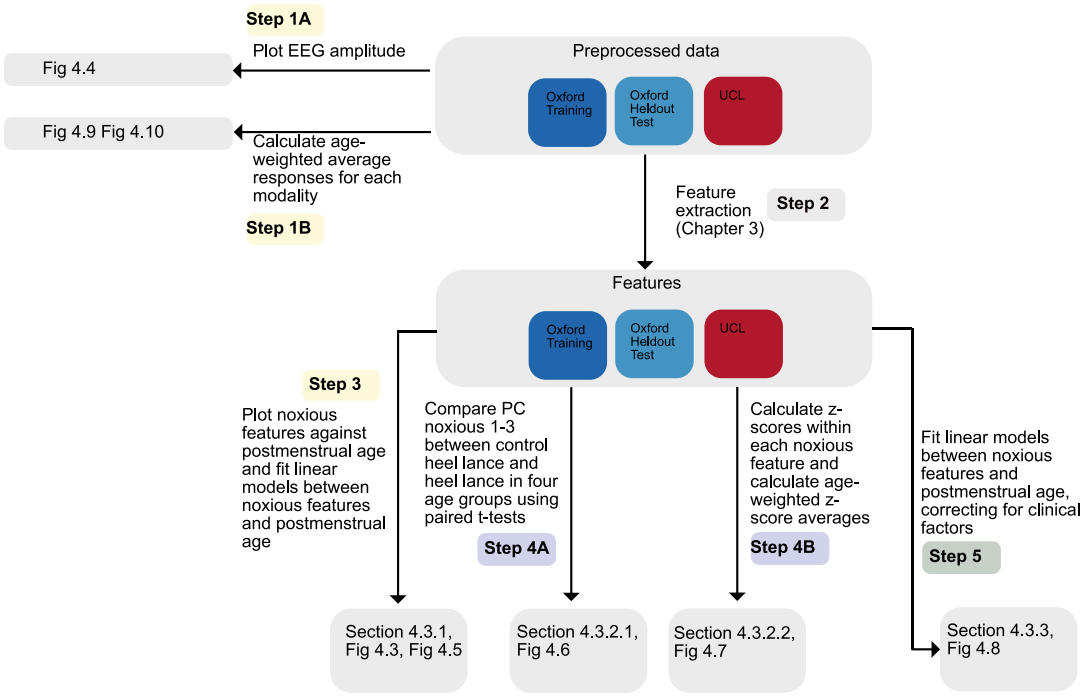


Figure 4.1: Graphical overview of the analysis steps in Chapter 4. Part 1 (Steps 1A, 1B, 2, 3): The development of the modalities included in the classification model in Chapter 3 was qualitatively examined. Step 1A: The amplitude of baseline-corrected EEG trials at Cz was visually examined (Fig. 4.4). Step 1B: Age-weighted average responses were constructed for each modality (Fig. 4.9, Fig. 4.10). Step 2: Features were extracted from each modality (see Chapter 3). Step 3: A subset of noxious features were related to postmenstrual age (Fig. 4.3, Fig. 4.5). Part 2 (Steps 4A and 4B) : The developmentally-distinct noxious-evoked event-related potentials were further investigated. Step 4A: To assess the discriminative ability of PC 1-3 noxious, the magnitude of PC 1-3 noxious was compared between the control heel lance and heel lance procedure in different age groups (Fig. 4.6). Step 4B: To compare developmental trajectories of the noxious-evoked ERPs with those from other modalities, z-scores were calculated within each feature and age-weighted average z-scores were plotted (Fig. 4.7). Part 3 (Step 5): Finally, the impact of clinical factors, such as postnatal age, was explored by fitting linear models between the noxious features and postmenstrual age, while correcting for these factors (Fig. 4.8).

4.2.2 Datasets

The Oxford Heel Lance Dataset (n = 70) and the UCL Dataset (n = 74, see Chapter 2 and Chapter 3) were used in this chapter. In brief, these datasets included infants aged at least 28 weeks PMA and less than 40 weeks PMA who underwent a clinically-required noxious heel lance and a non-noxious control heel lance. Behavioural, cardiovascular and neurophysiological recordings were obtained

around each stimulus. Datasets, recording methods and stimuli are described in more detail in Chapter 2 and Chapter 3. For statistical analyses, the two datasets were combined to increase sample size. In figures, the two datasets are plotted in different colours in the same axes to show the (dis)similarity of the datasets collected at different sites. The consistency of the results across the two datasets is discussed in Section 4.2.5.

4.2.3 Part 1: Developmental trajectories of multimodal noxious-evoked responses

The development of the multimodal noxious-evoked responses described in Chapter 3 was investigated both qualitatively and quantitatively. Modalities and the features derived from them are described in detail in Chapter 3, Table 3.2. In brief, features were derived from the following modalities: cardiovascular (heart rate increase), limb withdrawal (EMG RMS magnitude in one second post-stimulus), facial expressions (duration of brow bulge) and cerebral ($ERSP_{log}$ in five time-frequency windows; ERPs, summarised as the magnitude of PCs and the predefined template of noxious-evoked EEG activity [131]).

The development of each feature was investigated with respect to PMA at test occasion, which is the sum of GA (time spent *in utero*) and PNA (time elapsed since birth). PMA was deemed a clinically more relevant age metric compared to GA (which ignores rapid postnatal brain development) and PNA (which ignores the significant impact of preterm birth on brain development). The contribution of PNA specifically to development was investigated in an exploratory analysis in Section 4.2.5.

For qualitative assessment, the developmental trajectories of ERPs, heart rate, reflex withdrawal and facial expression activity were visually assessed using smoothed group responses constructed around 29–39 weeks PMA. For the ERP and heart rate data, age-weighted averages were calculated in 4-week sliding windows across the centre PMA. For the limb withdrawal data, age-weighted medians were calculated in 4-week sliding windows centred around 29–39 completed weeks PMA using the

function *weightedMedian* [126]. Within a window, each infant's data was assigned a weight between 0 and 1 based on the difference in the corresponding infant's PMA in days and the window's centre PMA. Smoothed group responses for the magnitudes of PC 1 noxious, PC 2 noxious, PC 3 noxious and the ERSP features were also constructed around 28–40 weeks PMA, for visualisation purposes only (see Fig. 4.5 and Fig. 4.3). The MATLAB function *gausswin* was used to derive weights w for each day of difference n from a gaussian window. In a given sliding window, weights were computed as follows:

$$w(n) = e^{-\frac{1}{2}\left(a\frac{n}{(L-1)/2}\right)^2}$$

a is the width factor which determines the width of the gaussian curve and L is the sliding window length in days + 1. For this analysis $a = 2.5$, which is the default value in MATLAB's *gausswin*, $L = 29$ and $-14 \geq n \leq 14$. Fig. 4.2 shows that infants who were were one week younger or older than the centre PMA were assigned a weight of approximately 0.45. Infants who were more than 2 weeks older or younger than the centre PMA were assigned a weight of 0. Brow bulge duration was expressed as the proportion of infants displaying a brow bulge response of any length in a cohort of infants in a 3-week sliding window. These plots were generated for the noxious and non-noxious responses separately.

Next, relationships between PMA and a subset of the noxious features (noxious-evoked heart rate, PC 1 noxious, PC 2 noxious, PC 3 noxious, early delta, early alpha, late delta, late alpha, late beta, ipsilateral reflex and contralateral reflex, see Table 3.2) were quantitatively investigated with general linear models with p-values derived non-parametrically using 10000 permutations in FSL PALM [350]. The relationship between PMA and brow bulge duration was not formally tested in this way, as these results were previously reported for a partly overlapping cohort [120]. Tests were not conducted for PC 1-3 non-noxious and PC 1-3 all data, as these features were derived (partly) from the non-noxious data, and the template of noxious-evoked EEG activity was not tested because this waveform resembles the latter part of PC 3 noxious (see Fig. 4.5). As each of the 11 features was

tested separately, the p-values were corrected for multiple comparisons using the Holm-Bonferroni method [140] implemented in R (version 4.1.0, number of tests = 11). Corrected p-values below 0.05 were deemed statistically significant.

The magnitude of PC 1-3 noxious was compared between the noxious and the non-noxious stimuli using two-tailed paired t-tests with p-values derived non-parametrically in FSL PALM [350] using 10000 sign-flips. Holm-Bonferroni correction was used to correct for multiple testing (number of tests = 3).

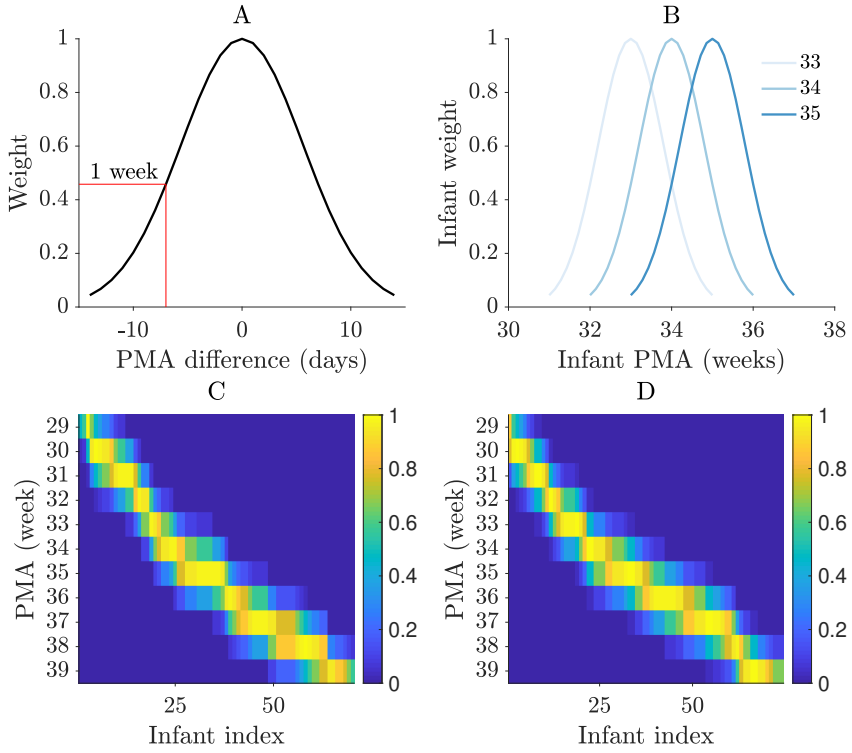


Figure 4.2: Age-related development of the continuous features was qualitatively assessed by constructing age-weighted averages. A) A gaussian window was used to assign weights to construct the age-weighted average for each feature. The window size was 4 weeks with 14 days on either side of the centre postmenstrual age (PMA). The weight corresponding to a 1-week difference from the centre PMA is indicated in red. B) Example: weights for the sliding windows centred around 33, 34 and 35 weeks PMA by infant PMA. For the centre PMA of 34 weeks, the sliding window includes data from infants aged 32+0 weeks PMA up to and including 36+0 weeks PMA. Infants with a PMA of 34 weeks have a weight of 1. The gaussian window is shifted across each centre PMA. C) Weights for each infant in each sliding window centred around 29–39 weeks PMA in the Oxford Dataset. Infant index is obtained by sorting infants by their PMA at study. Colour bar indicates the assigned weight. D) Weights for each infant in each sliding window centred around 29–39 weeks PMA in the UCL Dataset. Colour bar indicates the assigned weight. For some features, some infants had missing data (see Table 3.3). This is not depicted here.

4.2.4 Part 2: Investigation of morphological changes in noxious-evoked ERPs and their relationships

4.2.4.1 Discriminative ability of noxious-evoked ERPs

The second part of the chapter focuses on the morphological changes in noxious-evoked ERPs. First, I qualitatively investigated whether the observed noxious-evoked ERPs were discriminative by comparing the ERPs evoked by the heel lance to those evoked by the control heel lance. I visually compared the noxious and non-noxious evoked ERPs in the weighted average plots and in the four non-overlapping 3-week age bins which were introduced in Chapter 3: 28–31 (non-inclusive) weeks, 31–34 (non-inclusive) weeks, 34–37 (non-inclusive) weeks and 37–40 (non-inclusive) weeks PMA. The number of infants in each bin was n = 19, n = 26, n = 56, n = 43, respectively.

To quantitatively investigate whether the discriminative ability of the 3 PCs derived from the noxious data (PC 1 noxious, PC 2 noxious and PC 3 noxious) changed with age, PC magnitude to the heel lance was compared to PC magnitude to the control heel lance in the four age bins. This was achieved by using paired t-tests with one-tailed p-values derived non-parametrically in FSL PALM [350] using 10000 sign-flips. One-tailed tests were conducted because the goal was only to test if the noxious PC magnitudes would be higher in response to the heel lance compared to the control heel lance. As there were 3 PCs and 4 age groups, this resulted in a total of 12 statistical tests, and p-values were corrected using the Holm-Bonferroni method (number of tests = 12). Corrected p-values below 0.05 were deemed statistically significant. As the PCs were derived from the noxious data in the Oxford Training Dataset, there was a risk of bias towards higher PC magnitudes for the noxious stimulus compared to the non-noxious stimulus in this dataset, which affects the generalisability of the findings. The differences are therefore presented both for the combined Oxford + UCL dataset and separately for the fully independent UCL Dataset, in which this bias is not present.

4.2.4.2 Multimodal patterns of noxious-evoked responses

In order to qualitatively compare the developmental trajectories of the noxious-evoked PCs and the other features, noxious-evoked responses were normalised within feature to z-scores by subtracting the mean from each observation and dividing this by the standard deviation. An age-weighted average of the z-scores was calculated in 4-week sliding windows as described above for the individual features (see Section 4.2.3). Regression coefficients were calculated between PMA and z-scores and features were plotted in the order of the regression coefficient magnitude. This allowed the visual identification of clusters of features with similar trajectories.

4.2.4.3 Relationship between discriminative PCs and behavioural and autonomic domains

To investigate whether the presence of discriminative PCs was related to changes in the behavioural and autonomic domains, the PCs that were discriminative in at least one age category were correlated with the features extracted from the non-neural domains to investigate whether there was a relationship between the magnitude of the PCs and other modalities. This was achieved by performing Pearson correlations between the magnitude of the discriminative PCs to the noxious stimulus and the noxious-evoked heart rate, facial and limb withdrawal features, while correcting for PMA. p-values were derived non-parametrically in FSL PALM [350] and p-values were corrected using the Holm-Bonferroni method (number of tests = 8). Corrected p-values below 0.05 were deemed statistically significant.

4.2.5 Part 3: Exploratory Analyses: Reproducibility of developmental trajectories and effect of PNA and other clinical factors

As described above, the main aim of this study was to investigate the role of PMA in the development of noxious-evoked responses across all available data. However, as PMA is composed of gestational age + postnatal age, some effects may be specifically attributed to either PMA, GA or PNA. In addition, prior pain exposure and sex [338, 331] may influence noxious-evoked responses. Therefore, additional

exploratory analyses were performed to investigate effects that could be attributed to PMA specifically when adjusting for PNA, prior pain and sex. To identify the effect of PNA, prior pain exposure and sex on the primary PMA association results, multivariate linear models were fit as described above with PMA, PNA, estimated number of skin-breaking blood tests and sex as covariates. For sex, female sex was set as the default. As these analyses were exploratory, p-values are presented to provide information about the strength of the relationships between each clinical factor and the relevant features, but these were not corrected for multiple testing. Theoretically, GA (irrespective of PMA) could also be of influence on developmental trajectories. However, as PNA was restricted to 8 weeks, GA and PMA were forced to be highly correlated and it was impossible to investigate GA specifically.

Differences in developmental trajectories may be present across the two datasets. The consistency of the developmental trajectories across datasets was assessed by visual comparison of the Oxford and UCL average plots. Linear models were also fit in each dataset separately to assess whether the regressors had the same sign in both datasets (equivalent to the analyses performed across the combined dataset in part 1). The goal was not to precisely estimate parameters in the two datasets separately, but to identify any large discrepancies between the two datasets, and therefore p-values are not presented.

4.3 Results

4.3.1 Part 1: Developmental trajectories of multimodal noxious-evoked responses

A summary of noxious-evoked developmental trajectories is displayed in Fig. 4.9 and described in detail in the following sections.

4.3.1.1 Development of noxious-evoked changes in facial expressions, heart rate, reflex withdrawal and ERSP features

Noxious-evoked heart rate responses showed an increase with PMA ($n = 130, beta = 1.69, t = 7.08$, Holm-corrected $p = 0.0011$, Fig. 4.9). In contrast, reflexive limb withdrawal magnitude in the first 1000 ms post-stimulus was stable across 28–40 weeks PMA (ipsilateral limb: $n = 68, beta = -0.073, t = -0.39$, Holm-corrected $p = 1.0$; contralateral limb: $n = 67, beta = 0.21, t = 0.90$, Holm-corrected $p = 1.0$, Fig. 4.9). The development of facial responses in the Oxford Dataset was previously described in in a larger-scale analysis that included a large subset of the infants included in the Oxford dataset [120] and was therefore not tested here. Of the ERSP features included in the model, early alpha ($n = 66, beta = -0.56, t = -3.12$, Holm-corrected $p = 0.017$) and early delta ($n = 66, beta = -0.53, t = -2.77$, Holm corrected $p = 0.047$) significantly decreased with PMA, while the other features remained relatively stable (Fig. 4.3).

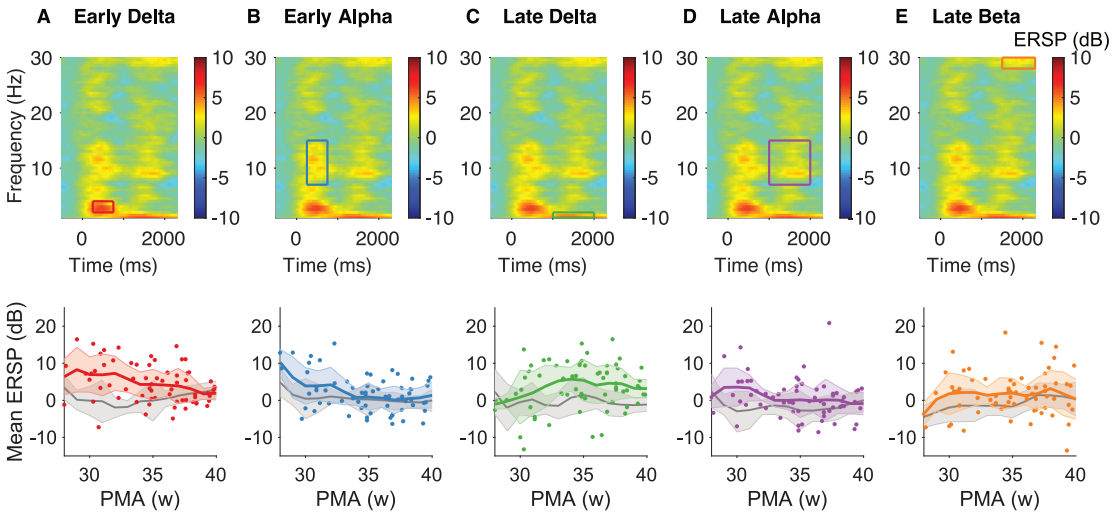


Figure 4.3: Developmental trajectories of the five event-related spectral perturbations (ERSP) features. Upper row shows the average $ERSP_{log}$ in the Oxford Dataset. Time windows corresponding to the five ERSP features are indicated by boxes. Lower row shows the developmental trajectories for each of the features. Each dot represents a single infant's noxious-evoked response. Colour and grey lines are the smoothed averages for each feature for the noxious and the non-noxious data, respectively. Shaded areas depict the age-weighted standard deviation corresponding to the colour lines. This figure is my own work, and was reproduced from [321] under the terms of a CC-BY license.

4.3.1.2 Development of noxious-evoked ERP morphology

Qualitative examination of noxious-evoked EEG amplitude changes showed a continuous transition from a predominantly negative amplitude pattern beginning around 400–500 ms post-stimulus to a predominantly positive amplitude pattern just after 250 and 500 ms post-stimulus (Fig. 4.4A). This change was also visible in the age-weighted average traces (Fig. 4.9), which showed a slow negative wave in the very youngest infants, a negative deflection approximately 450 ms post-stimulus from 30 weeks PMA onwards and an early (approximately 250 ms post-stimulus) and late (approximately 500 ms post-stimulus) positive deflection in the older infants.

These changes were captured in the three PCs that explained most of the variance in the noxious-evoked EEG data (Fig. 4.5). The first stage (approximately 29 weeks PMA and below) was captured by PC 1 noxious, which was a slow wave in the delta range with superimposed activity in the higher frequencies (Fig. 4.5, A, D). The second stage was captured by PC 2 noxious, a negative deflection with a peak latency at approximately 445 ms (Fig. 4.5, B, E). Finally, the last stage was visible in PC 3 noxious, which was composed of an early positive-negative deflection (latency: 214 ms and 429 ms respectively) followed by a second positive deflection at approximately 595 ms and resembling the previously published template of noxious-evoked EEG activity in term infants [131] (Fig. 4.5, C, F) and other reports of noxious-evoked ERPs in term infants [290, 89]. PC magnitudes were different between the noxious and non-noxious events. PC 1 was significantly smaller for the heel lance trials (paired t-test, n = 144, t = 2.70, Holm-corrected p = 0.0082), while PC 2 and PC 3 were significantly larger in the heel lance trials compared with the control heel lance trials (paired t-test, PC 2; n = 144, t = -5.32, Holm-corrected P = 0.0003, PC 3; n = 144, t = -5.14, Holm-corrected P = 0.0003),

The magnitudes of the three PCs were used to quantitatively investigate infant noxious-evoked ERP development. As can be seen in Fig. 4.9 and Fig. 4.5G-H, PC 1 noxious significantly decreased in magnitude with increasing PMA (GLM, $n = 144, beta = -0.016, t = -5.30$, Holm-corrected $p = 0.0011$), while PC 2 was maximal at approximately 30 weeks PMA and decreased in magnitude until term age

(GLM, $n = 144, beta = -0.023, t = -6.89$, Holm-corrected $p = 0.0011$). In contrast, PC 3 noxious increased in magnitude with age (GLM, $n = 144, beta = 0.013, t = 3.74$, Holm-corrected $p = 0.0040$, Fig. 4.5I, [321]).

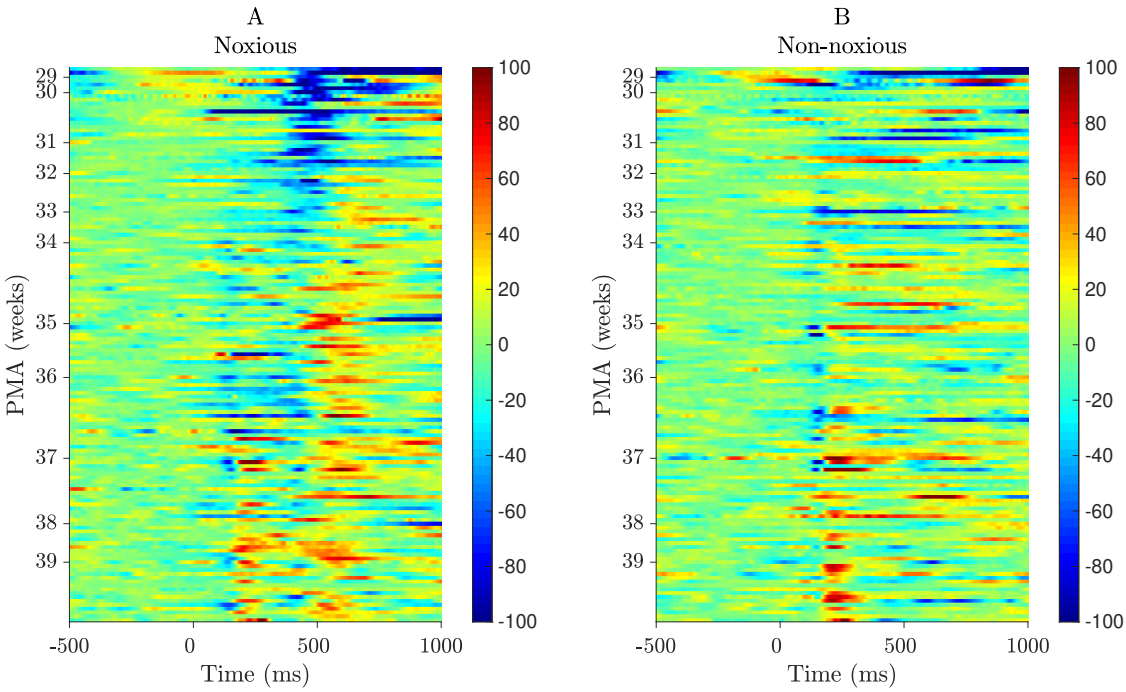


Figure 4.4: Amplitude of individual baseline-corrected trials for each infant at the Cz electrode, sorted by postmenstrual age (PMA). Each row corresponds to one infant, colour bar depicts amplitude in μV. The stimulus occurred at time = 0 ms. A) Noxious heel lance trials. In the youngest infants, a negative amplitude change is present around 450 ms. In the oldest infants, a positive amplitude change is present around 100–300 ms and around 500 ms. B) Non-noxious control heel lance trials. In the youngest infants, predominantly negative amplitude changes are visible. In the oldest infants, a positive amplitude change is present around 100–300 ms post-stimulus. For both the noxious and the non-noxious trials, the evoked amplitude changes from a predominantly negative pattern to a predominantly positive pattern of activity.

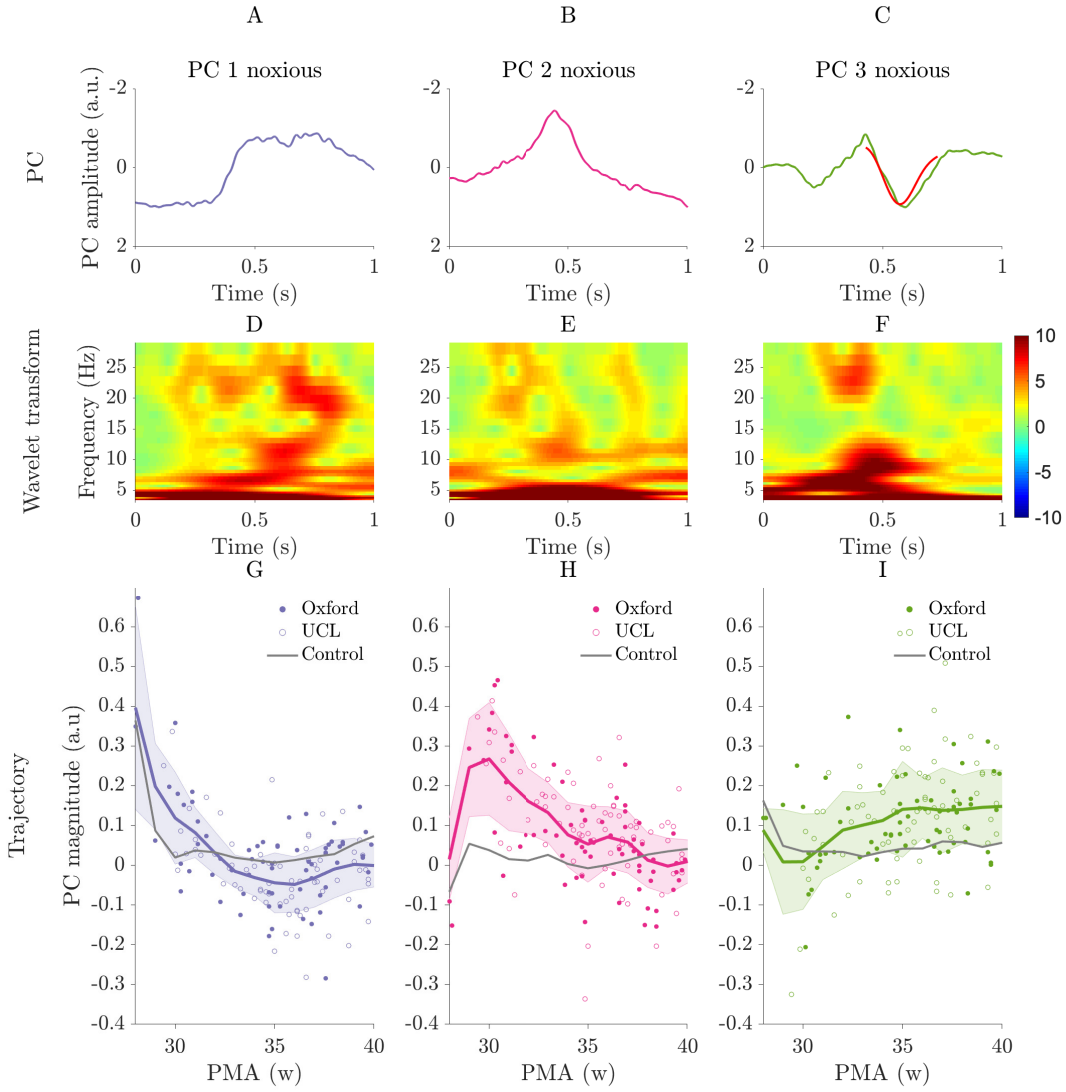


Figure 4.5: Characterisation of the three PCs that explained most of the variance in the noxious-evoked EEG data. Upper (A-C): PC 1 noxious, PC 2 noxious and PC 3 noxious, scaled to the maximum value within each PC. The template of noxious-evoked EEG activity [131] is scaled to fit PC 3 noxious and overlaid in red to show the similarity between the latter part of PC 3 noxious and the template. Middle (D-F): corresponding wavelet transforms. Colour bar represents power in dB. Lower (G-I): corresponding developmental trajectories. Each dot represents an individual infant's noxious-evoked PC magnitude. Solid lines indicate smoothed age-weighted averages for the PC coefficients evoked by the noxious heel lance across 28–40 weeks PMA, while the shaded area depicts the corresponding age-weighted standard deviation. Grey lines indicate smoothed age-weighted averages for the PC coefficients evoked by the control heel lance (non-noxious) stimulus. a.u. = arbitrary units. This figure is my own work, panels A-C and G-I were published in edited format in [321] and reproduced under the terms of a CC-BY license.

4.3.2 Part 2: Investigation of morphological changes in noxious-evoked ERPs and their relationships

ERPs with a similar morphology to PC 1 and PC 3 noxious have been previously described [89, 290], whereas PC 2 represents a newly identified noxious-evoked waveform. The three ERP stages were further investigated with regards to their discriminative ability and their relation to other noxious-evoked responses.

4.3.2.1 Differences in ERPs evoked by noxious and non-noxious stimuli were present throughout development

To assess whether the three ERP stages were discriminative between the heel lance and the control heel lance at different stages of development, the noxious-evoked ERPs were compared to the ERPs evoked by the non-noxious control heel lance. As depicted in Fig. 4.4B and Fig. 4.10, EEG activity evoked by non-noxious stimulation also changed morphology with PMA. The response in the youngest infants consisted of a slow wave with superimposed high-frequency activity, which transitioned to a single early negative peak around 150 ms from approximately 33 weeks PMA onwards, and finally a N-P complex between approximately 100–300 ms from 36 weeks onwards. Qualitatively, the negative peak between 33–36 weeks was less consistently present across individual infants than the positive peak from 36 weeks onwards (Fig. 4.4B). Each of the peaks identified in the non-noxious traces could also qualitatively be observed in the noxious traces (Fig. 4.4A, Fig. 4.9). Thus, the early peaks (before 300 ms post-stimulus) evoked by the non-noxious stimuli were common to both stimuli, whereas the late peaks evoked by the noxious stimulus (from 450 ms post-stimulus) described in Section 4.3.1.2 were only present in response to the noxious stimulus.

A direct comparison of the ERPs evoked by the control heel lance and the heel lance in 3-week age bins highlighted qualitative differences between the noxious and non-noxious responses in every age group (Fig. 4.6, upper rows). When comparing the magnitude of the noxious PCs between the heel lance and the control heel lance in the same 3-week age bins using paired t-tests, it was revealed that PC 2 noxious

magnitude was significantly larger for the heel lance compared to the control heel lance in the 28–37 week bins, while PC 3 noxious magnitude was higher for the heel lance than for the control heel lance in the bins from 34 weeks onwards (one-tailed paired t-tests, p-values depicted in Fig. 4.6). The period from 34 to 37 weeks PMA was thus a transitory period during which both PC 2 noxious and PC 3 noxious were discriminative (Fig. 4.6). PC 1 noxious magnitude to the heel lance was not significantly larger than PC 1 noxious to the control heel lance in any age group.

The three PCs were derived in the noxious data from the Oxford Dataset, which may bias the magnitude of the PC differences in this dataset. Therefore, the same analysis was also conducted in the fully independent UCL Dataset only. The results were largely the same (Fig. 4.6), except PC 2 was not significantly discriminative in the 28–31 week age group (paired t-test, Holm-corrected $p = 0.19$). Reproducibility between the Oxford and UCL dataset is further addressed in Section 4.3.3.

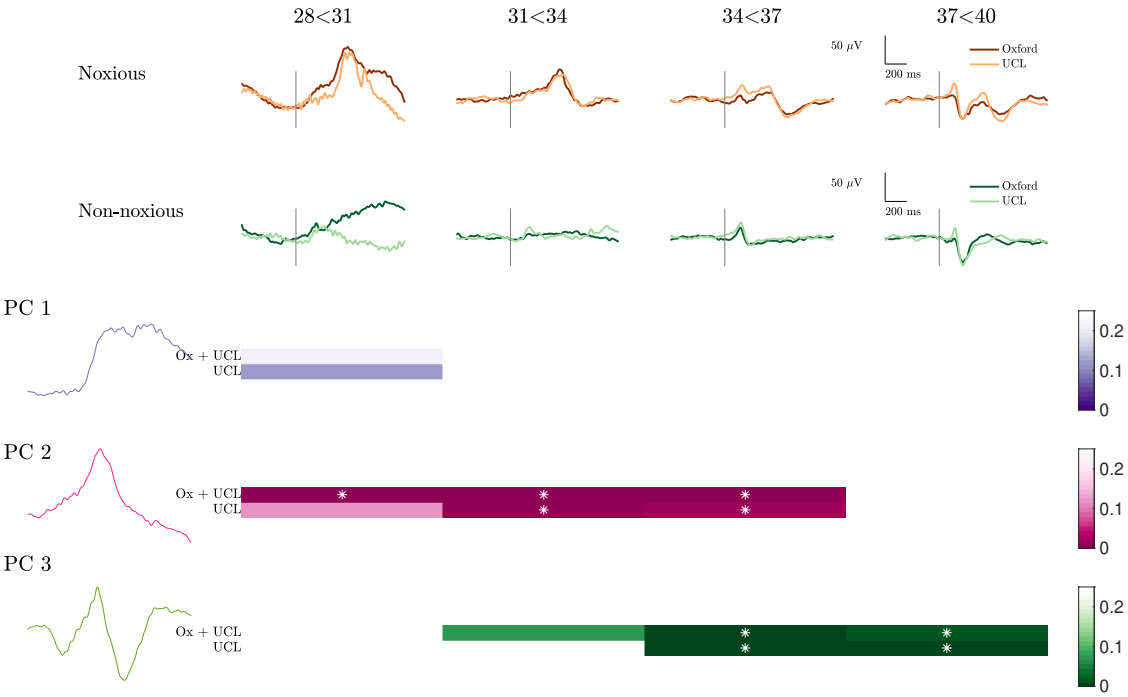


Figure 4.6: Comparison of noxious and non-noxious evoked event-related potentials in different age groups: 28–31 (non-inclusive) weeks, 31–34 (non-inclusive) weeks, 34–37 (non-inclusive) weeks and 37–40 (non-inclusive) weeks postmenstrual age (PMA). Upper rows show the mean noxious and non-noxious traces in the Oxford Dataset and the UCL Dataset in four non-overlapping age bins. Qualitative differences between the two stimuli can be observed in every age group. The noxious-evoked ERP in the youngest age groups resemble PC 1 noxious and PC 2 noxious, while the noxious-evoked ERP in the oldest age group resembles PC 3 noxious. In the three lower rows, the coloured bars indicate one-tailed Holm-Bonferroni corrected p-values for the comparison of the magnitude of PC 1, PC 2 and PC 3 noxious for the heel lance versus the control heel lance in the four age bins. Asterisks indicate the age bins in which $p < 0.05$. P-values are presented for the combined Oxford and UCL Datasets ("Ox + UCL") and for the independent UCL Dataset only ("UCL"). Only PC 2 is discriminative in the 31–34 week age group and only PC 3 is discriminative in the 37–40 week age group, whereas 34–37 weeks PMA is a transitory period in which both PC 2 and PC 3 are discriminative. PC 1 is not significantly larger in the heel lance responses compared to the control heel lance responses in any age group.

4.3.2.2 Multimodal patterns of noxious-evoked activity changed throughout development

As can be seen in Fig. 4.7, the discriminative PC 2 and PC 3 were part of two visually distinct clusters of features. PC 2 noxious was greater in infants below 33 weeks and decreased in magnitude with increasing PMA. The same pattern was apparent for PC 1 noxious, early alpha, early delta and late alpha. PC 3 noxious, heart rate increase, brow bulge duration, the magnitude of the template of noxious-evoked

EEG activity and late delta, showed the opposite pattern, with larger responses in infants above 36 weeks and smaller responses in infants below 33 weeks.

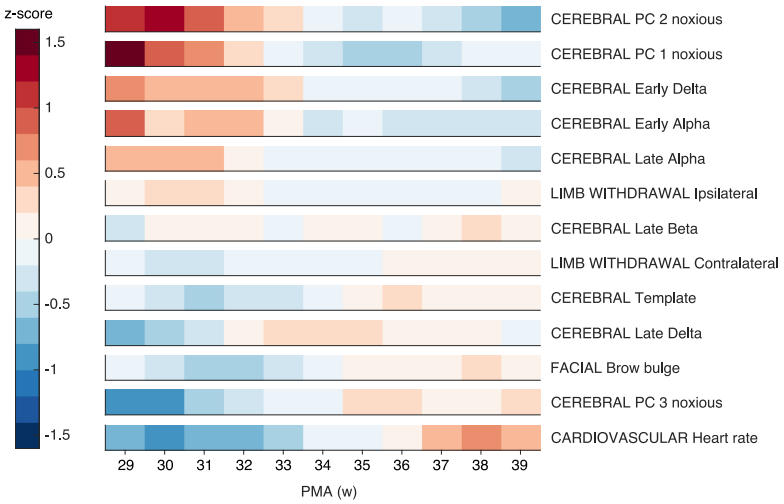


Figure 4.7: Normalised magnitude of noxious-related responses at 29–39 weeks postmenstrual age, separated by feature. Two clusters can visually be identified (see Section 4.3.2.2). This figure is my own work and was reproduced with permission from [321] under the terms of a CC-BY license.

4.3.2.3 Noxious-evoked PCs were not significantly correlated with cardiovascular and behavioural measures

To explore whether the emergence of the discriminative PC 2 and PC 3 noxious could potentially drive changes in the non-cerebral modalities modalities, the relationship of PC 2 and PC 3 to heart rate, reflex withdrawal and brow bulge duration was investigated using two-tailed partial Pearson correlations, corrected for PMA. All correlations were weak and not statistically significant. The highest correlation was found between PC 2 and PC 3 magnitude and the duration of brow bulge (PC 2: $n = 135, r = 0.14$, Holm corrected $p = 0.63$; PC 3: $n = 135, r = 0.18$, Holm corrected $p = 0.25$). All other correlation coefficients had Holm corrected p-values of 1.

4.3.3 Part 3: Exploratory Analyses: Reproducibility of developmental trajectories and effect of PNA and other clinical factors

Finally, I explored the consistency of the results presented across the Oxford and the UCL Dataset and the importance of PNA and other clinical variables

in age-related changes.

Developmental trajectories were qualitatively similar between the UCL and Oxford dataset. As can been seen in Fig. 4.9A, the noxious-evoked ERPs had similar morphology, polarity and latency in the UCL and Oxford dataset at the different developmental stages. Amplitude was more variable than morphology, especially when considering the first N-P complex between 100–300 ms post-stimulus. Noxious-evoked heart rate response morphology was similar for the youngest and oldest age groups, but qualitatively differed in magnitude between the two datasets between 33–35 weeks (Fig. 4.9B). The proportion of infants showing a facial expression response followed the same trend in the Oxford and the UCL Dataset, with low proportions in the youngest age groups and higher proportions in the middle and late age groups (Fig. 4.9E). In line with this, the relationship between PMA and the magnitude of the noxious features that were tested in part 1 and were available in both datasets (heart rate, PC 1 noxious, PC 2 noxious and PC 3 noxious) had the same direction in both datasets. Taken together, these observations indicated that although subtle differences were present between the datasets, overall trends were reproducible.

Adjusting for PNA, prior pain and sex in the linear models had no effect on the direction of the association between each feature and PMA across all data (Fig. 4.8), although the association between PMA and early delta became weaker after adjusting for PNA, prior pain and sex. PNA had the most impact on the development of the ERSP features, which increased with PNA. Sex and prior pain were of marginal influence on most features, but impacted PC 2 noxious and heart rate, respectively.

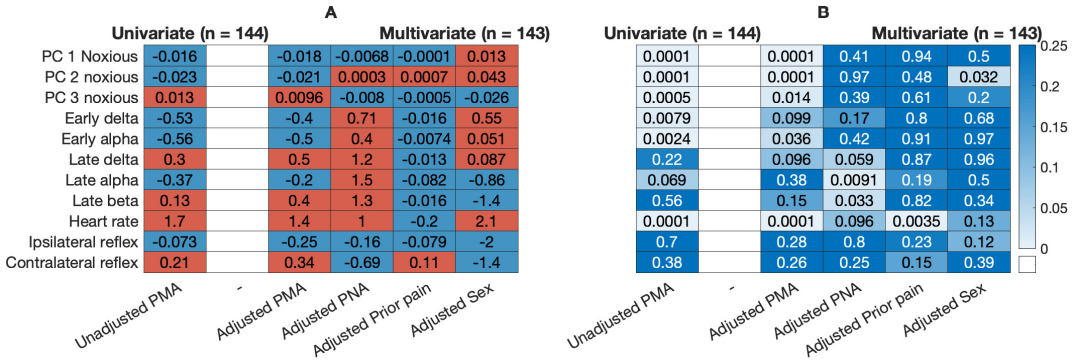


Figure 4.8: Exploration of the influence of clinical factors on postmenstrual age (PMA) – feature relationships. A) Regression coefficients for the relationship between several clinical variables and each feature. Red blocks indicate positive coefficients, blue blocks indicate negative coefficients. B) Uncorrected p-values corresponding to the models in A. The far left columns in A and B indicate the univariate linear models which only include PMA (*unadjusted*). The columns on the right (*adjusted*) indicate the parameters obtained from the multivariate linear model which included PMA, postnatal age (PNA), prior pain and sex as covariates. Regression coefficients and p-values refer to each covariate when adjusting for the other three covariates. As prior pain was missing for one infant, the multivariate models were derived in 143 infants instead of 144. Inclusion of these extra variables did not have an impact on the direction of the regression coefficients for the main relationship between each feature and PMA.

4.4 Discussion

Infant development is characterised by widespread changes across the nervous system. In this chapter, the development of multimodal noxious-evoked responses was characterised in infants aged 28–40 weeks PMA, with a special focus on noxious-evoked ERPs.

4.4.1 Noxious-evoked ERPs change morphology

Noxious-evoked brain responses are characterised by a shift from non-discriminative delta brush activity to noxious-evoked ERPs [89]. In the group of infants studied here, it was confirmed that delta brush responses could be observed in the youngest infants in response to a heel lance, whereas from 34 weeks onwards a positive deflection between 400–700 ms post-stimulus was present. In addition, a previously unidentified intermediary noxious-evoked ERP was found in infants as young as 30–33 weeks PMA. This transitional ERP (captured by PC 2 noxious) consisted of

a negative deflection which peaked around 450 ms post-stimulus, which was not present in response to the control heel lance and decreased in magnitude with age.

The structural correlates and interpretation of these gradual changes in ERP morphology are unknown. One possible interpretation is that each of the three EEG responses reflect a distinct sensory process which uniquely relates to the transitory connections and structures that are present in the preterm infant brain at the time during which it is present. Delta brush responses have been linked to the subplate [339, 236, 11], a transitory structure that underlies the cortical plate during fetal development [183, 180, 127]. During the time period in which PC 2 noxious is present, thalamocortical connections are established through subplate neurons [183, 180] and the cortical plate differentiates into its adult-like configuration with six layers [180]. It is possible that PC 2 noxious represents this early thalamo-cortical signalling through the subplate. The subplate begins to dissolve after direct thalamo-cortical connections are formed [180]. Therefore, the more mature positive ERP (PC 3 noxious) may represent signalling through these direct connections.

PC 2 noxious and PC 3 noxious could also be more closely related. Indeed, when qualitatively observing the smoothed changes in ERPs over time, it appears that the negative deflection (PC 2 noxious) is an early, high-amplitude version of the negative peak of the more mature ERP (PC 3 noxious). These findings fit with previous reports of age-related decreases in sensory ERP amplitude [130, 307, 195] and the notion that early components of tactile ERPs develop before later components [345]. In this context, PC 2 noxious may represent early thalamo-cortical signalling, whereas the positive deflection may be the result of further processing involving inter-hemispheric connections, which form between 35–37 weeks gestation [180]. This would parallel adult laser-evoked potentials, in which the negative (N2) and positive (P2) peaks co-vary with activity in different brain areas [224]. The developmental decrease in PC 2 noxious magnitude could be due to pruning of excess connections in the infant brain [127, 339], which would lead to less neurons being recruited and smaller responses being measured at the scalp [201]. Non-neural anatomical

differences in very young infants, such as fontanelle size, could also contribute to the decrease in amplitude [339].

Another interpretation is that the negative (PC 2 noxious) and positive (PC 3 noxious) waveforms represent the same process but that ERP polarity reverses over the course of early development. The polarity of event-related potentials is generally not considered to be meaningful, because various factors contribute to ERP polarity and it is usually not possible to disentangle them [201]. These factors include for instance the location of the ERP generator with respect to the recording electrode and the polarity of the underlying neurotransmission (excitatory or inhibitory) [201, 177]. Some of these factors may change with development. For instance, the orientation of the cortical neurons with respect to the recording electrode may change as a result of gyrification [325] and a switch occurs in the properties of the neurotransmitter GABA, which changes from excitatory to inhibitory in early development [26, 326]. Yet, the latency difference between PC 2 noxious and PC 3 noxious, and the fact that both negative and a positive deflection can be present in the same infant, suggests that they are complementary instead of one and the same response.

Interestingly, a similar shift from delta brush responses to an early predominantly negative ERP (approximately 150 ms post-stimulus) and finally a slightly later predominantly positive ERP (between approximately 100–300 ms post-stimulus) was seen in the responses evoked by the non-noxious control heel lance, although the predominantly negative ERP appeared less consistently and at a later stage in development compared to PC 2 noxious. This suggests that a common developmental process, which is not specific to nociception, underlies the changes in both noxious and non-noxious ERPs. The apparent timing difference in maturation between responses to noxious and non-noxious stimuli could be due to the vibrotactile receptors developing slower compared to nociceptors [92] (see Section 1.7.3). It is also possible that non-noxious-evoked responses have a different topography compared to noxious-evoked responses [170], and that early developmental stages were therefore not visible at the Cz electrode. The development of tactile-evoked

responses was not investigated in detail here, because the focus was on noxious-evoked responses at the vertex. Nonetheless, in the future, it would be of interest to investigate the development of several sensory responses at multiple electrode sites, to investigate whether a general developmental pattern can be distinguished, or whether these changes are limited to the nociceptive system.

4.4.2 Discriminatory ERPs are present in premature infants

In contrast to delta brush responses, which in preterm infants are not differently evoked by noxious and non-noxious stimuli, PC 2 noxious magnitude was larger for the heel lance compared to the control heel lance in a group of preterm infants below 34 weeks PMA. This suggests that differential neural encoding of noxious and non-noxious stimuli can occur at an earlier stage than originally thought [89], and potentially before the emergence of discriminatory behavioural responses, which has been reported to occur around 33 weeks PMA [120].

In this study, the relationship between PC 2 noxious and PC 3 noxious magnitude and other modalities was weak and not statistically significant, after adjusting for PMA. This suggests that other factors, such as neonatal infection and other illnesses and sleep state are also likely to play a role in the generation of these types of activity. For instance, previous work has shown that the relationship between neural and behavioural noxious-evoked responses is disrupted in infants under stress [167]. Investigation of these factors did not fall within the scope of this project. In the future, it would be of interest to investigate the relationship between PC 2 noxious and other types of activity as has previously been done for the late noxious-evoked ERP in term infants [167, 132, 120].

As was touched upon in Chapter 3, the presence of discriminatory ERPs should not be interpreted as evidence that preterm infants consciously perceive noxious stimuli in the same way as adults, or even that term infants experience stimuli in the same way as preterm infants. In the absence of verbal report, we can only speculate about the subjective experience of pain in infants. However, the findings presented

in this chapter provide evidence for the theory that even in very preterm infants, noxious stimuli are processed in the brain, and thus emphasise the importance of adequate pain assessment and management in infants of all ages. Additionally, this highlights the possibility of using brain-derived measures of pain in clinical trials of analgesics in very preterm infants, analogous to the use of the pre-defined template of noxious-evoked activity in infants aged 34 weeks PMA and over [134, 44, 131].

The fact that PC 2 noxious is transient makes the interpretation of its magnitude more complicated. In studies of the previously derived template of noxious-evoked EEG activity [131], the underlying assumption is that larger responses correspond to more nociceptive processing, and smaller responses correspond to less nociceptive processing. However, as PC 2 noxious magnitude declines with age, small values for PC 2 noxious could be interpreted both as little nociceptive processing and advanced stages of development. PC 2 noxious magnitude should thus always be interpreted within the context of infant PMA. It would be wise to also take into account the magnitude of PC 3 noxious or the template of noxious-evoked EEG activity, and other multimodal responses. Novel methods to estimate infant "brain age" from EEG recordings [248] may prove helpful in interpreting noxious-evoked responses within the context of an infant's individual brain maturity stage.

It is possible that discriminative brain activity arises even before the emergence of PC 2 noxious. PC 1 noxious, which resembled delta brush activity, was not discriminative in the group of infants studied here, but this could be due to the low number of very young infants in the dataset, or because the PC summary measure was not sensitive enough to capture differences between noxious and non-noxious responses. For instance, PC 1 noxious magnitude is largely based on the amplitude of the slow wave. Differences in the frequency or amplitude of the fast oscillatory activity in delta brush responses may exist between noxious and non-noxious events. Tactile-evoked delta brushes have been characterised in very preterm infants [195, 324], and it would be of interest to compare these with noxious-evoked delta brushes. However, this would require a larger group of very preterm infants than was studied here.

4.4.3 Developmental changes in noxious-evoked activity are also present across other modalities

The noxious-evoked ERPs matured concomitantly with responses from other modalities.

Noxious-evoked heart rate responses showed a strong positive association with PMA, which may reflect the increase in autonomic tone which occurs in the first weeks of life [235]. Although heart rate responses are potentially discriminative even in very young infants, this developmental pattern is important to consider when designing or interpreting pain assessment tools. Bespoke cut-off values should be implemented based in infant age, as suggested for the Bernese pain scale by other authors [274].

Two ERSP features (noxious-evoked late alpha and late delta) decreased in magnitude with PMA, whereas the other ERSP features and reflex withdrawal magnitude did not change with age. For the reflexive responses, only the first 1000 ms post-stimulus was considered for inclusion in the classification model because activity in this time period was not significantly different between term and preterm infants in previous work [53]. This may explain the apparent discrepancy with previous studies that utilised a longer recording duration and reported a decrease in limb withdrawal activity with age [132, 53]. The decrease in noxious-evoked late delta activity and late alpha activity with age may represent a decrease in noxious-evoked delta brush activity, which consists of delta waves with overriding oscillations in the alpha frequency range [346]. It may mark the start of a shift towards a more adult-like noxious-evoked spectral response to pain, in which noxious-evoked activity is predominantly found in the gamma range and alpha activity is reduced [90]. Further investigations of noxious-evoked spectral changes of infants at increasing postnatal ages should be undertaken in the future. The ERSP and limb withdrawal features were chosen to maximise discrimination in the development of the classification model in the previous chapter and are not necessarily the best physiological representation of the modalities under investigation. Other aspects of the reflexive response, such as duration or amplitude [133], and the noxious-evoked

EEG spectrum, may have different developmental trajectories. These were not considered here, as the focus was on the features included in the classification model in Chapter 3 and the development of noxious-evoked ERPs.

4.4.4 Developmental trajectories are reproducible across datasets

Large sample sizes are rare in infant EEG research because of the challenges associated with data collection in this group. In this study a pooled dataset from two different sites was used so that a reasonable sample size (n = 144) could be achieved. Exploratory analyses showed that the main findings were reproducible across the two datasets. Qualitatively, differences were observed in the amplitude of the noxious-evoked ERPs. These may arise through differences in recording equipment or stimulation methods. For instance, electrode impedances may be different between different sites and the sound which is generated by a heel lance device may vary from brand to brand. This cross-site variation is important to consider when implementing EEG-derived measures in practice. As an increasing number of centres across the world have embarked on the use of EEG in neonatal pain research [27, 205, 173, 169], replication studies should be carried out to assess this cross-site variability. Further quantitative investigations into the reproducibility between noxious-evoked EEG recordings at different research sites are already underway [13].

4.4.5 Limitations and future work

As these results are from a cross-sectional observational design, it is not possible to make causal inferences on the role of PMA in the observed noxious-evoked responses. It is known that age trajectories derived from cross-sectional developmental studies can differ from those derived from longitudinal studies [9]. Yet, the cross-sectional design was implemented so that a relatively large sample size was available, which allowed a thorough investigation of noxious-evoked responses. It would have been infeasible to record noxious-evoked activity at multiple test occasions in such a large group of infants and therefore small effects might have been missed.

The current study focused on PMA, which is the sum of gestational age (time spent in utero) and postnatal age (time spent ex utero). Some aspects of development may be attributable more specifically to GA or PNA than to PMA overall, or to experiences in postnatal life, such as prior exposure to painful procedures. To explore whether postnatal life experiences affected the findings in this study, exploratory analyses were conducted in which PNA and prior pain exposure were added as covariates to the linear models investigating the relationship between PMA and multi-modal noxious-evoked responses. This had minimal impact on the main association between PMA and the individual features, which suggests that PMA was an informative variable in this cross-sectional analysis.

The impact of PMA, postnatal experiences and GA, should further be investigated in a longitudinal study with several cohorts grouped by GA and followed up at regular intervals. This would allow the comparison of the cross-sectional trajectories with the development within infant. It would also be of interest to compare the trajectories of infants born at increasing GAs, as previous studies have found a relationship between GA and accelerated development of brain-derived responses to sensory stimuli [204, 275]. Nonetheless, one must bear in mind that even in a longitudinal study with multiple cohorts, it is not straightforward to define "normal" brain development, because infants born prematurely are by definition not typically developing fetuses and have altered functional brain connectivity compared to fetuses in utero [60].

Sex and a variety of other factors such as length of hospitalisation, body weight and the presence of co-morbidities, may also play a role in the shaping of noxious-evoked responses. The impact of sex on the primary PMA association was explored in this study and no strong effects were found, except for a positive association between male sex and higher PC 2 noxious responses, when adjusting for PMA, PNA and prior pain. These preliminary findings should be further investigated in a follow-up study, as a previous study did not find any sex-related differences in noxious-evoked ERPs at the vertex [331]. Other factors were not individually

investigated here, as many of these are are linked to PMA and PNA. However, it would be of interest to investigate these factors further in longitudinal studies.

As noxious-evoked activity in term infants is largest at the vertex [290], only the Cz electrode site was considered in this work. It is possible that additional lateralised responses are present at other electrode sites. For instance, spontaneous delta brushes often occur in temporal areas [11]. Future work should thus investigate whether ERP and ERSP responses can be distinguished at other electrode sites. In addition, an important next step is the investigation of the spatial extent of noxious-evoked brain activity using techniques such as BOLD fMRI. In term infants, noxious-evoked BOLD activity is remarkably similar to that of adults [116, 80], but it is currently unclear how it develops or how it relates to noxious-evoked ERPs. Enhanced knowledge of the brain areas which are involved in generating noxious-evoked ERPs can further shape our understanding of infants' noxious-related experiences.

4.4.6 Summary and next steps

In this chapter I used data from 144 infants across two datasets to demonstrate that multimodal noxious-evoked responses gradually change in early life. Importantly, a previously unreported discriminatory noxious-evoked ERP was identified in infants as young as 30–33 weeks PMA, which suggests that the neonatal brain differentially processes noxious and non-noxious events even at this stage of life. The results in this chapter underline the need for developmentally-sensitive pain assessment and management in neonates of all ages, and provide a first step towards the implementation of brain-derived measures of pain in very preterm infants. Before brain-derived measures of pain can be implemented in very young infants, further neurophysiological understanding of noxious-evoked ERPs is necessary. Therefore, the next chapter investigates the neuroscientific validity of the previously derived template of noxious-evoked EEG activity.

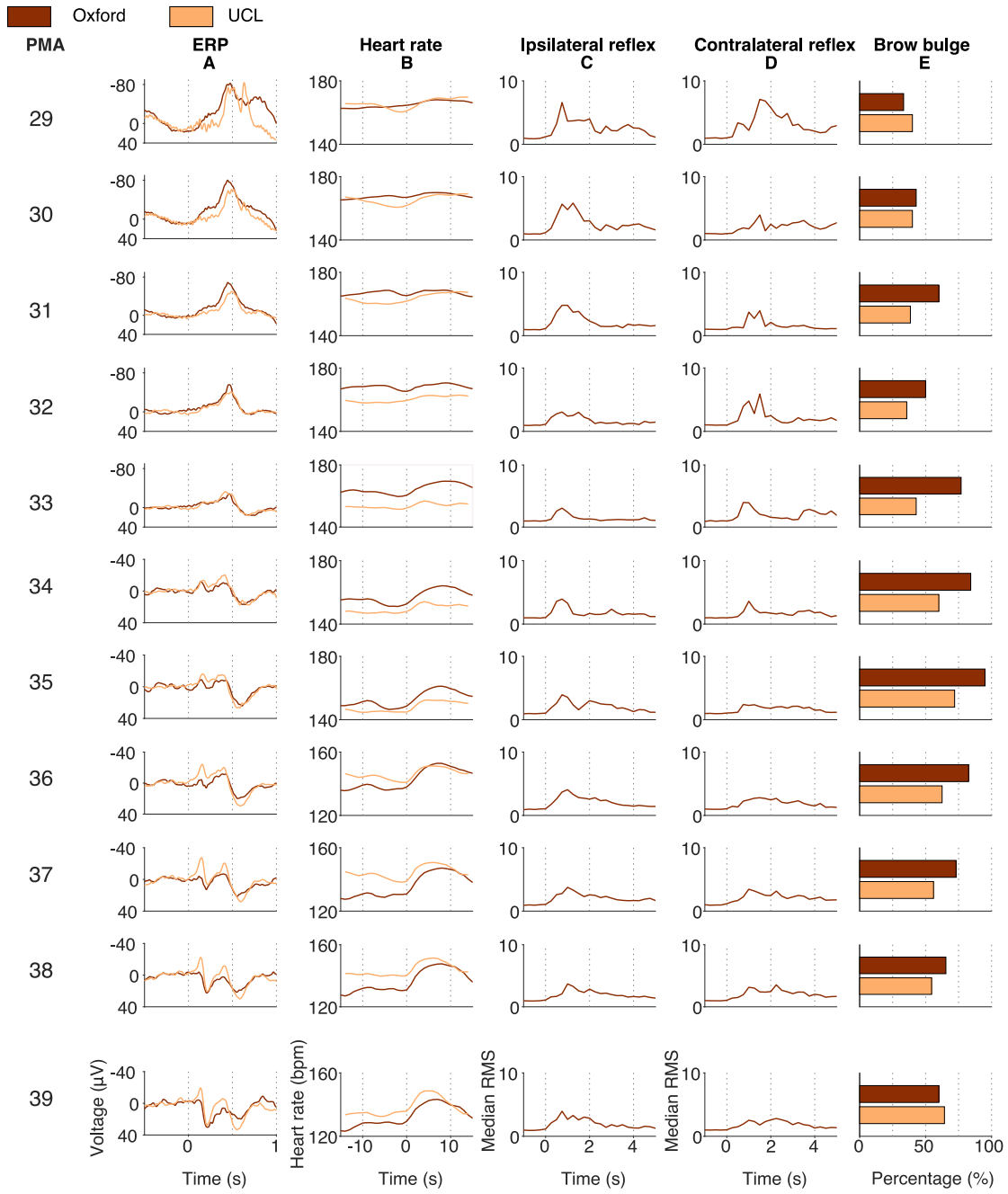


Figure 4.9: Overview of the development of multimodal noxious-evoked responses. Smoothed averages for the noxious-evoked A) event-related potentials (ERP) at Cz, (B) heart rate, (C) ipsilateral reflex, (D) contralateral reflex and (E) brow bulge responses in neonates from 28–40 weeks postmenstrual age (PMA), split by PMA. Column A) ERP morphology changed from a high-amplitude slow wave at 29 weeks PMA to a negative peak around 450 ms at 30–33 weeks PMA, and finally to a positive waveform around 600 ms from 33 weeks onwards. From 34 weeks onwards, an early ERP appeared around 100–300 ms post-stimulus, which was also present in the non-noxious responses (Fig. 4.10). Column B) Resting heart rate decreased with age, while noxious-evoked responses increased with PMA. Columns C-D) Noxious-evoked reflex activity in the first second post-stimulus did not significantly change across development. Column E) The proportion of infants displaying a brow bulge response increased with PMA. This figure is my own work and was reproduced with permission from [321] under the terms of a CC-BY license.

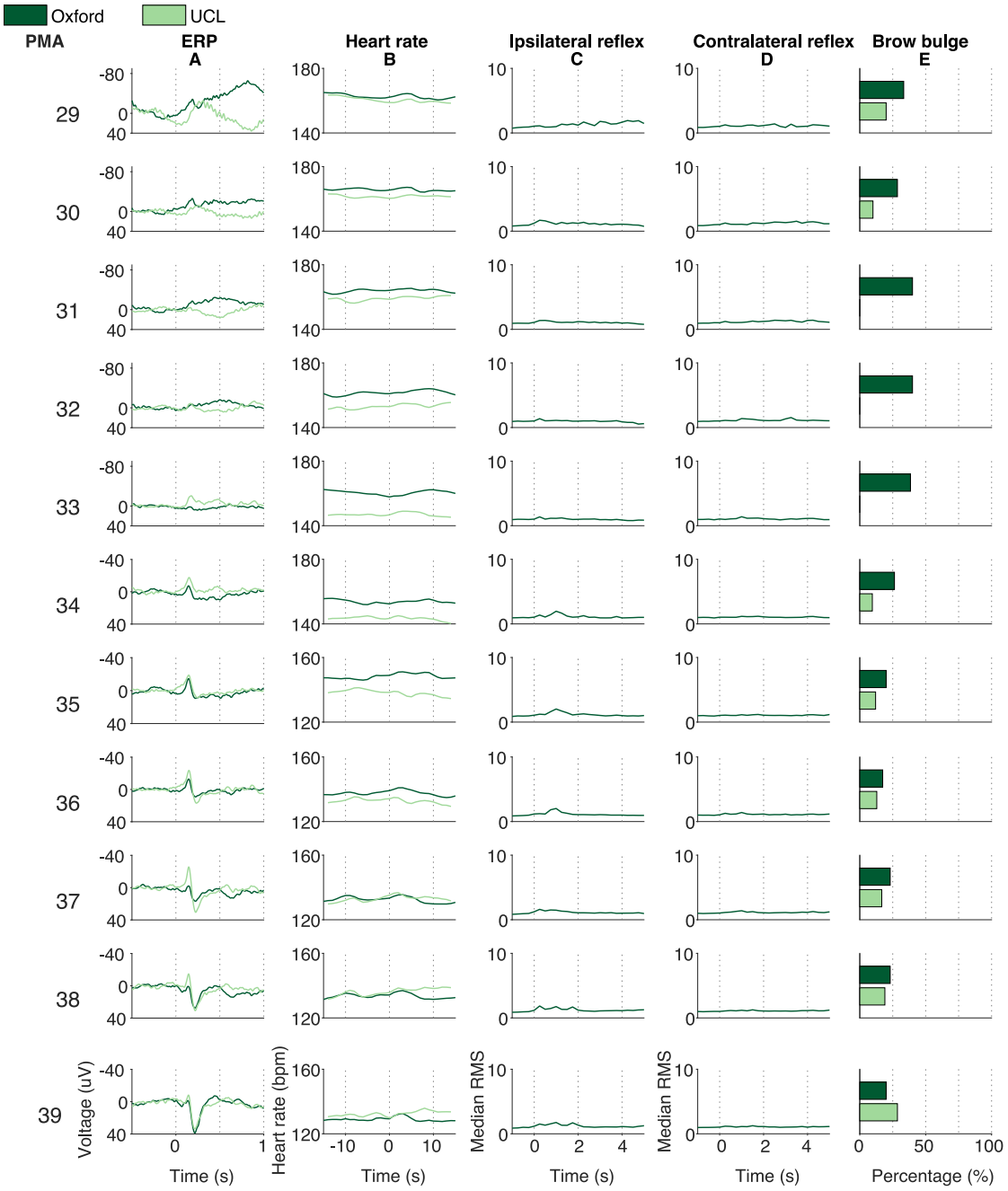


Figure 4.10: Overview of the development of multimodal responses evoked by a non-noxious stimulus. Smoothed averages for non-noxious-evoked A) event-related potentials (ERP) at Cz, (B) heart rate, (C) ipsilateral reflex, (D) contralateral reflex and (E) brow bulge responses in neonates from 28–40 weeks postmenstrual age (PMA), split by PMA. Column A) ERP morphology changed from a slow waveform at 29 weeks PMA, to a negative waveform between 100–300 ms, which was also present in the noxious-responses, from 33 weeks onwards (Fig. 4.9). Finally, from 36 weeks onwards, a positive deflection was visible between 100–300 ms. Column B) Heart rate responses. Columns C-D) Reflex activity. Column E) The proportion of infants displaying a brow bulge response at increasing PMA. This figure is my own work and was reproduced with permission from the Supplementary Materials published with [321] under the terms of a CC-BY license.

Part III

Investigating spatial correlates of a measure of noxious-evoked EEG activity

5 Introduction

In Chapter 3 and Chapter 4, I investigated multimodal neonatal responses to a noxious clinical procedure. Noxious-evoked neural responses were present from early stages in prematurity and were qualitatively similar between two datasets collected at different sites. These findings are an encouraging step forward in the implementation of EEG-derived metrics for infant pain assessment. In this final chapter I present an exploratory pilot study that aims to improve the neuroscientific interpretation of a measure of noxious-evoked EEG activity in term infants.

5.1 A template of noxious-evoked EEG activity

This chapter focuses on the previously derived template of noxious-evoked EEG activity [131], which captures the magnitude of a noxious-evoked ERP between 400–700 ms after a noxious stimulus and was developed for use in infants aged 34 weeks PMA and over. I chose to investigate this template over the noxious-evoked PCs identified in the previous chapters, because the behaviour of this template has been characterised under various clinical conditions and therefore already has a higher degree of validity, reliability and responsiveness compared to the EEG-derived measures first explored in this thesis. For instance, the fact that EEG template magnitude is larger in response to noxious heel lances [131, 125, 320], immunisations

[46] and experimental noxious stimuli [46], compared to non-noxious salient stimuli [131] is indicative of this template's construct validity. Its responsiveness was demonstrated in a study investigating its modulation by gentle touch [125] and a plan exists to test its generalisability to new sites [13].

5.2 Neuroscientific validity

Surrogate measures of pain, such as the template of noxious-evoked EEG activity, are a necessity when assessing analgesic efficacy in non-verbal infants. However, in recent years, interest has grown for the development of brain-derived measures of pain even in instances were verbal report is available [309]. The use of complementary neuroimaging biomarkers in clinical trials could improve mechanistic understanding of the processes that drive particular types of pain, and could be an opportunity to assess the potential of a new analgesic to modulate pain before clinical changes manifest [341]. The use of biomarkers in such a way requires — besides evidence on validity, reliability and responsiveness — an understanding of their physiological basis and qualitative meaning.

For neuroimaging biomarkers, the importance of *neuroscientific validity* has therefore been highlighted [351]. A brain-derived surrogate measure of pain has neuroscientific validity, when it is deemed physiologically plausible [351]. The measure should be embedded within existing knowledge about the neural basis of pain and supported by evidence from a wide range of studies involving complementary research techniques (e.g. fMRI, EEG and single-cell recordings) and populations (e.g. animals and humans) [59]. This concept relates to COSMIN's content validity (the degree to which an instrument reflects the construct of interest [227, 66]) and interpretability (the degree to which an instrument's scores can be qualitatively interpreted [227, 63]). In this part of my thesis I will focus on improving the neuroscientific validity of the infant template of noxious-evoked EEG activity [131].

5.3 Understanding noxious-evoked EEG activity

As described in Chapter 1, it is uncertain what aspect of pain or nociception the template of noxious-evoked EEG activity [131] captures. As a result, it is not always clear how to interpret changes in EEG template magnitude brought about by analgesics. For instance, if the EEG template magnitude were a reflection of nociceptive input to sensory areas, it would be particularly useful in investigations of interventions that are designed to decrease nociceptive signalling, such as local anaesthetics [303], and in situations where the main goal is to limit nociceptive input to the brain. In situations where the goal is to decrease the construct of pain, as defined in Chapter 1, it would have to be used in conjunction with other modalities. On the other hand, if the EEG template magnitude were to reflect the full range of sensory, emotional, autonomic and motor responses to pain, any drug modulating its magnitude would be a highly promising candidate for clinical care.

As described in Chapter 1, several studies have attempted to understand how to interpret noxious-evoked ERPs in adults. Mediation analysis linking contact heat-evoked potentials to perceptive, autonomous and motor responses showed that the negative and positive components of the adult noxious-evoked ERP are differentially involved in each aspect of the overall pain response [304]. Other studies have linked noxious-evoked ERPs to stimulus intensity rather than pain perception [233]. However, as study designs like these necessitate a gold standard measure for pain, they do not easily translate to neonatal pain research.

5.4 Spatial organisation of noxious-evoked EEG activity

Another approach to understanding the template of noxious-evoked EEG activity is to investigate its anatomical source, and to use existing knowledge about this brain region's role in nociceptive processing to understand which aspects of the pain experience the EEG response is likely to represent (reverse inference, see Chapter 1). In adults, numerous EEG studies have reported that the anatomical

origin of laser- and contact-heat-evoked potentials is located in the ACC and the supra-sylvian region, including the parietal operculum and insula [106], and more recent EEG-fMRI studies corroborate these findings by showing correlations between laser- or contact-heat-evoked BOLD activity and the strength of EEG generator activity in these areas [225, 43]. As the ACC and parietal operculum interact with a network of brain regions in nociceptive processing, it is not surprising that simultaneous EEG-BOLD measurements additionally identified a range of other brain areas in which within-subject trial-to-trial BOLD activity is related to variation in noxious-evoked EEG potentials. Besides the ACC and the parietal operculum, these include the cerebellum, bilateral insula, midcingulate cortex, bilateral secondary somatosensory cortex, bilateral primary somatosensory cortex, supplementary motor area (SMA), precuneus [210], amygdala, suprarmarginal gyrus and bilateral thalamus [224]. While these areas may not be the direct anatomical source of the noxious-evoked EEG potentials, they may contribute indirectly to the EEG signal or may be activated by a common pathway.

While simultaneous EEG-fMRI recordings have been achieved in infants [11] and MRI-safe EEG electrodes exist [323], they are challenging to conduct in terms of safety, practicality and data processing. For instance, the equipment used to record EEG and time-lock noxious procedures to the recording in the previous chapters is not MRI safe and clinical procedures cannot be carried out in the scanner. Additionally, simultaneous EEG-fMRI induces artefacts in both recording modalities [1]. EEG recordings are contaminated by pulse and gradient artefacts, and by vibrations inside the scanner environment [1]. In turn, the fMRI recordings will be affected by the presence of EEG equipment in the scanner bore [1]. As infant neuroimaging is already more prone to movement artefacts than adult data, it is questionable whether this would be helpful, or whether it would only introduce more uncertainty about the nature of the recorded responses. Therefore, as a first step, relationships between noxious-evoked EEG template magnitude and noxious-evoked fMRI responses should be recorded in separate sessions.

This implies that it will not be possible to assess within-individual trial-to-trial variability in EEG template magnitude, but that it will be possible to investigate between-individual variability. Neonatal clinical trials are likely to involve between-subject designs rather than within-subject designs [46], as infants are likely to only require the procedure of interest once during the study period. Therefore understanding how between-subject variability in EEG template magnitude relates to spatial patterns of neuronal activity will be of great interest.

5.5 Inter-individual variability in noxious-evoked brain activity

Investigation of between-subject variability relies on the assumption that, when studied under the same conditions, some infants inherently have a higher noxious-evoked EEG template magnitude than others. Natural variability in pain sensitivity (as measured by subjective reports) exists amongst adults and is related to variation in EEG and BOLD responses [48, 277]. While part of this variation may be explained by effects that can rapidly fluctuate over time, such as attention, increasing evidence suggests that another part is more stable and unique to the individual. For instance, resting-state BOLD activity is related to an individual's response to a noxious stimulus both within [210, 293, 237] and across [311] sessions. Grey matter volumes in the parietal operculum [237], as well as grey matter density in the posterior cingulate cortex, precuneus, intraparietal sulcus and inferior parietal lobule [84], and cortical thickness in the primary somatosensory cortex [87] and morphological connectivity [361] are correlated with behavioural pain sensitivity. White matter micro-structure in the medial forebrain bundle [108] and cingulum [342] is related to clinical pain scores. A study in adults reported that noxious-evoked EEG responses were correlated with BOLD activity in the parietal operculum across two separate sessions [128]. This indicates that noxious-evoked responses are not only stable across time, but also across modalities.

Inter-individual variation in infant neural responses has been less extensively investigated, however recent work suggests that term infants display varying degrees

of noxious-evoked EEG and BOLD activity, and that these responses are at least partly predictable within modality. For example, infants' EEG responses to a heel lance are correlated to their response to mild experimental noxious stimuli [46]. Noxious-evoked BOLD responses to mild experimental noxious stimuli are predictable from resting state BOLD activity [21] and relate to functional connectivity in descending pain modulatory pathways [115] and white matter integrity in somatosensory pathways [21]. It therefore seems reasonable to assume that inherent differences in pain processing exist between infants, and that these differences can be exploited to investigate how inter-individual variability in infant noxious-evoked EEG template magnitude relates to spatial patterns of neural activity as measured with BOLD fMRI.

5.6 Aims and study overview

In this part of my thesis, I present the results of a pilot study, which will inform a study to investigate the fMRI correlates of inter-individual variability in, and spatial organisation of, noxious-evoked EEG template responses in infants. The goals of the pilot study were to

1. Evaluate the feasibility of (non-simultaneously) recording noxious-evoked EEG and fMRI in the same infant, and

2. Explore analytical approaches to find correlations between noxious-evoked EEG template magnitudes and noxious-evoked fMRI responses.

To do so, noxious-evoked EEG and fMRI data were recorded in seven healthy term infants in two separate sessions. The EEG template magnitude evoked by a heel lance was then related to noxious-evoked fMRI responses in the same infant using three different approaches. First, EEG template magnitude was correlated to mean fMRI activity in a spatial pattern that is related to pain in adults. Second, EEG template magnitude was correlated with fMRI activity in regions of interest (ROI) that co-varied with noxious-evoked EEG potentials in a parallel adult study. These ROIs were generated in a parallel adult study, because it was expected that

SNR would be higher in adults than in infants. This adult-infant comparison was deemed appropriate because adults and infants display noxious-evoked potentials [90] and because noxious-evoked fMRI patterns in infants are similar to those in adults [116, 80]. Third, correlations between EEG template magnitude and BOLD responses were investigated in a voxelwise manner.

Chapter 6 outlines the adult study which was conducted to derive ROIs for the infant study. Chapter 7 details the main infant study. The main findings from the pilot study and recommendations for a larger-scale follow-up study will be discussed in Chapter 7, Section 7.3. Whilst the initial aim was to conduct a larger-scale infant and adult study, this was not achievable due to a long-term pause in infant MRI studies because of the COVID-19 pandemic between March 2020 and December 2021. The first infant MRI scan in our research group after this pause took place only in April 2022.

Graphical summaries of the analysis steps taken in Chapter 6 and Chapter 7 can be found in Fig. 5.1 and Fig. 5.2.

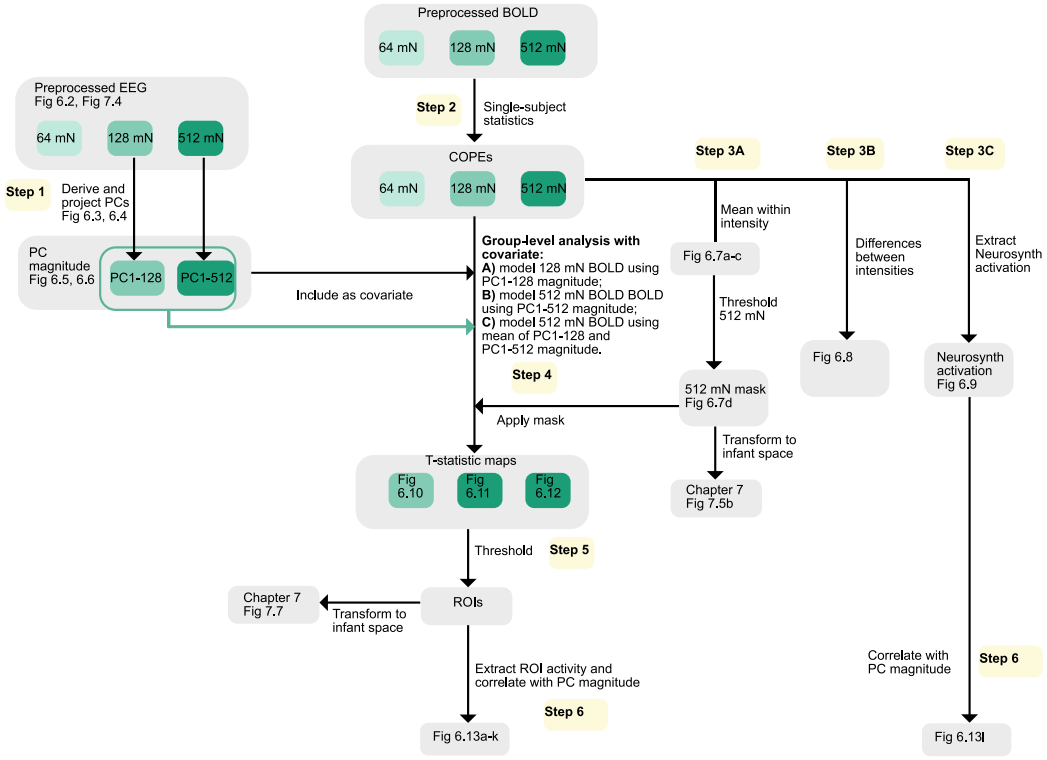


Figure 5.1: Graphical representation of the analysis steps and figures in Chapter 6. Step 1: PCA was applied to the preprocessed EEG 128 mN and 512 mN data at Cz (Fig. 6.2) to obtain PC magnitudes for each participant (Fig. 6.3, Fig. 6.4, Fig. 6.5, Fig. 6.6). Step 2: For each participant a contrast of parameter estimates (COPE) image was extracted from the preprocessed fMRI data which shows which voxels are activated by the pinprick stimuli. Step 3A: The COPEs were fed into a group analysis to extract the mean activation across participants (Fig. 6.7). The 512 mN mean activation map was thresholded to create a mask. Step 3B: Differences in BOLD activity between the 64, 128 and 512 mN were investigated (Fig. 6.8). Step 3C: Activity in the Neurosynth signature of pain was extracted for each COPE (Fig. 6.9). Step 4: To investigate EEG-BOLD relationships, group level analyses were carried out using the EEG PCs as covariates, using the mask derived in step 3A. Step 5: The group-level t-statistic map were thresholded to delineate regions-of-interest (ROIs, Fig. 6.10, Fig. 6.11, Fig. 6.12). Step 6: BOLD activity in the ROIs and the neurosynth signature was correlated with the EEG PCs Fig. 6.13). Step 7: the ROIs and 512 mN mask were taken forward to Chapter 7).

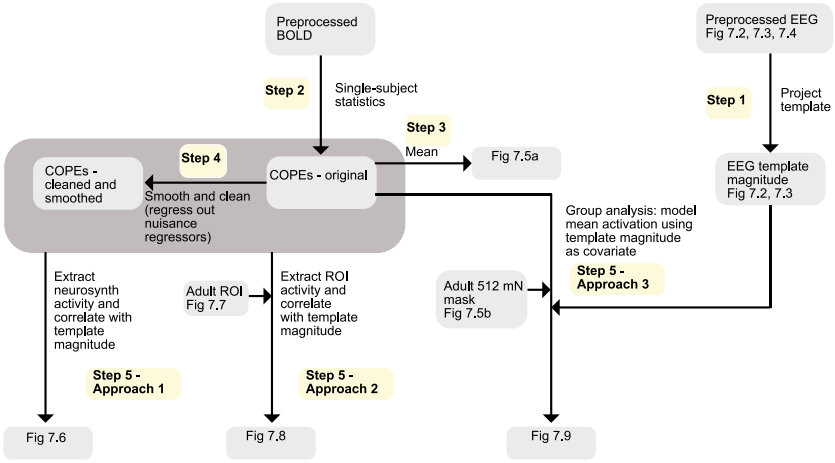


Figure 5.2: Graphical representation of the analysis steps and figures in Chapter 7. Step 1: the template of noxious-evoked EEG activity was projected to the preprocessed EEG data at Cz to obtain EEG template magnitudes (Fig. 7.2, Fig. 7.3). Step 2: for each infant a contrast of parameter estimates (COPE) image was extracted from the preprocessed fMRI data which shows which voxels are activated by the noxious stimuli. Step 3: The COPEs were fed into a group analysis to extract the mean activation across infants Fig. 7.5a. Step 4: To explore the effects of extra cleaning steps, nuisance regressors were regressed from individual COPEs and smoothing was applied. Step 5: Three approaches were used to investigate EEG-BOLD relationships. Approach 1: Neurosynth activity was extracted from both the original and the cleaned and smoothed COPEs and correlated with EEG template magnitude Fig. 7.6). Approach 2: BOLD activity was extracted for each of the adult ROIs for both the original and and the cleaned and smoothed COPEs (Fig. 7.7) and correlated with EEG template magnitude (Fig. 7.8). Approach 3: a voxelwise analysis was carried out by using EEG template magnitude as a covariate in a group level analysis (Fig. 7.9). The 512 mN mean activation was used as a mask (Fig. 7.5b).

6

Noxious-evoked EEG-BOLD relationships in adults: derivation of ROIs

This chapter describes a small scale EEG-fMRI pilot study in adults. The goal of the study was to identify ROIs in which noxious-evoked BOLD activity may correlate with noxious-evoked EEG potentials. These ROIs could then be carried forward to the infant study (Chapter 7).

6.1 Methods

Research governance and technical study details are described in Chapter 2, Section 2.2.3.

6.1.1 Participants and design

This study included six healthy adults aged over 18 years of age. Five adults underwent both a complete EEG and MRI session which were 16–32 days apart, one adult only took part in the EEG session. Two additional healthy adults were recruited for studies to pilot the MRI sequences and practice the EEG set-up. Data from these two initial studies were not included in the analyses as they did not utilise the final design. Whilst the initial aim was to conduct a larger-scale study,

this was not achievable in the timeline of my DPhil due to a pause in recruitment because of the COVID-19 pandemic.

6.1.2 Experimental procedures

6.1.2.1 EEG and psychophysics

Participants were seated in a comfortable chair with their feet propped up and were asked to relax. EEG data were recorded using a 21 channel EEG system (Synamps, Neuroscan) placed according to the 10-20 system, with a reference at Fz and a ground electrode on the forehead. Pinprick stimuli were applied to the dorsum of the left foot using punctate probes that deliver a constant force (PinPrick, MRC GmbH, see Chapter 2, Section 2.2.1). The stimuli were applied in two blocks, which each contained three sub-blocks. Each sub-block contained stimuli of only one intensity (64, 128 or 512 mN). The order of the sub-blocks was determined by randomisation and the same order was applied in both the first and the second block within participant. During each sub-block a train of approximately 10 stimuli was applied with an inter-stimulus interval of at least 10 seconds to match the long inter-stimulus intervals used in infants. During the the second block (thus when participants had already experienced all three intensities), after each sub-block, participants were asked to rate their pain intensity across the previous sub-block on a numerical rating scale (NRS) of 0–10. They were also asked to circle any descriptors they associated with the current stimulus intensity on a McGill Questionnaire form [214]. During stimulation, participants were asked to keep their eyes open and focus on a point in the distance while a curtain obscured their view of the stimuli. The 64 mN and 512 mN stimuli were time-locked to the EEG using a high-speed camera and the 128 mN stimuli were time-locked with an automated stimulus detection interface.

6.1.2.2 MRI

During the MRI session, participants were asked to lie still in the scanner while pinprick stimuli were delivered to the dorsum of the left foot with an inter-stimulus interval of approximately 25 seconds to match the infant paradigm (Section 2.2.2.3).

Trains of stimuli of 5 different intensities (32, 64, 128, 256 and 512 mN) were applied in five sub-blocks. As with the EEG session, each sub-block only contained stimuli of the same intensity, and the order of the sub-blocks was randomised. The stimulus timings were marked by the investigator with a button press. The MRI acquisition parameters can be found in Section 2.2.3.2. The 32 mN and 256 mN stimuli were not analysed for this thesis.

6.1.3 Pyschophysics analysis

The relationship between numerical pain ratings and stimulus intensity was investigated by modelling each participants' pain scores by using linear regression with stimulus intensity as a predictor. A one-sample two-tailed t-test was conducted to test whether the mean slope across participants was different from 0, with p-values derived non-parametrically in PALM using exhaustive sign flips [350].

6.1.4 EEG preprocessing

The EEG data were preprocessed using EEGlab [68]. For comparison, "background" dummy event labels were added to the EEG 5 seconds prior to each true stimulus. EEG data were filtered between 0.1 Hz and 70 Hz (using a separate high-pass and low-pass filter, values denote lower and upper pass-band edge, respectively) and a notch filter at 50 Hz was applied using Hamming windowed sinc FIR filters implemented in EEGlab [68]. To remove artefacts, 5-second epochs with 2 seconds before the stimulus and 3 seconds after the stimulus were extracted and independent component analysis was applied to the epoched data using the EEGlab function *runica* [68]. Components were automatically labelled using the pre-trained classifier ICLabel [249] in the EEGlab environment. Components and labels were then inspected visually and components that resembled muscle activity, eye blinks or physiological artefacts were regressed out from the data, using the automatically assigned labels as a guide. ICA was performed on the epoched data, because the inter-stimulus intervals were too noisy to obtain a clean ICA decomposition. Individual trials were also visually observed to confirm that no gross movement artefact was

present. None of the epochs were rejected. Data were finally filtered with an 8 or 30 Hz low-pass filter (using Hamming windowed sinc FIR filters implemented in EEGlab [68], values denote upper pass-band edge) for data visualisation and principal component analysis (see Section 6.1.5.1).

6.1.5 Summarising pinprick-evoked potentials (PEP)

To allow the comparison of noxious-evoked EEG and fMRI activity, a summary measure of time-locked pinprick-evoked EEG activity needed to be derived. Adult pinprick-evoked potentials (PEP) are often summarised in terms of their N- and P-peak amplitude and latency [147, 265, 318, 317, 316, 315]. However, as the goal was to compare the adult data to the infant data, an "adult EEG template" summary measure was extracted using PCA, analogous to the derivation of the infant EEG template of noxious-evoked EEG activity [131].

The majority (4/6) of the participants picked "no pain" as a descriptor for the 64 mN pinprick and the average pain score for this intensity was below 1. Therefore, it was deemed that the 128 mN and 512 mN stimulus EEG data were better representatives of noxious stimuli, and the 64 mN EEG data were not used to investigate EEG-BOLD responses and no EEG template was derived for this intensity. The 128 mN was time-locked to the EEG recording using a precise automated event-detection interface, while the 512 mN stimuli were time-locked in a less precise way by using a high-speed camera. Because of the uncertainty in the time-locking precision for the 512 mN EEG data, the 128 mN pinprick stimuli and 512 mN pinprick stimuli were investigated separately.

6.1.5.1 Principal component analysis

Before conducting PCA, a low-pass filter of 30 Hz was applied to the data (see Section 6.1.4) and data were epoched with 500 ms before the stimulus and 1000 ms after the stimulus. The baseline mean was subtracted from each epoch. Within each intensity, topographic plots were created with the EEGlab *STUDY* functionality in conjunction with the EEGlab function *topoplot* [68], to investigate the spatial

localisation of the pinprick-evoked potentials. Data were averaged within intensity across participants at the Cz electrode to qualitatively evaluate the presence and shape of pinprick-evoked potentials at the vertex, where they have been previously identified [147, 265].

Analogous to the analysis used to generate the infant template of noxious-evoked EEG activity [131], PCA was used to extract the main waveforms that explained variance in the adult data at the Cz electrode. The time window of interest was set to 0–500 ms based on previous literature on pinprick-evoked potentials, which report P-peak latencies around 250–350 ms [147, 315, 316].

Participant-level average EEG traces were computed for the 128 mN stimulus epochs and epochs of background data. The number of background epochs was set to be equal to the number of 128 mN stimulus epochs. The within-participant average trials for all participants and both conditions were concatenated into one matrix. The within-participant average trials were iteratively aligned to the cross-trial mean between 0–500 ms post-stimulus using Woody-filtering and PCA was then conducted in the same time window. The resulting principal components were visually inspected and the coefficients were compared across the background and 128 mN stimulus averages using two-tailed paired t-tests with p-values derived non-parametrically in PALM using exhaustive sign-flips [350]. The "adult EEG template" of noxious-evoked activity was then chosen as the first PC that visually reflected the main waveform in the data and that had different coefficients for the background and the 128 mN within-participant averages. This PC was projected back to the individual trials for each participant to obtain the PC magnitude for each trial, allowing each trial to jitter $+-$ 50 ms to best fit the PC. The PC magnitude was averaged across trials within each participant. (Note that for the purpose of clarity in this chapter, "coefficient" is used to describe the PC coefficient obtained for the within-participant average EEG trace. "Magnitude" is used to describe the PC coefficient obtained by projecting the PC to individual trials, after Woody-filtering each individual trial to align with the PC. PC magnitudes can then be averaged within participant.) To investigate whether PC magnitude was

graded with stimulus intensity, the PC was also projected to the 64 mN and 512 mN trials, now allowing the trials to jitter $+-150$ ms, to account for differences in latency between the three stimulus intensities. Intensity grading was investigated by modelling each participant's average PC magnitudes by using linear regression with stimulus intensity as a predictor. A one-sample two-tailed t-test was conducted to investigate whether the mean slope across participants was different from 0, with p-values derived non-parametrically in PALM using exhaustive sign flips [350].

The above procedure was repeated for the 512 mN stimulus with one additional step. As the 512 mN stimuli were not automatically time-locked to the EEG, individual trials were first aligned within participant to account for differences in response latency that were due to variance in time-locking. This was done by iteratively aligning the trials to the within-sample mean, within the background condition and the 512 mN condition separately. These participant averages were then subjected to the procedures described above for the 128 mN averages.

6.1.5.2 Comparison of PCs with amplitude-based and behavioural measures

To investigate whether the "adult EEG templates" (PCs) were a good representation of the PEPs, they were compared with peak-to-peak amplitudes of the N-P complex in the participant-average responses of the same intensity. For the 128 mN stimuli, the raw average was taken for each participant, whereas for the 512 mN stimulus, individual trials were first aligned within participant as described above to account for inadequacies in time-locking. To quantify peak-to-peak amplitudes, trials were low-pass filtered at 8 Hz to remove small deflections that would influence peak amplitude, and local maxima and minima were identified using the MATLAB function *findpeaks*. "N" was defined as the first local minimum after stimulus onset, "P" was defined as the first local maximum that differed at least one 1 μV from surrounding peaks, and "N2" was the first local minimum occurring after P. The peak-to-peak amplitude was calculated between N and P ("N-P") and P and N2 ("P-N2").

Pearson correlation coefficients were calculated to assess the relationship between the "adult EEG template" (PC) magnitude and the peak-to-peak amplitudes and

NRS pain scores with two-tailed p-values derived non-parametrically in PALM [350]. This was done for the 128 mN and the 512 mN stimulus separately, i.e., the magnitude of the "adult EEG template" (PC) derived from the 512 mN data was correlated with the NRS scores for the 512 mN stimulus and the peak-to-peak amplitudes for the 512 mN traces. Correlation analyses were not corrected for multiple comparisons as this was an exploratory pilot study.

6.1.6 fMRI Single-subject analysis

fMRI data were analysed in FMRIB Software Library (FSL) [160, 354] using *FEAT* (version 6.00, [355]). Single-subject preprocessing consisted of motion correction using FSL's *MCFLIRT* [157, 159], highpass temporal filtering (cut-off 100 s) and 5 mm full-width half max (FWHM) spatial smoothing [291]. The functional data were nonlinearly registered using FSL's *FNIRT* [159] to MNI152 standard space in a two-step process that used the subject-specific structural T1 scan as an intermediate registration target. FSL's *MELODIC* [23] was used to conduct independent component analysis and a pre-trained version of *FIX* (*UKBiobank.RData*, provided with FIX) [272, 122] was used to automatically delineate motion, CSF and susceptibility artefacts. All components and the labels assigned by FIX were visually reviewed and labels were adjusted if necessary [122]. Artefactual components were then regressed out of the data using FSL's *regfilt*. For each participant, stimulus timing was convolved with a single gamma haemodynamic response function (HRF) and then fit to the BOLD time series in each voxel with a general linear model (GLM). Fitting was performed separately for each stimulus intensity, as the the different stimuli were applied in separate fMRI runs (not interleaved). The output of the single-subject analysis was a COPE image for each participant and each intensity, which shows the GLM parameter estimate (in other words, the haemodynamic response amplitude for the pinprick stimuli versus baseline) for each voxel.

Neurosynth signature To obtain a summary measure of each participant's noxious-evoked BOLD activity, mean BOLD activity in a spatial pattern commonly associated with pain was extracted. The spatial template was a meta-analysis association map from Neurosynth [358] (key word: pain, `https://www.Neurosynth.org/analyses/terms/pain/`, tresholded at a false discovery rate of p = 0.01). Neurosynth automatically generates meta-analysis maps of z-statistics based on keywords found in research papers. The association map which was used in this study shows in which voxels BOLD activity is preferentially associated with the mention of the word "pain" in study abstracts. The Neurosynth template is thresholded and thus contains voxels with positive z-statistics (voxels that are more commonly activated in papers whose abstracts report the word "pain" compared to papers that do not), voxels with negative z-statistics (voxels that are less commonly activated in papers whose abstracts report the word "pain" compared to papers that do not) and voxels with a value of 0 (voxels that are equally commonly activated in papers whose abstracts report the word "pain" compared to those that do not). The weighted mean BOLD activity in this signature was extracted from the single-subject COPE maps using FSL's *fslmeants* [160, 354], using the z-statistic values in each voxel as weights. All negative z-statistic values were set to have a weight of 0. This weighted mean activity is referred to in the rest of this chapter as "Neurosynth activity". The relationship between Neurosynth activity and stimulus intensity was investigated by modelling each participant's Neurosynth activity by using linear regression with stimulus intensity as a predictor. A one-sample two-tailed t-test was conducted to test whether the mean slope across participants was different from 0 with p-values derived non-parametrically in PALM using exhaustive sign flips [350].

6.1.7 fMRI Group level analysis

The single-subject COPE maps for each intensity were carried forward to the group level analysis. Group level analysis was performed with FSL's *randomise*, which allows non-parametric permutation statistics [350]. Variance smoothing (6

mm FWHM) was applied as it can increase power when the number of participants is low [24].

6.1.7.1 Mean activation

First, the group mean activation was modelled with a group-level GLM analysis within each intensity. This was done 1) to assess whether activations were present that were consistent with previous literature and 2) to derive a mask of noxious-evoked BOLD activity for the BOLD-EEG comparisons. The mask was derived directly from the data instead of using an externally derived mask, because this would allow it to be specific to the nature (mechanical) and location (left foot) of the stimulus. This was the only analysis in which statistical thresholding was used, all other outcomes are presented as unthresholded maps because of the exploratory nature of the work.

Group level maps were statistically thresholded using threshold-free cluster estimation (TFCE) [292]. The advantage of TFCE is that it is less stringent than voxelwise bonferroni family-wise error (FWE) correction, but does not require setting an arbitrary threshold like in cluster analysis [292]. The TFCE-statistic was computed at each voxel and FSL's randomise [350] was used to identify voxels that were significantly activated on average using exhaustive sign-flips. Group maps were thresholded at FWE-corrected $p < 0.05$ for each intensity.

Comparison of BOLD activity at three intensities In order to understand whether BOLD activation was related to stimulus intensity, the three intensities were directly compared using three paired two-sided t-tests (512 mN - 128 mN, 512 mN - 64 mN and 128 mN - 64 mN). The TFCE thresholded activation maps from the 512 mN stimulus were used as a mask for the difference analyses. The 512 mN average activation was pragmatically used as a mask because it spanned a larger area than the 128 mN and 64 mN activation, and thus it would be less likely to miss any areas that were important in the processing of the pinprick stimuli. To aid visual interpretation, clusters with the largest t-statistics were delineated and FWE corrected p-values were calculated for each. The initial cluster-forming threshold

was set at $t = 2.78$, which corresponds to $p = 0.025$ at 4 degrees of freedom (df) ($n = 5$, GLM with 1 regressor) and FWE corrected p-values were non-parametrically derived using FSL's randomise [350] based on cluster mass. The cluster-forming threshold was arbitrary and much lower than the commonly implemented threshold of $z = 3.1$ (corresponding to $p = 0.001$), however these analyses were only used to aid interpretation of the t-statistic maps. The use of permutation testing ensures that the family-wise error rate (FWER) is controlled.

6.1.7.2 PEP-BOLD relationships: identifying ROIs

To identify areas in which the BOLD response positively co-varied with the EEG response, a group-level GLM analysis was conducted using the average "adult EEG template" (PC) magnitude as a covariate (as well as a mean regressor). This was done three times: first, using the magnitudes of the template derived from the 128 mN EEG responses to model the BOLD activity evoked by the 128 mN stimulus; second, using the magnitudes of the template derived from the 512 mN EEG responses to model the BOLD activity evoked by the 512 mN stimulus; and third, using the mean of the 128 mN and 512 mN template magnitude within each participant to model the BOLD activity evoked by the 512 mN stimulus. As the main interest was to identify areas that positively co-varied with the EEG response and were related to noxious processing, the COPEs were masked with the TFCE thresholded mean activation maps from the 512 mN stimulus (see Section 6.1.7.1). Unthresholded t-statistic maps were visually inspected to identify areas that positively co-varied with EEG magnitude.

In addition, to pragmatically delineate the areas in which BOLD activity was most strongly related to the EEG response and that could be used as ROIs in the infant study, clusters of voxels that exceeded a given threshold were extracted. The cluster-forming threshold was arbitrarily set at $t = 3.1$, which corresponds to $p = 0.025$ at 3 *df* ($n = 5$, GLM with 2 regressors). To further understand which clusters were most strongly associated with the PCs, permutation tests in FSL's randomise [350] were used to derive FWE-corrected p-values for each cluster based

on cluster mass. The cluster-forming threshold was arbitrary and much lower than the commonly implemented threshold of $z = 3.1$ (corresponding to $p = 0.001$). However, the goal of the current analysis was to identify any potentially interesting areas that could inform the infant study, as opposed to making statistical inferences. The use of permutation testing ensures that the FWER is controlled.

FSL's *atlasquery* was used to map the clusters to anatomical brain regions, based on the Harvard-Oxford Subcortical Structural Atlas and the Harvard-Oxford Cortical Structural Atlas [207, 103, 71, 103]. *Atlasquery* finds the probability of each voxel in the cluster being part of an area defined in the atlas and then averages this probability within all the voxels in the cluster [158]. Brain regions in grey matter with at least 10% probability are presented to aid interpretation.

Clusters of voxels that showed a positive relationship with the "adult EEG template" (PC) magnitude in the group analysis were considered an ROI. The mean activity in the ROI was extracted for each participant using FSL's *fslmeants* and plotted against the EEG "adult EEG template" (PC) magnitude to further understand the strength of the relationship between the EEG response and the BOLD activity in that ROI. Neurosynth activity for the highest intensity (512 mN) was also plotted against the two "adult EEG template" responses.

6.2 Results

6.2.1 Behavioural responses to pinprick stimuli

Mean pain scores were 0.3/10, 1.8/10 and 3.3/10 for the 64, 128 and 512 mN stimulus, respectively, corresponding to mild pain on the numerical rating scale. Pain ratings significantly increased with stimulus intensity (two-tailed one-sample t-test, n = 6, mean slope = 0.0057, t = 8.90, p = 0.031, Fig. 6.1). All participants chose the words *sharp* and *pricking* to describe the 128 mN and 512 mN stimuli, whereas the 64 mN was only considered *sharp* by four participants and *pricking* by three. Four participants picked *no pain* to describe the 64 mN pinprick, whereas two did so for the 128 mN pinprick and none did so for the 512 mN pinprick.

As the 64 mN pinprick was not considered painful by the majority of participants and was given an average pain score of less than 1 on a scale of 0–10, it was not deemed a good model to investigate noxious-evoked activity in this population. Therefore, only the 128 mN and 512 mN intensity were used to investigate EEG-BOLD relationships. Mean EEG and BOLD responses to the 64 mN pinprick are shown in the remainder of the Results section for completeness.

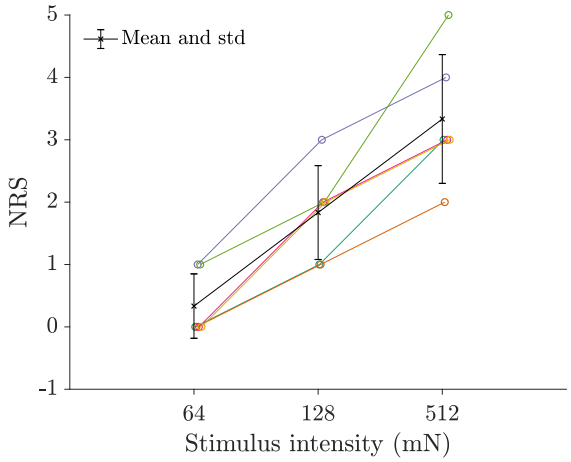
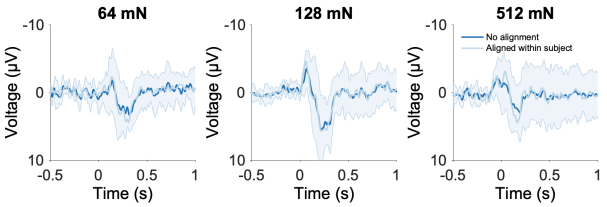


Figure 6.1: Numerical rating scale (NRS) scores (range 0–10) for the 64, 128 and 512 mN pinpricks increased with intensity. Each coloured line represents one participant, black lines represent the mean and standard deviation (std) at each intensity.

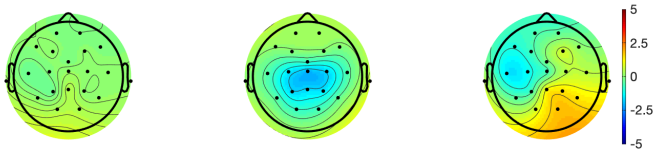
6.2.2 Pinprick-evoked potentials

The pinprick stimuli elicited EEG amplitude changes with a central predominance in the first 500 ms post-stimulus (Fig. 6.2). The average pinprick-evoked potential across participants consisted of a negative-positive complex for all stimulus intensities, as described in previous publications [147, 265, 318] but the latency and amplitude of the response varied between the three intensities (Fig. 6.2a). The average PEP evoked by the 128 mN stimulus had visually larger amplitude than the PEPs evoked by the 64 mN and 512 mN stimuli, and was more consistently visually present in within-participant averages (Fig. 6.5, "Raw") compared to the 64 mN stimuli (not shown) and 512 mN stimuli (Fig. 6.6, "Aligned"). Whilst these differences in amplitude and latency may represent true differences in processing of these different intensity events, it could not be excluded that they were an artefact of

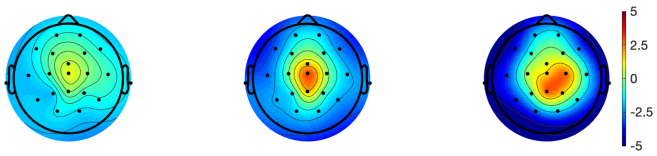
the time-locking procedures used. The 128 mN pinprick stimulus was time-locked to the EEG recording using an automated event-detection interface, while the 64 and 512 mN stimuli were time-locked by using a high-speed camera, which is less reliable, both due to intra-rater variability and due to variation in video quality. Aligning the individual trials within participant prior to averaging across participants increased the amplitude of the 512 mN response, but not the 128 mN response, which suggests less precise time-locking accounted for some of the amplitude difference (Fig. 6.2a). Given this uncertainty in the time-locking precision for the 512 mN EEG data, the 128 mN pinprick stimuli and 512 mN pinprick stimuli were investigated separately.



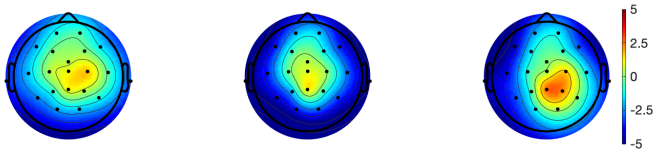
(a) Average responses.



(b) 0–100 ms.



(c) 100–300 ms.



(d) 300–500 ms.

Figure 6.2: Group EEG responses to pinprick stimuli in adults (n = 6). a) Mean time-locked EEG activity to the 64 mN, 128 mN and 512 mN stimuli at the Cz electrode. The pinprick-evoked potentials at the vertex consist of a negative-positive complex within 500 ms post-stimulus. Group-level means are created by averaging participant-level means. Dark blue traces indicate the cross-participant mean when individual trials are not aligned within participant, whereas light blue traces indicate the mean response when individual trials are first aligned within participant. This affects mainly the 512 mN stimulus, which is not automatically time-locked to the EEG. Shaded area represents across-participant standard deviation for the unaligned mean (dark blue traces). b-d) Average topographic plots across 0–100, 100–300 and 300–500 ms post-stimulus showing the central predominance of the pinprick-evoked potentials for all intensities.

6.2.3 Derivation of a summary measure of pinprick-evoked potentials

In order to quantitatively compare the pinprick-evoked potentials with noxious-evoked BOLD activity, a summary measure of the PEPs needed to be derived, analogous to the template of noxious-evoked EEG activity in infants. This was achieved by conducting PCA to extract the main waveforms that explained variance in the 128 mN and background data (Fig. 6.3, panel A). This yielded 12 PCs, of which the first three together accounted for approximately 90% of the variance in the data, with the first PC accounting for 76% of the variance (Fig. 6.3, panel B, C). The shape of the first PC (hereafter termed *PC1-128*) corresponded to the waveform that could visually be observed in the 128 mN EEG response (Fig. 6.3, panel A, C). In addition, the coefficients for PC1-128 were higher for the 128 mN condition compared to the background, whereas this was not the case for PC 2 and PC 3 (n = 6, two-tailed paired t-test, t = 7.62, p = 0.031, Fig. 6.3, panel D-F). When projected back to individual 128 mN epochs, PC1-128 fit the data well for each participant (Fig. 6.5). Mean PC1-128 magnitude was positively correlated with the amplitude of both the positive-negative peak and the negative-positive peak (n = 6; N-P amplitude: Pearson r = 0.81, two-tailed p = 0.05; P-N2 amplitude: Pearson r = 0.82, two-tailed p = 0.047).

This process was repeated for the 512 mN stimulus epochs (Fig. 6.4). Here, the first PC (hereafter termed *PC1-512*), which explained 43% of the variance, was also larger for the stimulus epochs compared to the background epochs (n = 6, two-tailed paired t-test, $t = 12, p = 0.031$) and fit the 512 mN data well when projected back to the data (Fig. 6.6). PC1-512 magnitude was non-significantly positively correlated with N-P amplitude (Pearson r = 0.76, two-tailed p = 0.078), and weakly non-significantly positively correlated with P-N2 amplitude (n = 6, Pearson r = 0.30, two-tailed p = 0.58), suggesting it only moderately captured the PEP waveform.

Numerical pain scores were weakly negatively correlated with PC1-128 magnitude (n = 6, Pearson r = -0.30, two-tailed p = 0.53) and not correlated with PC1-512 magnitude (n = 6, Pearson r = -0.06, two-tailed p = 0.93). Neither PC1-128 nor

PC1-512 was graded with stimulus intensity (PC1-128: mean slope = $-6.4 * 10^{-5}$, two-tailed p = 0.75; PC1-512: mean slope = $-3.7 * 10^{-5}$, two-tailed p = 0.81).

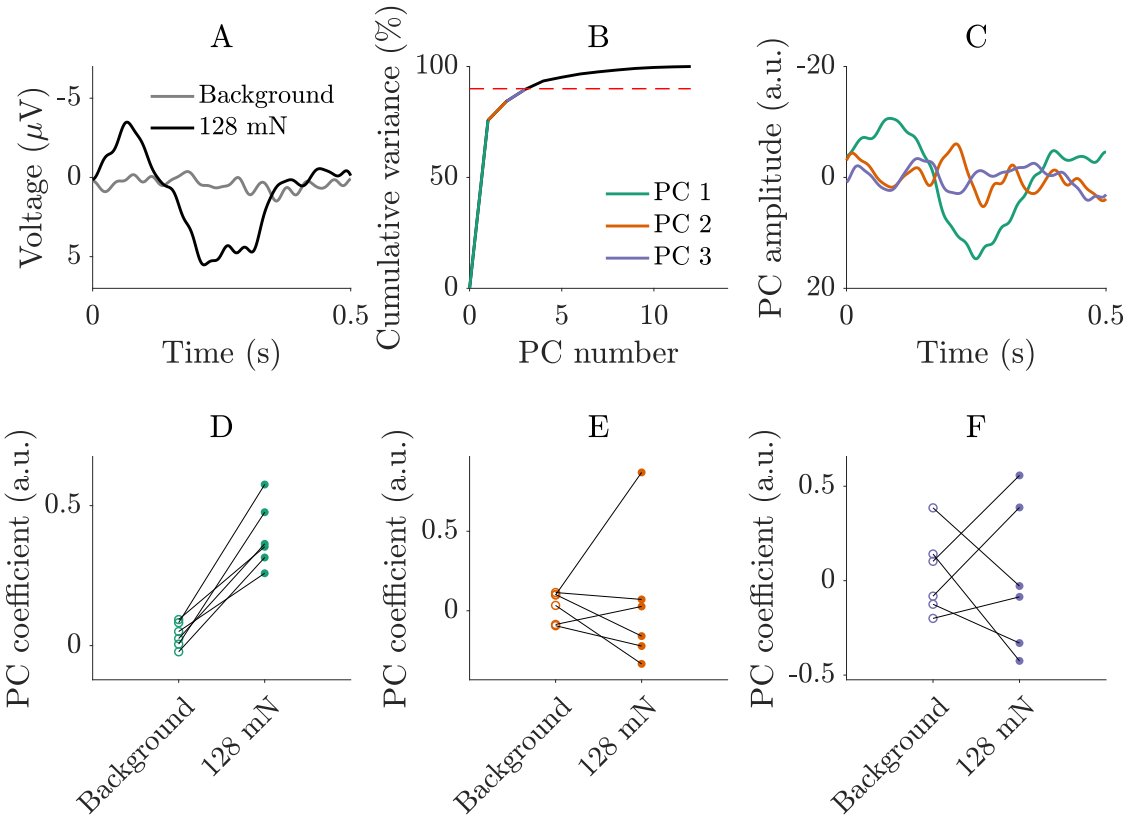


Figure 6.3: Derivation of a summary measure of pinprick-evoked potentials in adults (128 mN). A) Across-participant average of background epochs and responses to the 128 mN stimulus. Principal component analysis was conducted across concatenated participant-average epochs of these conditions to extract the main waveforms that explained variance in the data. B) Cumulative variance explained by the 12 principal components (PCs). The first PC (PC1-128) accounted for over 75% of the variance. Dashed red line indicates 90% cumulative variance. C) The three PCs that cumulatively explained approximately 90% of the variance. The first PC (PC1-128) resembled the raw data. D-F) PC coefficients for the within-participant average background and within-participant average 128 mN data for PC 1, PC 2 and PC 3 respectively. In all participants, the coefficients for the first PC (PC1-128) were higher for the 128 mN data compared to the background data. a.u. = arbitrary units.

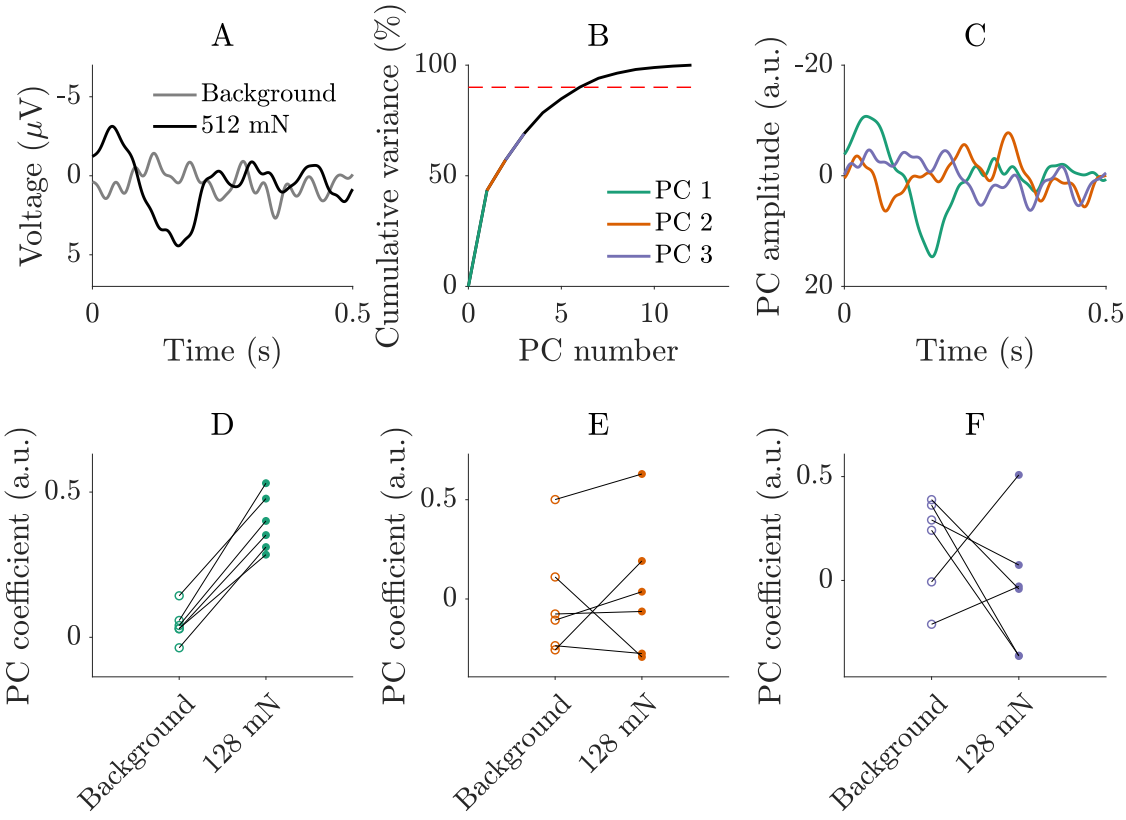


Figure 6.4: Derivation of a summary measure of pinprick-evoked potentials in adults (512 mN). A) Across-participant average of background epochs and responses to the 512 mN stimulus. Principal component analysis was conducted across concatenated participant-average epochs of these conditions to extract the main waveforms that explained variance in the data. B) Cumulative variance explained by the 12 principal components (PCs). The first PC accounted for approximately 43% of the variance. Dashed red line indicates 90% cumulative variance. C) The first three PCs. The first PC (PC1-512) resembled the raw data. D-F) PC coefficients for the within-participant average background and within-participant average 512 mN data for PC 1, PC 2 and PC 3 respectively. In all participants, the coefficients for the first PC (PC1-512) were higher for the 512 mN data compared to the background data. a.u. = arbitrary units.

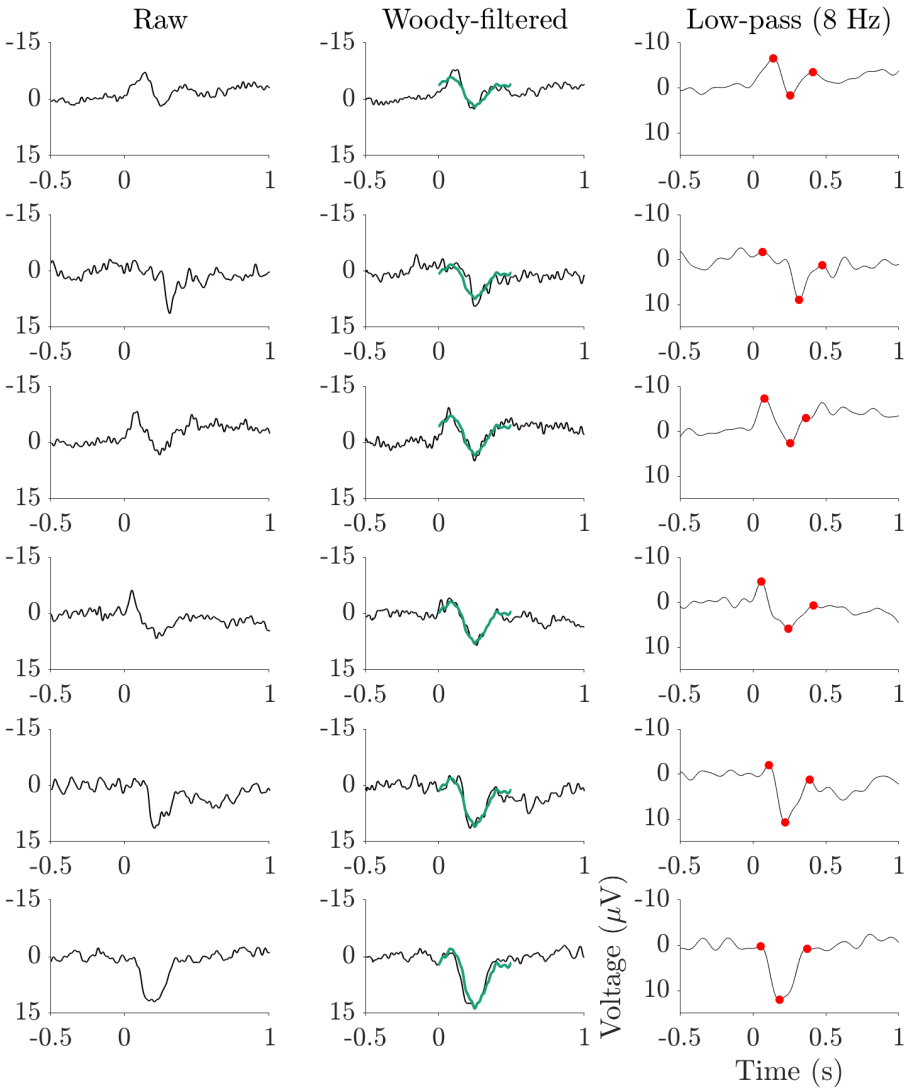


Figure 6.5: Within-participant average pinprick-evoked potentials (PEP) for the 128 mN data sorted by PC1-128 magnitude. Left column ("Raw") shows the raw average responses for each participant with a 30 Hz low-pass filter. Pinprick-evoked potentials are visually present in each participant. Middle column ("Woody-filtered") shows the within-participant average, where each individual trial was Woody-filtered to allow maximum correlation with PC1-128 before averaging. PC1-128 is overlaid in green. Right column ("Low-pass") shows the same averages, now filtered with a low-pass filter at 8 Hz to allow the identification of local maxima and minima. The identified peaks are indicated in red and peak-to-peak amplitudes were taken as the difference in amplitude between the two pairs of adjacent peaks. PC1-128 captures the pinprick-evoked potential identified in the low-pass filtered data.

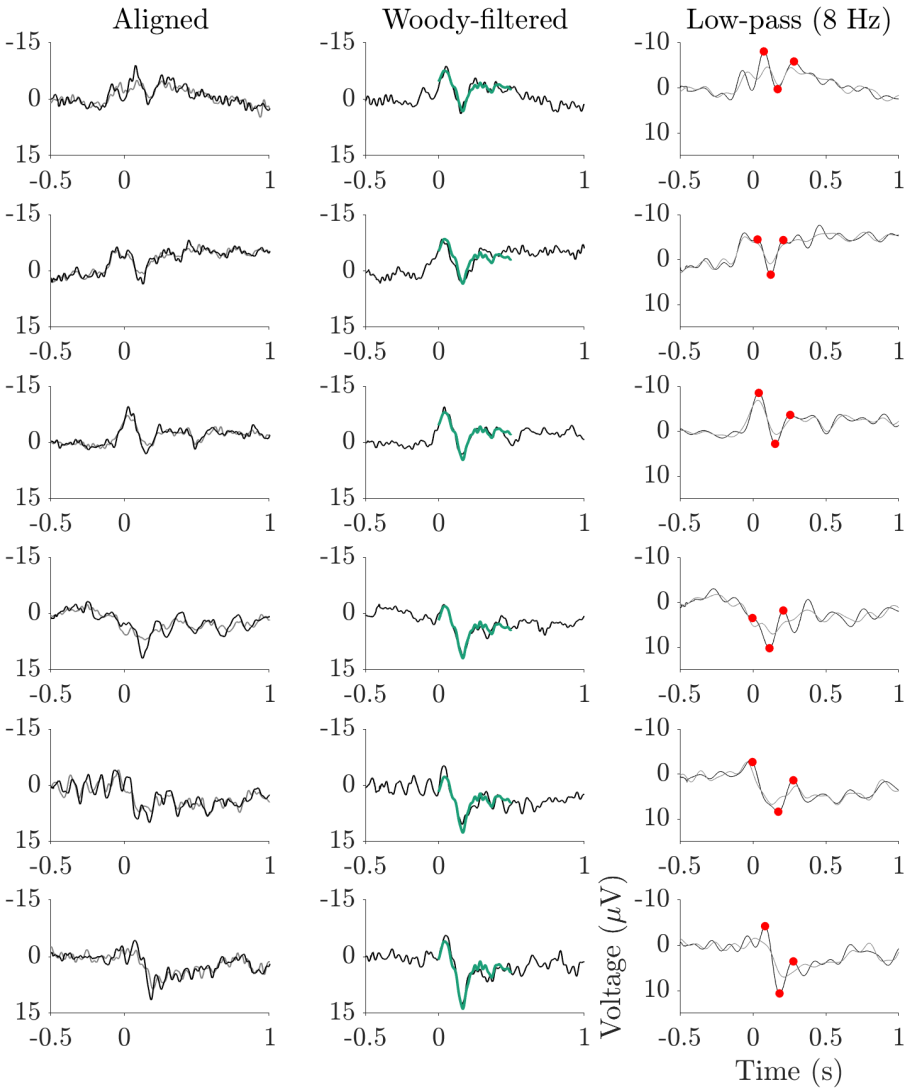
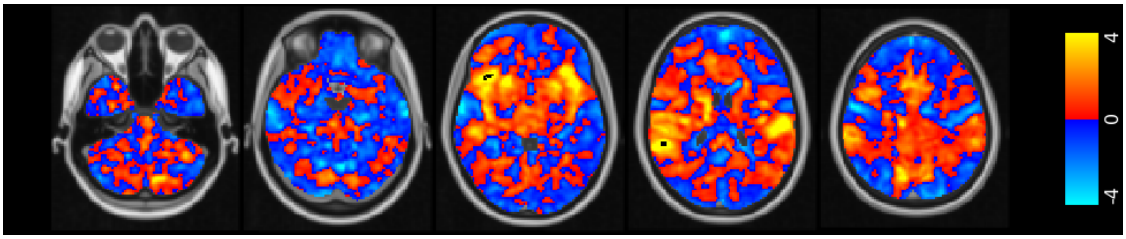


Figure 6.6: Within-participant average pinprick-evoked potentials (PEP) for the 512 mN data sorted by PC1-512 magnitude. Left column ("Raw") shows the within-participant aligned averages (black) and raw averages (unaligned, grey) with a 30 Hz low-pass filter. Middle column ("Woody-filtered") shows the within-participant average, where each individual trial was Woody-filtered to allow maximum correlation with PC1-512 before averaging. PC1-512 is overlaid in green. Right column ("Low-pass") shows the same within-participant aligned averages, filtered with a low-pass filter at 8 Hz to allow the identification of local maxima and minima. Grey lines indicate the raw, unaligned traces. The identified peaks are indicated in red and peak-to-peak amplitudes were taken as the difference in amplitude between the two pairs of adjacent peaks. PC1-512 moderately captures the PEP shape.

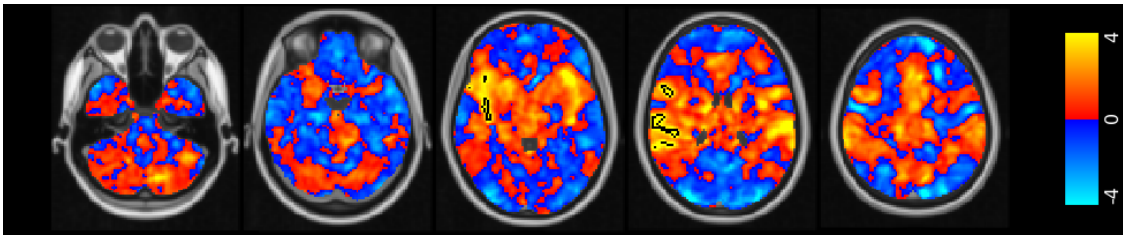
6.2.4 Noxious-evoked BOLD activity

The 64 mN, 128 mN and 512 mN pinprick stimuli were associated with BOLD activity in areas commonly associated with pain, including the bilateral insula, parietal operculum, central operculum, anterior cingulate cortex, post-central gyrus, and thalamus (Fig. 6.7). Activation patterns were similar across the three stimulus intensities, however the extent and magnitude of the activation varied with stimulus intensity and a larger number of voxels survived FWE for the 512 mN stimulus compared to the two lower intensities. Direct comparisons revealed that there was a trend for the 512 mN pinprick to evoke higher activity compared to the 128 and 64 mN stimuli, especially in the ACC and the opercular and insular regions (Fig. 6.8). None of the voxels in these comparisons survived FWE correction, but several sub-threshold clusters of activity ($p < 0.1$) were present. In line with these findings, there was a non-significant trend for Neurosynth activity to increase with intensity (Fig. 6.9, two-tailed one-sample t-test $n = 5$, mean slope $= 0.40, t = 4.13, p = 0.063$). Inter-individual variation was present at each intensity.

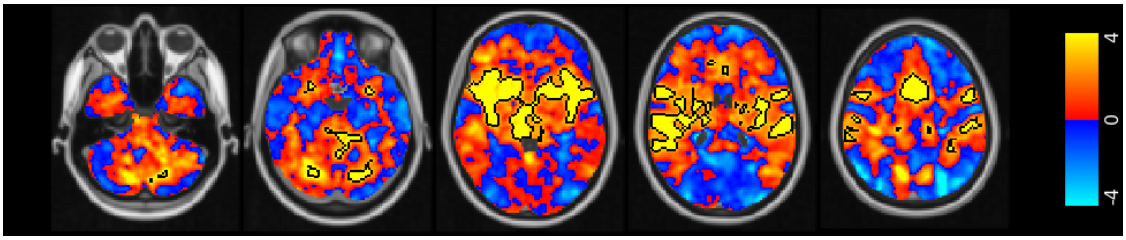
Figure 6.7: Average BOLD activation to the 64, 128 and 512 mN pinprick stimuli (n = 5). a-c) Images are shown as unthresholded t-statistic images overlaid on a standard brain (Fig. 6.7d), with clusters of significant activity (threshold-free cluster enhancement, family-wise error corrected at $p < 0.05$) indicated in black. d) Mask derived from the average activation map for the 512 mN pinprick (threshold-free cluster enhancement, family-wise error corrected $p < 0.05$). Areas commonly associated with pain are indicated. ACC = Anterior Cingulate Cortex; OFC = Orbitofrontal Cortex; S1 = postcentral gyrus; S2 = parietal operculum.



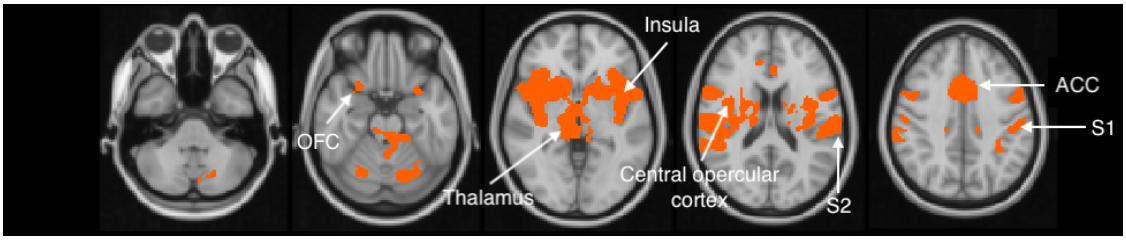
(a) Average activation to the 64 mN pinprick stimulus.



(b) Average activation to the 128 mN pinprick stimulus.

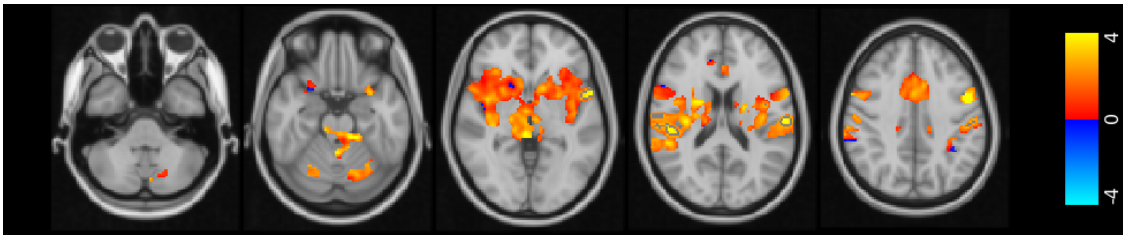


(c) Average activation to the 512 mN pinprick stimulus (family-wise error corrected $p < 0.05$).

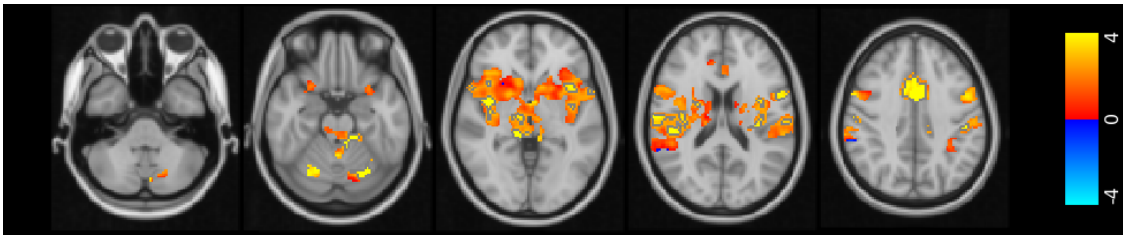


(d) Thresholded average activation to the 512 mN pinprick stimulus.

Figure 6.8: Direct comparisons of BOLD activity evoked by the 512 mN stimulus and the lower intensity stimuli, masked by mean 512 mN activation (see Fig. 6.7d). Data are shown as untresholded t-statistic maps, overlaid on a standard brain. No voxels survived family-wise error rate correction. Areas where family-wise error corrected $p < 0.1$ (cluster-based thresholding using cluster mass and an initial cluster forming threshold of $t = 2.78$, $p = 0.025$), are indicated by grey lines to aid visualisation. There was a trend for the 512 mN pinprick to evoke higher activity in the depicted areas compared to the 64 and 128 mN pinprick stimuli.



(a) 512 mN activation versus 128 mN activation.



(b) 512 mN activation versus 64 mN activation.

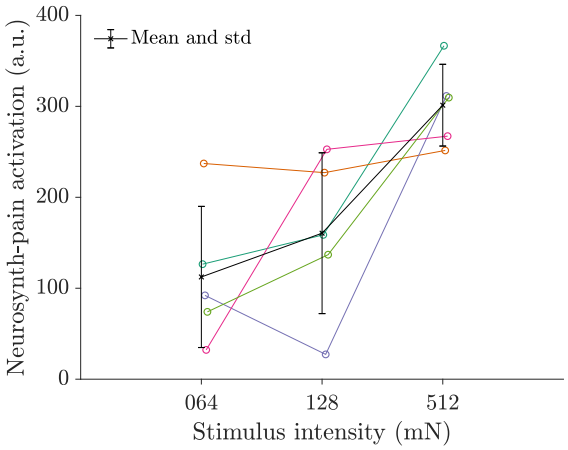


Figure 6.9: Mean weighted Neurosynth activity for the 64, 128 and 512 mN pinpricks increases with intensity. Each coloured line represents one participant, black lines represent the mean and standard deviation (std) at each intensity.

6.2.5 PEP-BOLD relationships

Having established that the pinprick stimuli evoked noxious-evoked activity in both the EEG and the fMRI recordings, relationships between the two modalities were investigated to identify ROIs for the infant study. Three group level analyses were performed.

First, the group BOLD activation to the 128 mN stimulus was modelled with PC1-128 magnitude as a covariate to identify voxels in which noxious-evoked BOLD activity positively co-varied with PC1-128 magnitude. The activity map was masked by the areas which on average were significantly activated by the highest intensity pinprick stimulus (512 mN, see Fig. 6.7d) to limit the analysis to brain areas involved in processing these stimuli. As expected with this small sample size, no voxels survived FWE correction. The two largest clusters were found in the bilateral orbitofrontal cortex (OFC) (Table 6.1,Fig. 6.10, Fig. 6.13A-B).

Next, the analysis was repeated for the 512 mN BOLD responses, this time using PC1-512 magnitude as a covariate. No voxels survived FWE correction. The largest clusters were present in the contralateral putamen and contralateral postcentral gyrus (Table 6.2, Fig. 6.11, Fig. 6.13C-D). None of the clusters overlapped with the clusters identified in the 128 mN analysis.

Third, to investigate whether there were any areas that co-activated with *both* the PC1-128 and PC1-512 magnitude, a group level analysis was carried out using the mean of the PC1-128 and PC1-512 magnitude as a covariate. This analysis was only carried out for the 512 mN BOLD responses, as the overall BOLD signal for this stimulus was higher and it was expected that this would increase SNR. Here, the largest clusters were found in the bilateral thalamus, bilateral putamen and caudate, the ACC and in opercular areas. Visually, BOLD activity appeared more strongly associated with PC1-128 magnitude than with PC1-512 magnitude in these areas (Fig. 6.13E-K).

The largest clusters identified in each of the three approaches (arbitrarily defined as $p < 0.6$) were taken forward as ROIs in the infant analysis (see Chapter 7).

Besides a voxelwise analysis, I also investigated whether Neurosynth activity evoked by the 512 mN pinprick stimulus was related to EEG responses. Visually, a positive relationship could be observed between Neurosynth activity and PC1-128 and PC1-512 magnitude (Fig. 6.13L).

Table 6.1: Clusters in which BOLD activity evoked by the 128 mN pinprick was associated with variation in PC1-128 magnitude. Size is given in number of voxels. P refers to cluster-mass corrected p-values (arbitrary initial cluster-forming threshold: t = 3.1, p = 0.025). None of the clusters survived FWE correction. The two largest clusters were located in the bilateral orbitofrontal cortex. ROI = region of interest.

Size	P	Anatomical areas	Side	ROI name
14	0.58	Frontal orbital cortex, Insula	R	OFC R
12	0.58	Frontal orbital cortex, Inferior frontal gyrus, Frontal operculum	L	OFC L
5	0.79	Thalamus	R	
2	0.92	Brain stem		
1	0.94	Frontal Orbital Cortex, Temporal Pole	L	
1	0.94	Central Opercular Cortex, Heschl's gyrus, Insular Cortex	L	

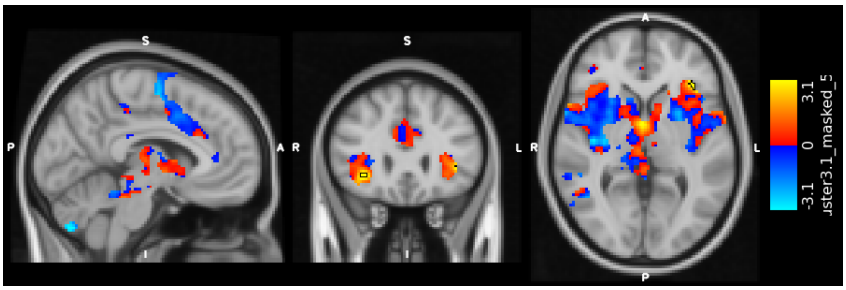


Figure 6.10: Unthresholded t-statistic maps showing areas in which BOLD activity evoked by the 128 mN stimuli co-varies with PC1-128 magnitude. The two clusters with the largest mass (see Table 6.1), located in the bilateral orbitofrontal cortex, are indicated in black.

Table 6.2: Clusters in which BOLD activity evoked by the 512 mN pinprick was associated with variation in PC1-512 magnitude. Size is given in number of voxels. P refers to cluster-mass corrected p-values (arbitrary initial cluster-forming threshold: t = 3.1, p = 0.025). None of the clusters survived FWE correction. The two largest clusters were located in the contralateral putamen and the contralateral postcentral gyrus. ROI = region of interest.

Size	P	Anatomical areas	Side	ROI name
47	0.23	Putamen, Caudate	R	Putamen R
17	0.59	Postcentral gyrus	R	Postcentral Gyrus R
10	0.84	Planum polare, Heschl's gyrus, Insular cortex	R	
9	0.88	Precentral gyrus, Middle frontal gyrus, Inferior frontal gyrus	R	
6	0.98	Thalamus	R	
5	0.99	Central Opercular Cortex	R	
4	0.99	Insular Cortex	R	
3	0.99	Frontal Orbital Cortex	R	
3	0.99	Middle Temporal Gyrus	R	

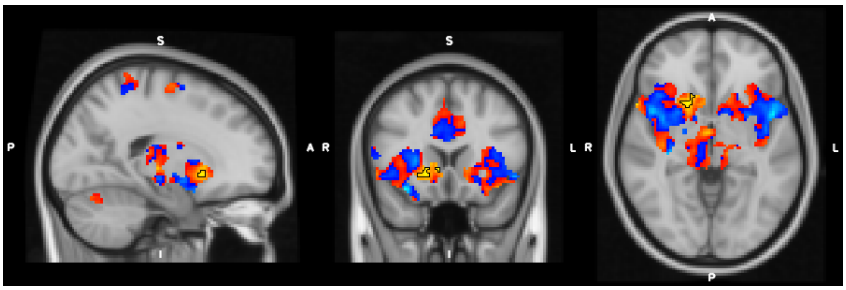


Figure 6.11: Unthresholded t-statistic maps showing areas in which BOLD activity evoked by the 512 mN stimuli co-varies with PC1-512 magnitude. The largest cluster (see Table 6.2), located in the contralateral putamen and caudate, is indicated in black.

Table 6.3: Clusters in which BOLD activity evoked by the 128 mN pinprick was associated with variation in mean PC1-128 and PC1-512 magnitude. Size is given in number of voxels. P refers to cluster-mass corrected p-values (arbitrary initial cluster-forming threshold: t = 3.1, p = 0.025). The two largest clusters were located in the bilateral thalamus. ROI = region of interest.

Size	P	Anatomical areas	Side	ROI name
114	0.033	Thalamus	R	Thalamus R
58	0.22	Thalamus	L	Thalamus L
53	0.24	Putamen, Caudate	R	Putamen R
33	0.33	Central opercular cortex	R	Central Opercular Cortex R
23	0.5	Anterior cingulate cortex, Paracingulate gyrus	L	ACC-PCC L
19	0.58	Caudate, Accumbens	L	Caudate L
17	0.58	Anterior cingulate cortex, Juxtapositional lobule	L	ACC-SMA L
14	0.77	Parietal operculum	L	
11	0.88	Postcentral gyrus, Superior parietal lobule	L	
8	0.93	Caudate	L	
7	0.96	(Cerebellum)	L	
6	0.98	Thalamus	R	
6	0.98	Juxtapositional lobule	R	
5	0.99	Parietal operculum, Supramarginal gyrus	R	
4	0.99	(Cerebrospinal fluid)		

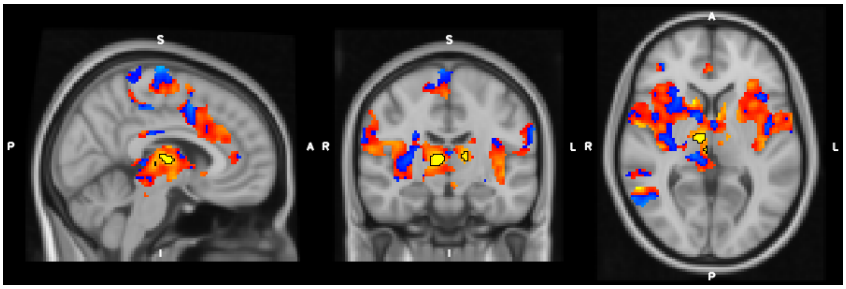


Figure 6.12: Unthresholded t-statistic maps showing areas in which BOLD activity evoked by the 512 mN stimuli co-varies with pinprick-evoked EEG activity (mean of PC1-128 magnitude and PC1-512 magnitude). The two largest clusters, located in the bilateral thalamus (see Table 6.3), are indicated in black.

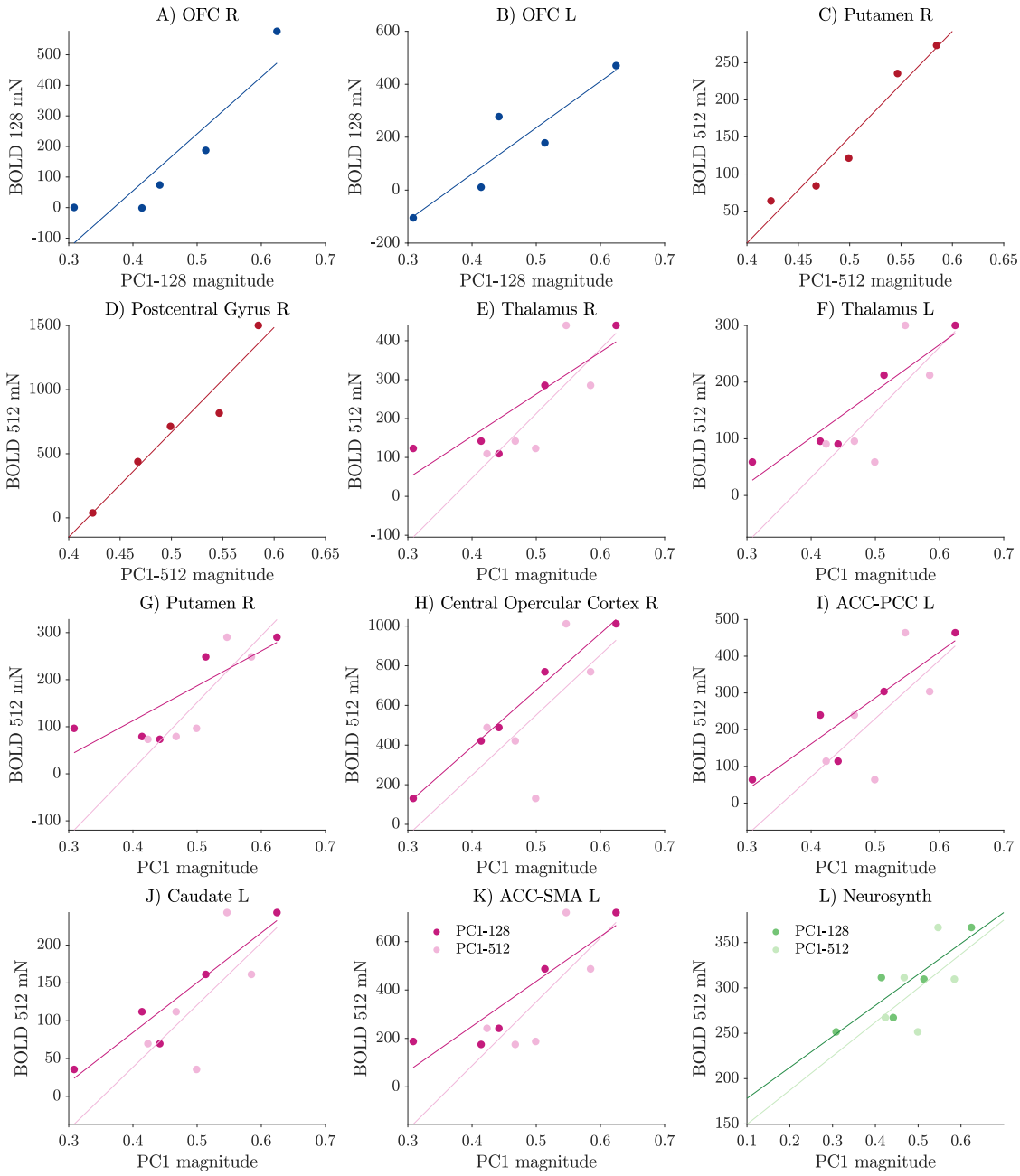


Figure 6.13: Correlations between PC1-128 magnitude (a.u) and PC1-512 magnitude and mean BOLD activity (a.u.) in all identified clusters with family-wise error corrected $p < 0.6$ and in areas making up the Neurosynth template. These are the ROIs that were taken forward to Chapter 7. Each dot indicates one participant's response, lines indicate lines of best fit. A-B) Correlation between EEG PC1-128 magnitude and BOLD activity evoked by the 128 mN pinprick in the areas that covary with PC1-128 magnitude (see Table 6.1). C-D) Correlation between EEG PC1-512 magnitude and BOLD activity evoked by the 512 mN pinprick in the areas that covary with PC1-512 magnitude (see Table 6.2). E-K) Correlation between EEG PC1-128 magnitude and EEG PC1-512 magnitude and BOLD activity evoked by the 512 mN pinprick in the areas that covary with the mean of PC1-128 and PC1-512 (see Table 6.3). L) Correlation between weighted mean Neurosynth activity and EEG PC1-128 magnitude and EEG PC1-512 magnitude. L = left; R = right; OFC = Orbitofrontal Cortex; ACC = Anterior Cingulate Cortex; PCC = Posterior Cingulate Cortex; SMA = Juxtapositional Cortex / Supplementary Motor Area.

6.3 Discussion

The pinprick stimuli evoked BOLD activity and pinprick-evoked potentials that were consistent with earlier reports [357, 8, 147, 265, 318, 317, 316, 315]. Neurosynth BOLD activity and BOLD activity in clusters in several brain regions showed a positive association with pinprick-evoked potential responses. While - unsurprisingly, considering the small sample size - none of these associations survived FWE correction and should be interpreted with caution, these trends can be used to inform the infant pilot study.

6.3.1 PEP-BOLD relationships in adults

Some of the potential ROIs were located in brain regions that have been linked to noxious-evoked EEG responses by other authors and are consistently activated in studies of human pain [357], such as the thalamus, postcentral gyrus, ACC and opercular cortex. BOLD activity in clusters in the bilateral thalamus were related to PEPs evoked by both the 128 and 512 mN experimental stimuli. In addition to regions identified by other authors, the largest clusters found in the 128 mN EEG-BOLD comparisons were located in the bilateral orbitofrontal cortex, and the largest clusters for the 512 mN EEG-BOLD comparisons were found in the putamen and caudate. The caudate and putamen are part of the basal ganglia, which form a relay between the thalamus and cortex to modulate motor, cognitive and affective responses [29]. Preclinical and clinical studies have shown that the basal ganglia play an important role in nociception [29], for instance the caudate has been linked to variation in subjective pain reports [221, 278], while the putamen is involved in both sensory and motor responses to painful stimuli [294]. The OFC plays a complex role in mediating the relationship between stimulus intensity and subjective reports [111] and sensitivity to visceral pain [83]. Neurosynth activity was also visually positively correlated with PC1-128 and PC1-512. It is important to note that areas which are correlated with PEP magnitude are not necessarily the source of PEPs, instead, their activity simply co-varies with an individual's PEP magnitude. If statistically significant relationships between BOLD activity in

these ROIs and infant EEG responses were found in a larger-scale infant study, this would imply that inter-individual differences in EEG template magnitudes may be related to differences in various aspects of pain. However, this cannot be confirmed based on the weak associations and small sample size in this study.

As the goal of this study was to identify ROIs for the infant study, any potential ROIs were pragmatically delineated as the areas with the strongest relationship between BOLD activity and PEP amplitude, as summarised by PC1-128 and PC1-512. Several ambiguities should be pointed out before the results are taken forward to the infant study.

6.3.2 Ambiguous interpretation of adult PEP amplitude

Verbal pain reports and BOLD activity were graded with intensity within participant, whereas PC1-128 and PC1-512 magnitude were not graded with intensity. This may be due to the use of intensity-specific templates, which have a better fit in the data in which they were derived compared to data from the other intensities. However, visual comparison of raw amplitudes also suggested that PEP amplitude was not graded by intensity. This lack of intensity grading in PEP amplitudes could be due to the sub-optimal time-locking of the 512 mN stimulus or the small sample size, but has also been reported by other authors [318, 265] and might thus be physiological.

In addition, pinprick-evoked potentials were not correlated with pain scores within intensity, which has also been previously reported [147]. In this study, this may have to do with the small sample size or the set-up of the study, where participants were asked to give an overall rating in the form of a rounded number across a train of 10 stimuli, instead of trial-to-trial pain ratings. Effectively this lead to the pain scores being categorical variables and subtle inter-individual differences in perception may have gone unnoticed. In addition, participants had experienced all three stimulus intensities before being asked to provide pain ratings and may have felt inclined to give each intensity a different score to indicate a stimulus' intensity relative to the other stimuli. As the lowest possible score was 0, and the

lowest intensity was rated as 0 or 1 by all participants, this would leave little room for variability in pain scores for the 128 mN stimulus due to a floor effect.

Another explanation is that variability in PEPs is driven by inter-individual differences in salience, arousal or attention, rather than pain. While the pinprick stimuli probe Aδ fibers and spinothalamic tract neurons [147], responses of similar morphology can also be observed after exposure to other salient procedures [233, 234], and similar EEG responses to laser stimulation can be influenced by top-down processes such as attention [102]. There may be other aspects of the evoked potential, such as latency to peak [265], that are graded with intensity or pain reports, but these were not considered here because the infant template is amplitude-based. The interpretation of PEP amplitude differences in adults is thus ambiguous, and may not represent variation in experienced pain. In addition, it may not represent the same physiological process as the infant noxious-evoked EEG template magnitude, which is graded with stimulus intensity [132].

6.3.3 Inconsistencies in PEP-BOLD relationships across intensities

The areas in which PEP amplitude co-varied with BOLD activity were different for the 128 mN and the 512 mN stimulus. This is most likely the result of noisy estimates for both the BOLD and EEG responses due to the small sample size. The 512 mN pinprick was not automatically time-locked to the EEG recordings, which may have introduced additional variance in the 512 mN EEG response data.

Besides a problem in power, it is possible that there is a physiological difference in the neural processing of the 128 and 512 mN stimuli. The 512 mN was perceived as mildly painful on the NRS, yet was deemed more painful than the 128 mN stimulus. There was also a trend for the 512 mN to evoke higher BOLD activity than the 128 mN especially in the insular and opercular regions, both areas which have been linked to the preferential processing of pain compared to non-painful, salient stimuli [142, 279]. The 128 mN stimulus BOLD activity might thus be a representation of somatosensory processing that is common to non-painful and

painful stimuli alike, while the activity evoked by the 512 mN reveals inter-individual differences underlying natural variation in pain processing. In line with this, higher stimulus intensities have been linked to more consistent BOLD activity within individuals across sessions [129]. Studies with larger sample sizes would be needed to investigate this specifically, but this is beyond the scope of the current pilot study. For this study, this difference in results between the two stimulus intensities adds to the uncertainty about the robustness of the identified ROIs.

6.3.4 Summary and next steps

In the adult pilot study, I confirmed that experimental pinprick stimuli evoke patterns of noxious-evoked EEG and BOLD activity that are consistent with previous studies, and which vary from individual to individual. I did not find statistically significant relationships between PEP amplitude and BOLD activity in any brain region. However, I identified areas in the orbitofrontal cortex, parietal operculum, central operculum, putamen, caudate and ACC as having the strongest relationship between PEP amplitude and BOLD activity. These areas will be used as ROIs in the infant study. Because the interpretation of PEP amplitudes in adults is ambiguous and not necessarily the same as in infants, and PEP-BOLD relationships were not robustly identified in the same areas across different stimulus intensities, other approaches should also be considered in the infant pilot study.

7

Exploring fMRI correlates of noxious-evoked EEG responses in infants

This chapter outlines the main pilot study in infants. The goals were to establish the feasibility of recording noxious-evoked EEG and fMRI activity in the same infant, and to evaluate methods to investigate correlations between fMRI activity and EEG responses, as measured by the template of noxious-evoked EEG activity.

7.1 Methods

7.1.1 Participants and design

This study included seven healthy infants aged 36–44 weeks PMA, who were all studied within 15 days of life. EEG was recorded during a clinically necessary heel lance and fMRI was recorded at a separate test occasion, during which 128 mN pinprick stimuli were applied to the left foot, as described in earlier publications [116]. Detailed eligibility criteria and recruitment process are described in Chapter 2, Section 2.2.2.1. In six infants, the EEG and the MRI test occasions took place on the same day or on consecutive days, in one infant the two test occasions were ten days apart.

7.1.2 Experimental procedures

The MRI test occasion included a structural scan, fieldmap scan and a task-fMRI scan (see Chapter 2, Section 2.2.2.3). During the task-fMRI scan, a train of 10 pinprick stimuli with an intensity of 128 mN were delivered to the dorsum of the infants' left foot. These stimuli do not break the skin and do not cause behavioural distress in term infants [132]. The investigator time-locked the stimuli to the scan by pressing a button. EEG was recorded during a clinically required heel lance and a control heel lance, and for 5 infants also during 10–20 stimulations with the 128 mN pinprick stimulus.

7.1.3 EEG analysis

7.1.3.1 Preprocessing and extracting EEG template magnitude

EEG data were preprocessed with the same pipeline used in Chapter 3. All the control heel lance and heel lance epochs were included after visual inspection. A total of four pinprick stimuli epochs across three infants were rejected because of movement artefact. The template of noxious-evoked EEG activity [131] was projected to each included data epoch to obtain the corresponding EEG template magnitude. Data were Woody-filtered by 100 ms prior to projection of the template instead of 50 ms, as this visually improved template fit in the small group of infants investigated here. For the pinprick stimuli, the template was fit to each stimulus epoch individually and then averaged within infant.

7.1.3.2 Statistics

The mean, standard deviation and coefficient of variation (standard deviation divided by the mean) of the EEG template magnitude for the heel lance, control heel lance and pinprick stimuli was calculated to evaluate group responses. If the data is of sufficient quality, it is expected that the EEG template magnitude evoked by the heel lance is larger than that evoked by the control heel lance [131], and that it is correlated with the response to the pinprick stimuli [46]. To test this, the EEG template magnitude evoked by the heel lance was compared to the EEG

template magnitude evoked by the control heel lance using a paired t-test and the Pearson correlation coefficient was calculated between the EEG template magnitude evoked by the heel lance and the EEG template magnitude evoked by the pinprick stimuli. P-values for both tests were derived non-parametrically in PALM using exhaustive sign-flips and permutations, respectively [350].

7.1.4 fMRI Single-subject analysis

7.1.4.1 Preprocessing

Prior to my DPhil, paired EEG and fMRI data for three infants had already been collected. The fMRI data for these infants were preprocessed by a colleague as part of the work published here: [20, 21]. I applied the same analysis pipeline to the prospectively collected fMRI data.

fMRI data were preprocessed using the developing Human Connectome (dHCP) fMRI pipeline v0.5.3, which is described in detail in [99, 20] and will be briefly outlined here. As part of this pipeline, the structural T2 scans were processed using the main dHCP structural pipeline v1.1, which performs bias field correction, brain extraction and tissue segmentation to facilitate functional to structural registration in the latter part of the fMRI pipeline [208]. Motion and distortion correction were applied simultaneously to the functional data using FSL's *EDDY* [6, 5, 4, 3]. Data were temporally filtered with a 100 second (i.e., 0.01 Hz) high-pass filter and ICA decomposition was performed using FSL's *MELODIC* [23]. FSL's *FIX*, which was previously trained on a larger infant dataset by a colleague as part of the work published in [20], was used to flag noise components [272, 122]. All components were then visually reviewed and relabelled if necessary, and FSL's FIX [272, 122] was used to regress out components flagged as noise, as well as head motion time series. Finally, a low-pass spatial 3 mm FWHM filter was applied using FSL's *SUSAN* [291] and grand mean scaling was applied. The functional data were registered to a standard 40-week GA structural template from the dHCP volumetric atlas [276] through the combination of several primary registrations. Functional data were registered to infants' native T2 structural space using FSL's

FLIRT [157, 159]. Structural images in native space were aligned to age-matched templates from the dHCP volumetric atlas [276], which were then aligned to the standard 40-week GA template from the dHCP volumetric atlas [276]. This was achieved via week-to-week transformations from the age-matched template to the standard 40-week template. Structural registration was performed using ANTs's *SyN* [14]. All transformations from infants' native functional space to standard space were combined and applied in a single step.

7.1.4.2 Single-subject statistics

At the single-subject level, the pinprick timings marked by the investigator were convolved with a double gamma haemodynamic response function suited for neonates [10, 20] and a general linear model was fitted to the BOLD time course in each voxel in FSL's FEAT (v6.00, [355]). This yielded one COPE map per infant, which depicted the amplitude of the haemodynamic response to the pinprick stimuli versus baseline in each voxel.

7.1.5 fMRI Group level analysis (mean activation)

To assess fMRI data quality, the group average BOLD response to the 128 mN pinprick stimuli was extracted. A group level GLM analysis was performed using FSL's *randomise* [350] by using the lower-level COPE images in 40-week GA infant standard space [276]. Variance smoothing of 6 mm FWHM was applied. For visual comparison with the adult data, the group level maps were transformed into adult MNI152 space using FSL's *applywarp* using the warps generated and distributed here: [276, 98]. As the goal of this study was not to test a specific hypothesis, unthresholded t-statistic maps are shown.

7.1.6 EEG template magnitude–BOLD relationships

Three approaches were used to investigate EEG template magnitude–BOLD relationships, which involve an increasing degree of spatial granularity. No statistical inferences were made due to the exploratory nature of this pilot study.

7.1.6.1 Approach 1: Neurosynth signature

As previously described, noxious stimuli evoke a widespread pattern of BOLD activity, which may underlie the pain experience. The first approach was therefore to correlate EEG template magnitudes with a summary measure of an MRI signature that is commonly associated with pain by using the same meta-analysis association map from Neurosynth as in Chapter 6. The Neurosynth signature was warped to infant space using FSL's *applywarp* using the warps generated and distributed here: [276, 98] (see Fig. 7.1) and the weighted mean activation within this spatial map was extracted from the single-subject COPEs using FSL's *fslmeants* [160, 354], as described for the adult study in Chapter 6. Neurosynth activity was plotted against the EEG template magnitudes evoked by the heel lance and the pinprick stimuli for each infant. To explore whether the removal of additional noise in the fMRI data would change the results, nuisance regressors (motion, CSF and stimulus-correlated motion) were extracted for each infant and regressed out of the individual COPE maps using *fslregfilt*, and the COPE maps were smoothed with a 3 mm FHWM spatial filter. These denoising steps are unconventional at the single-subject level and should not be implemented when the goal is to make statistical inferences. However, as the main goal of this study was to establish the feasibility and best approach to follow in a larger-scale study, these steps were useful to visualise patterns in the data.

The mean activation in the mask derived from the areas that were significantly activated by the 512 mN pinprick in adults (see Chapter 6) was extracted using FSL's *fslmeants* and plotted against the Neurosynth activity. This was done to assess whether there were any important differences in activation between the two patterns, which could be due to the nature of the pinprick stimulus compared to the stimuli used in the generation of the Neurosynth signature.

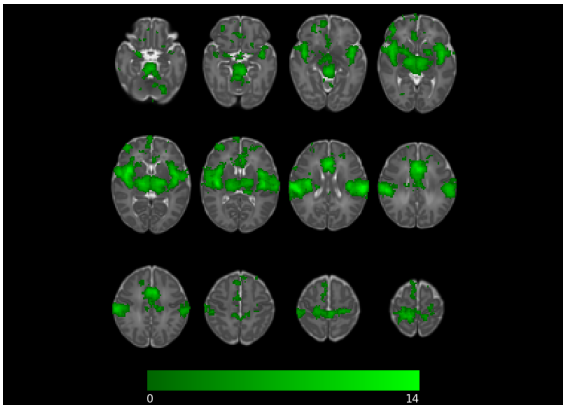


Figure 7.1: The Neurosynth signature in infant space, overlaid on a 40-week GA standard brain [276]. The Neurosynth signature is a z-statistic map. Neurosynth activity was extracted by taking the weighted mean activity in this map for each infant, using the (positive) z-statistics as weights.

7.1.6.2 Approach 2: ROI analysis

The Neurosynth pattern includes a wide range of areas. It is possible that EEG template magnitudes are only related to very specific brain regions, and that using a summary measure of the full pattern would obscure subtle effects in specific brain areas. To limit the number of comparisons, ROIs were pre-defined in the adult study in Chapter 6. These ROIs were warped to 40-week GA infant standard space [276] using using the warps generated and distributed here: [276, 98] (see Fig. 7.7). The mean activity in each ROI was extracted for each infant's COPE map (both the original and the smoothed and cleaned COPE) using FSL's *fslmeants* and plotted against the heel lance EEG template magnitude for each infant. No statistical tests were conducted due to the exploratory nature of the study.

7.1.6.3 Approach 3: Voxelwise analysis

As described in Chapter 6, the results from the adult study may not translate directly to the infant study, and it is possible that relationships are missed when only using the ROI approach. Therefore, a voxelwise GLM group analysis with the heel lance EEG template magnitude as a covariate (as well as a mean regressor) was also carried out using FSL *randomise* [350], analogous to the adult study in Chapter 6. The result was a voxelwise t-statistic map, which could be used to

understand in which voxels BOLD activity and EEG template magnitude were positively associated. The 512 mN adult activation was used as a mask for the group-level analysis, as I was interested in identifying areas that are activated by noxious stimulation and co-vary with the heel lance EEG template magnitude. To aid interpretation, clusters with strongest relationship between EEG template magnitude and BOLD activity were extracted. FWE corrected p-values for each cluster were derived with *randomise* [350] based on cluster mass (cluster-forming threshold t = 2.6, corresponding to p = 0.025 for 5 df). The cluster-forming threshold of t = 2.6 is arbitrary and much lower than would be used if the goal was to make statistical inferences. However, here p-values were only used to better understand the unthresholded t-statistic maps. The clusters were warped to adult MNI152 space, and *featquery* [158] was used to identify their anatomical location based on the Harvard-Oxford Subcortical Structural Atlas and the Harvard-Oxford Cortical Structural Atlas [207, 103, 71, 103].

7.2 Results

7.2.1 Feasibility of an EEG-fMRI study in infants

The goal of the infant study was to assess the feasibility of recording noxious-evoked EEG and fMRI activity in the same infant, and pilot potential EEG–fMRI relationships in infants, which can be explored further in a larger follow-up study. Noxious-evoked EEG and fMRI data was recorded in seven healthy infants aged between 36 and 44 weeks PMA. The EEG and the MRI session were 0 to 10 days apart.

7.2.2 Infant EEG and BOLD responses to clinical procedures and pinprick stimuli

The characteristic EEG response to heel lancing and pinprick stimuli was clearly distinguishable in the average EEG data for the seven infants in the study. The mean EEG template magnitude was 1.24 (n = 7, standard deviation = 0.92) for the heel lance, 0.50 (standard deviation = 0.39, n = 7) for the control heel lance and 0.64

(standard deviation = 0.15, n = 5) for the pinprick stimuli (Fig. 7.2). As a quality control measure, I confirmed that the the heel lance EEG template magnitude was significantly higher than the control EEG template magnitude (two-tailed paired t-test, mean difference = 0.74, t = 3.40, p = 0.016). The EEG template magnitude evoked by the heel lance was highly variable between infants (Fig. 7.3), and was correlated with the response to the pinprick stimuli (n = 5, Pearson r = 0.90, p = 0.017). Inter-individual variation in EEG amplitude across the entire epoch was qualitatively larger for the heel lance and control heel lance stimuli, which are single events, compared to the pinprick stimuli (Fig. 7.2). In line with this, the coefficient of variation (the ratio of the standard deviation to the mean) of the EEG template magnitude was larger for the control heel lance and heel lance compared to the pinprick stimuli (coefficients of variation 0.74, 0.79 and 0.24, respectively).

A direct comparison of the adult and infant noxious-evoked response to the 128 mN pinprick showed that the pinprick-evoked potential had a shorter latency and lower amplitude in adults than in infants (Fig. 7.4).

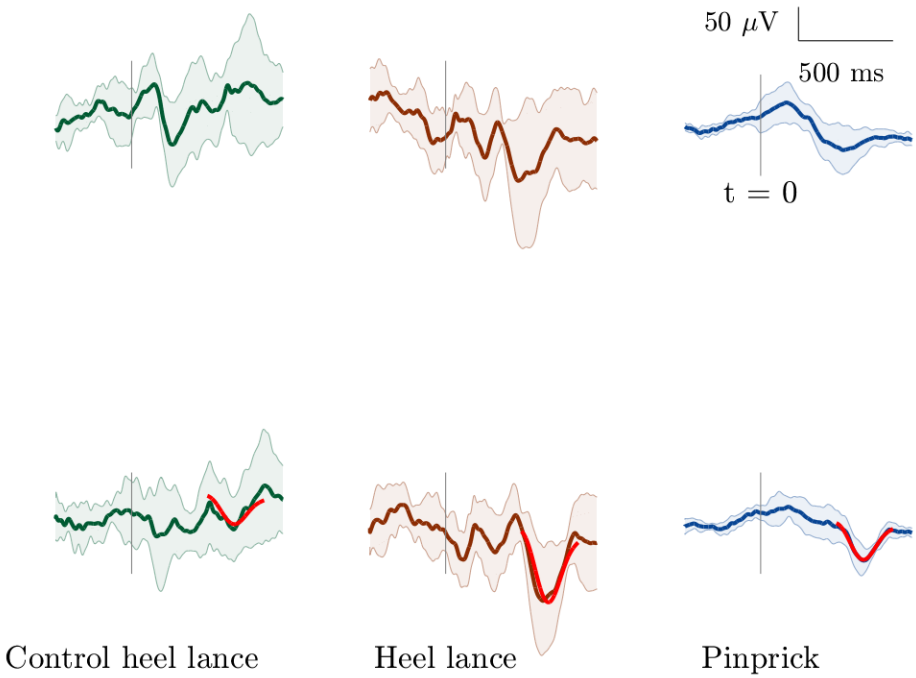


Figure 7.2: Mean infant EEG responses to the control heel lance (n = 7), heel lance (n = 7) and 128 mN pinprick stimuli (n = 5). Upper row represents the filtered and baseline-corrected epochs, before Woody-filtering, and lower row represents the Woody-filtered trials with the template of noxious-evoked EEG activity projected in bright red. The standard deviation (represented by the shaded area) appears smaller for the pinprick stimuli compared to the control heel lance and heel lance stimuli. The stimuli took place at t = 0.

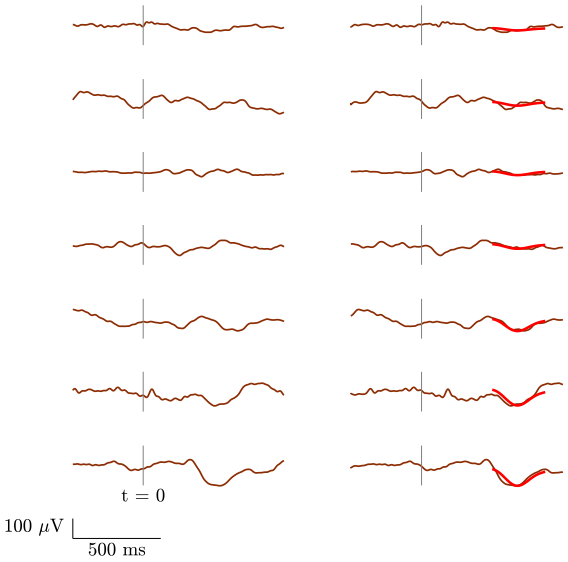


Figure 7.3: Individual infant time-locked EEG responses to the noxious heel lance. Left column shows the filtered and baseline-corrected epochs prior to Woody-filtering and projecting the template of noxious-evoked EEG activity. Right shows the Woody-filtered epochs with the template of noxious-evoked EEG activity overlaid in bright red. The heel lance took place at t = 0.

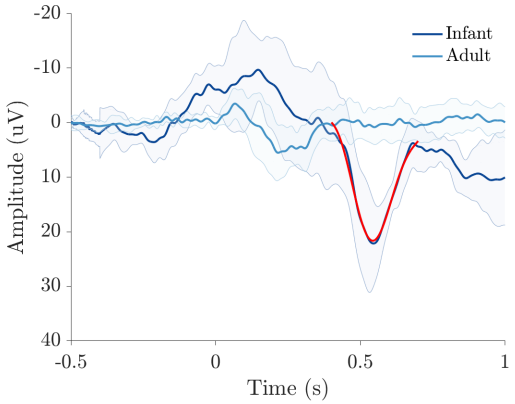


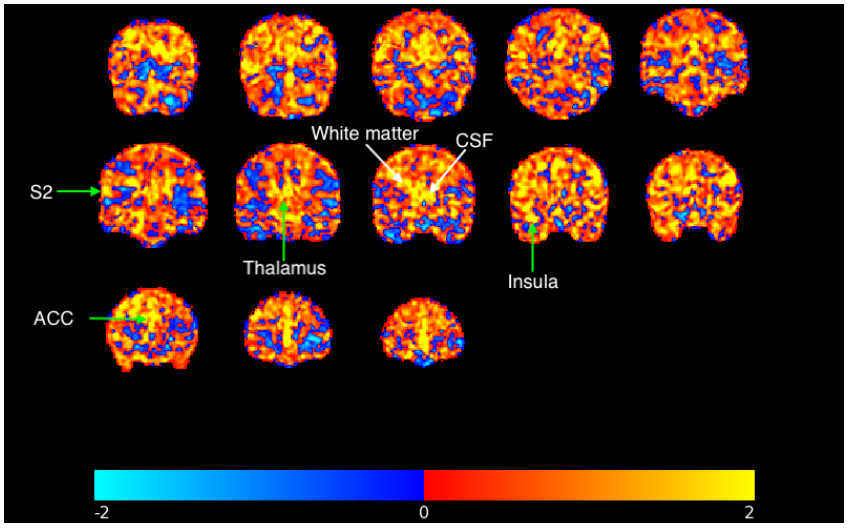
Figure 7.4: Mean time-locked EEG response to the 128 mN pinprick stimulus in infants (n = 5) and adults (n = 6, see Chapter 6). The infant data are Woody-filtered to the template of noxious-evoked EEG activity, which is overlaid in red. The latency and amplitude of the noxious-evoked EEG potential was larger in infants than in adults. The stimulus took place at time = 0.

Average BOLD activation (unthresholded) was visible in areas commonly associated with pain, such as the bilateral thalamus, ACC and contralateral parietal operculum (Fig. 7.5a). However, spurious signal was also present in white matter and CSF, which is likely to be the consequence of the small sample size. Qualitatively,

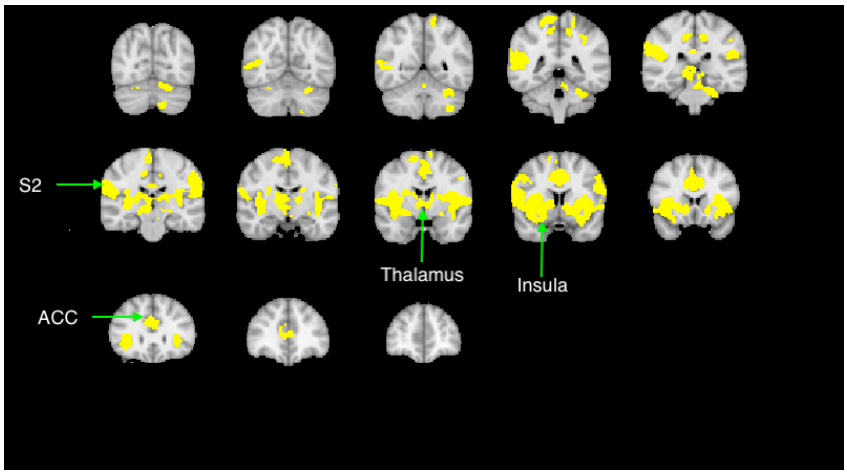
there was little overlap with mean BOLD activity in the adult study, except in the ACC, bilateral thalamus and contralateral parietal operculum (Fig. 7.5a and Fig. 7.5b).

Similarly to the variation in EEG responses, the extent and magnitude of the noxious-evoked BOLD activity was highly variable across infants, as can be seen in the wide range of Neurosynth activity (Fig. 7.6A). Neurosynth activity was highly correlated to activity in the areas that were significantly activated on average in the adult study (Fig. 7.6A).

Figure 7.5: Comparison of adult and infant BOLD activity evoked by the pinprick stimuli.



(a) Mean infant BOLD activity to the pinprick stimuli presented as unthresholded t-statistic maps and transformed to adult MNI152 standard space (see Fig. 7.5b for slices). Areas of relatively high activation include the ACC, contralateral parietal operculum and bilateral thalamus (green arrows), but high BOLD signal is also visible in CSF and white matter (white arrows). S2 = Parietal Operculum; ACC = Anterior Cingulate Cortex; CSF = cerebrospinal fluid.



(b) Adult anatomical standard MNI152 brain with significant BOLD activations found in the adult study overlaid in yellow for comparison with Fig. 7.5a. S2 = Parietal Operculum; ACC = Anterior Cingulate Cortex.

7.2.3 Exploring analytical methods to evaluate EEG template magnitude–BOLD relationships in infants

Three approaches were used to quantify EEG template magnitude–BOLD relationships.

First, a summary measure of fMRI activity, as captured by the Neurosynth

signature (see Fig. 7.1), was plotted against EEG template magnitudes. No correlation was observed between Neurosynth activity and EEG template magnitudes to the heel lance nor the pinprick stimuli (Fig. 7.6, panel B-C).

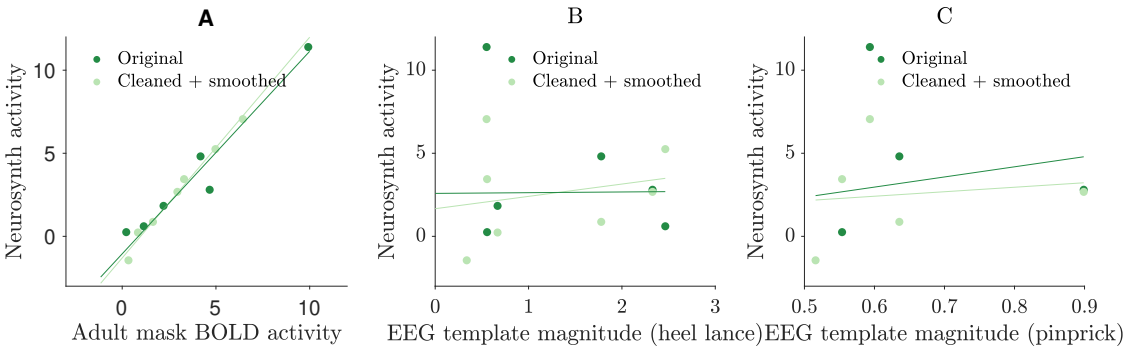


Figure 7.6: Correlations between Neurosynth activity (in arbitrary units) and 1) activity in the areas that were significantly activated on average in the adult study and 2) EEG template magnitudes (in arbitrary units). "Original" refers to raw contrast of parameter estimates (COPE) images, whereas "Cleaned + smoothed" refers to COPE images where additional single-subject level nuisance regressors were removed and smoothing was applied. A) Neurosynth activity was highly correlated with mean BOLD activity in the areas that were significantly activated in adults. B) Neurosynth activity was not correlated with EEG template magnitude evoked by the heel lance. C) Neurosynth activity was not correlated with EEG template magnitudes to the pinprick stimuli.

Second, mean BOLD activity in the adult ROIs (Fig. 7.7) was plotted against the EEG template magnitudes to the heel lance. BOLD activity in most of the ROIs showed weak negative correlations with EEG template magnitudes to the heel lance (Fig. 7.8), which disappeared when nuisance regressors were removed from the data and additional smoothing was applied at the single-subject level. The only weak relationships that remained were a positive relationship between the EEG template magnitude and the post-central gyrus ROI (Fig. 7.8D) and the ACC-SMA ROI (Fig. 7.8K) and a weak negative relationship between EEG template magnitude and the ACC-PCC ROI (Fig. 7.8I).

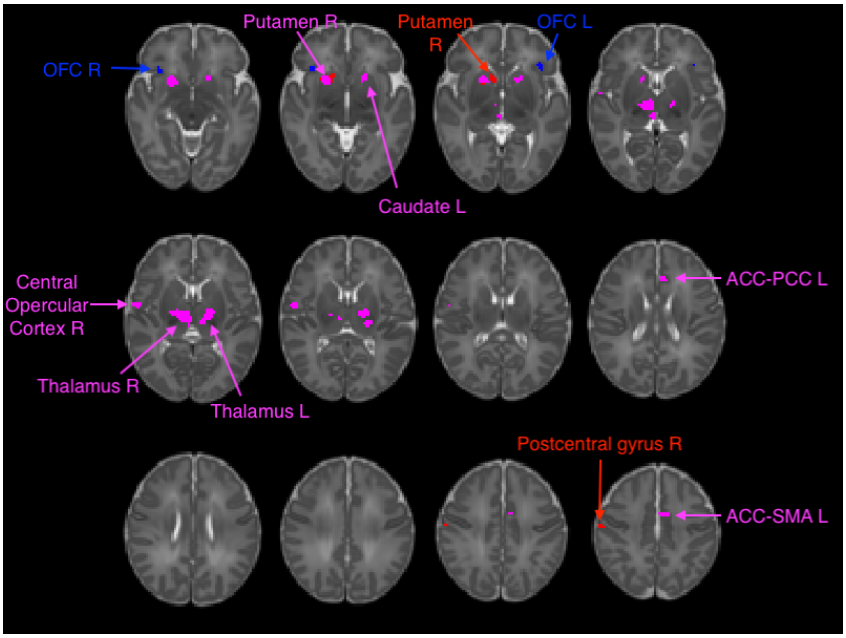


Figure 7.7: The 11 ROIs identified in the adult data in infant space, overlaid on a 40-week GA standard brain [276]. The two ROIs in the bilateral orbitofrontal cortex in which PC1-128 magnitude co-varied with 128 mN BOLD evoked activity in adults are depicted in blue. The two ROIs in the contralateral postcentral gyrus and the contralateral putamen in which PC1-512 magnitude co-varied with 512 mN BOLD evoked activity in adults are depicted in red. The ROIs in which both PC1-128 and PC1-512 magnitude co-varied with 512 mN BOLD evoked activity in adults are depicted in pink (see Section 6.2.5). Colours match those in Fig. 7.8.

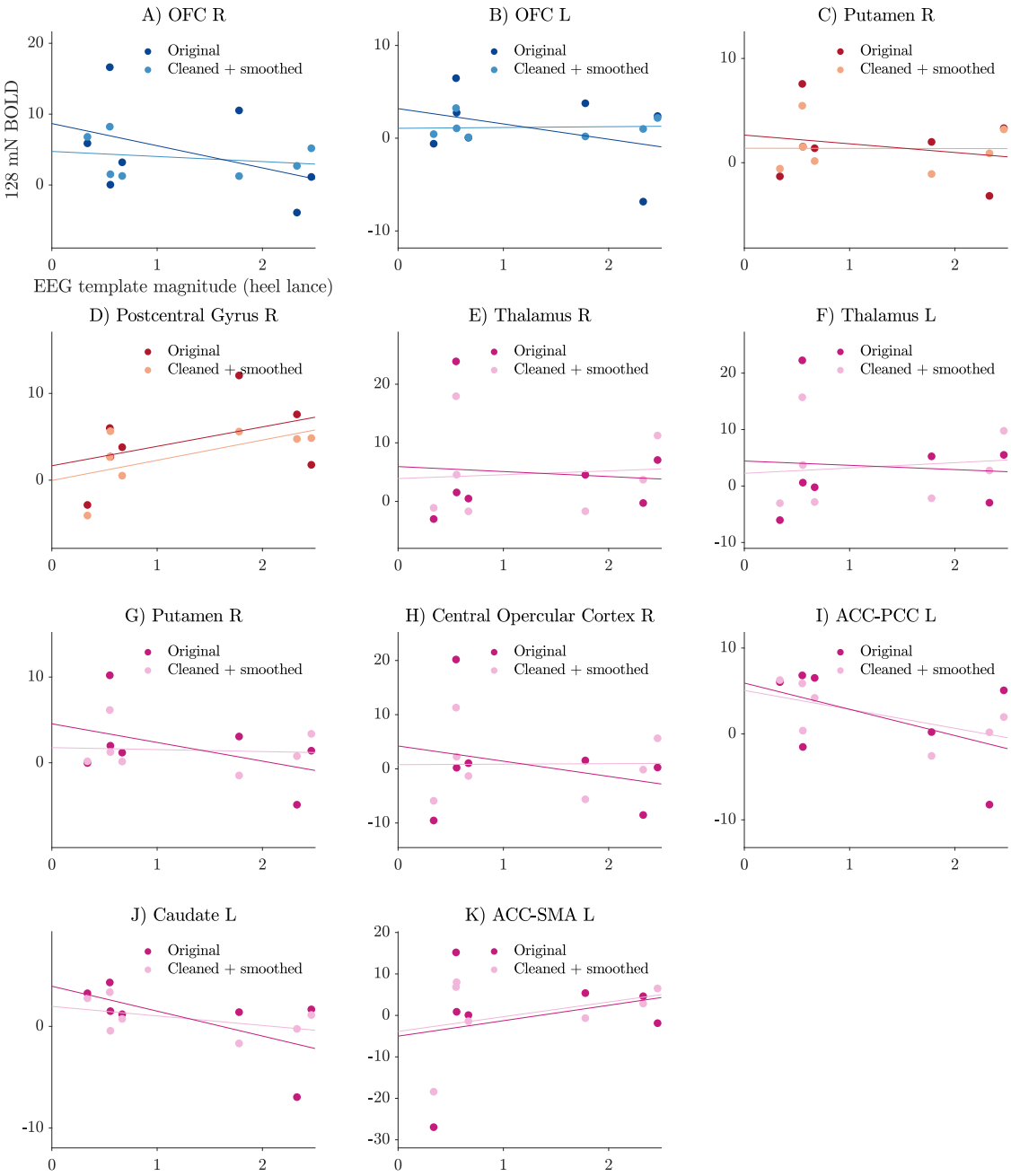
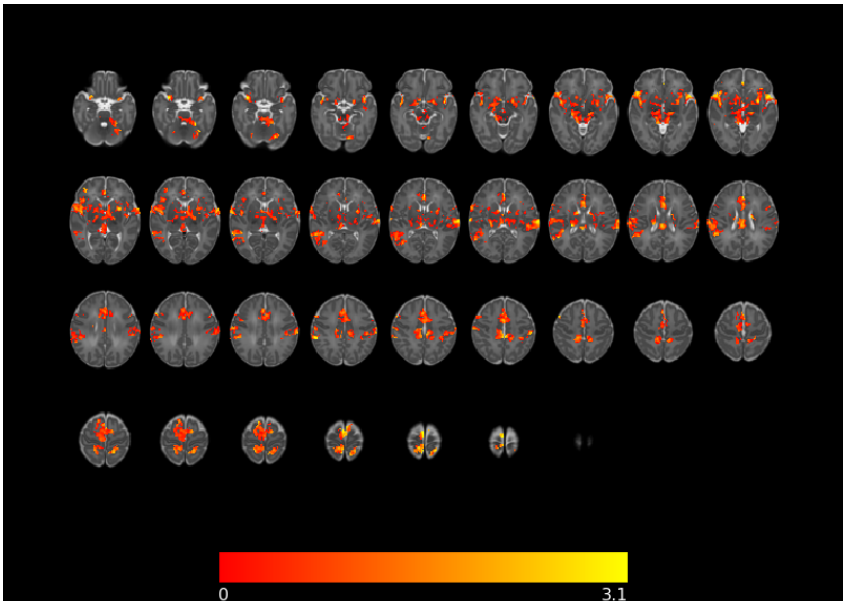
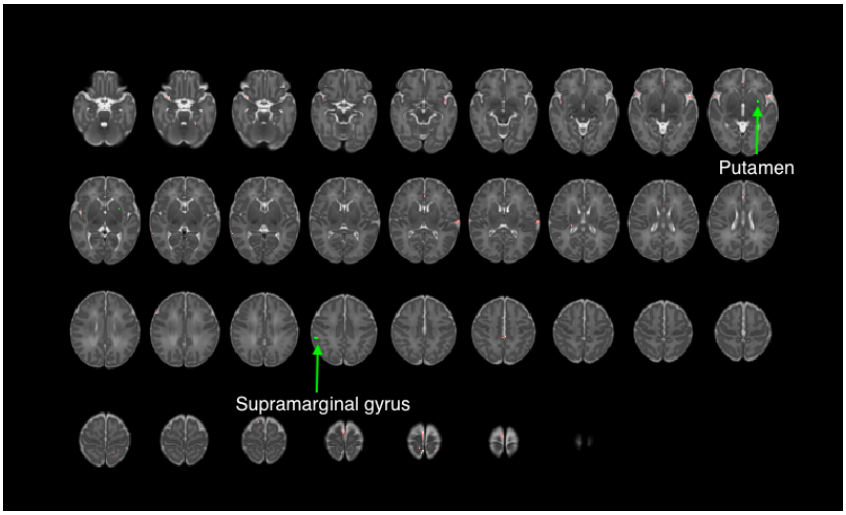


Figure 7.8: Correlations between heel lance EEG template magnitudes (arbitrary units) and activity in the adult ROIs (arbitrary units). ROIs in which PC1-128 magnitude co-varied with 128 mN BOLD evoked activity in adults are depicted in blue. ROIs in which PC1-512 magnitude co-varied with 512 mN BOLD evoked activity in adults are depicted in red. ROIs in which both PC1-128 and PC1-512 magnitude co-varied with 512 mN BOLD evoked activity in adults are depicted in pink (see Section 6.2.5). Colours match those in Fig. 7.7. "Original" refers to raw contrast of parameter estimates (COPE) images, whereas "Cleaned + smoothed" refers to COPE images where additional single-subject level nuisance regressors were removed and smoothing was applied. Mean BOLD activity in the adult ROIs was not correlated with heel lance EEG template magnitudes, except in the post-central gyrus ROI and the ACC-SMA ROI, in which a weak positive association was present, and in the ACC-PCC ROI, in which a weak negative correlation was apparent.

Finally, a voxelwise analysis was carried out to identify areas in which BOLD activity co-varied with the EEG template magnitude evoked by the heel lance (Fig. 7.9). To aid interpretation, the t-statistic maps were arbitrarily thresholded at $t = 2.6$ ($5\ df, p = 0.025$, analogous to the adult study) to visualise clusters of voxels with the highest t-statistics. Most voxels surviving this threshold were located in the CSF. However, two clusters were present in grey matter. One cluster (number of voxels = 7, FWE cluster-mass corrected p-value = 0.63) was located in the left (ipsilateral) putamen and a second cluster (number of voxels = 8, FWE cluster-mass corrected p-value = 0.60) was located in the right (contralateral) supramarginal gyrus (Fig. 7.9b).



(a) Untresholded.



(b) Thresholded.

Figure 7.9: Voxelwise analysis of areas in which BOLD activity co-varies with the heel lance EEG template magnitude. a) Unthresholded t-statistic maps (masked by the mean adult 512 mN BOLD activation) showing the areas that are positively associated with heel lance EEG template magnitude. b) Arbitrarily thresholded t-statistic map ($t = 2.6$, for 5 df, $p = 0.025$), presented to aid visualisation. Most of the voxels with high signal were located in cerebrospinal fluid (pink), but there were two clusters in grey matter (green, located in the putamen and the supramarginal gyrus).

7.2.4 Summary

Infants displayed variable levels of noxious-evoked EEG and BOLD activity. Three approaches were outlined to compare the two modalities. There was no clear relation-

ship between EEG template magnitude and BOLD activity in a pattern associated with pain in adults (the Neurosynth signature), nor in ROIs derived from a parallel adult study. A voxelwise analysis showed that the highest, albeit not statistically significant, relationship between EEG template magnitude and BOLD activity was located in the ipsilateral putamen and contralateral supramarginal gyrus.

7.3 Discussion

Further establishing the neuroscientific validity of the infant noxious-evoked template of EEG activity is important to improve infant pain assessment. One approach to improve neuroscientific validity is to compare noxious-evoked EEG responses with BOLD activity. The goal of the study presented in this part of my thesis was to establish the feasibility of investigating spatial neural correlates of inter-individual differences in EEG template magnitude using fMRI, and to identify potential analysis avenues for a future larger-scale study.

7.3.1 Feasibility and data quality

The current study included seven healthy term infants, who were recruited from the postnatal wards. All infants were studied during two test occasions and EEG and fMRI data were obtained and analysed for all, showing the feasibility of this approach in practice. The morphology of the time-locked EEG responses to the clinical procedures and pinprick stimuli were comparable to previous publications [290, 131, 125, 46] and the EEG template magnitude evoked by the heel lance was larger than the EEG template magnitude evoked by the control heel lance, which confirms that the EEG data are of sufficient quality to extract noxious-evoked signals. The fact that heel lance responses were correlated to responses to the pinprick stimuli provides some support for the hypothesis that variation in heel lance responses is a unique property of an individual [46].

The average noxious-evoked BOLD response was less comparable to previous work. Based on previous studies in the same population and using the same experimental stimulus [116, 115, 20], it was expected that the group average would

delineate a pattern of BOLD activity that was similar to that found in the adult study. While some areas, such as the thalamus and ACC showed (non-statistically significant) activity on average, it was evident that noise components remained in the data. This implies that the measures extracted from the individual BOLD activity maps were likely to also contain a significant amount of noise, which was confirmed by the finding that removal of motion regressors and CSF amplitude at the group level resulted in sometimes drastic differences in mean ROI activity. As a result, comparative analyses with the EEG data are likely to be unreliable. This finding was fully expected in this pilot study because of the small sample size, and it is likely that an increase in sample size will enhance the signal-to-noise ratio in the fMRI recordings.

The difference in SNR between the EEG data and the MRI data could be explained by several factors. First, the noxious pinprick stimuli are of lower intensity compared to the skin-breaking clinically required heel lances, and might thus elicit smaller brain responses. Second, infants are generally asleep inside the MRI scanner, while during the EEG session they are more likely to be naturally awake or to be awoken by the clinical procedure. Third, while EEG activity is recorded within a second after the stimulus, the BOLD signal is sluggish and lasts several seconds [10], which means it is more likely to be affected by head motion. Fourth, the EEG response is measured using a standardised template at a single electrode site, while BOLD responses are measured across thousands of voxels. Recommendations to increase data quality for future studies are discussed in Section 7.3.3.1.

7.3.2 Preliminary results

Three approaches were used to investigate EEG-BOLD relationships in infants. EEG template magnitudes evoked by the heel lance were not positively correlated with BOLD activity in the Neurosynth signature, nor in the the majority of the ROIs derived in the adult study, with the potential exception of the contralateral postcentral gyrus and the contralateral ACC-SMA ROI, in which visually a positive relationship was present. In the voxelwise analysis, BOLD activity was most strongly

related to EEG template magnitude in clusters in the supramarginal gyrus and the putamen. The contralateral putamen was one of the areas also identified in the adult study and the supramarginal gyrus has been linked to variation in EEG responses in adults in previous study [224]. As the data quality was low, and the number of infants was small, these results should not be over-interpreted, yet these findings are useful to generate hypotheses for future studies.

The lack of correlation between BOLD activity in the Neurosynth signature and the adult ROIs can be explained by several factors, which should be taken into account in follow-up studies.

1. Lack of power due to the small sample size and the relatively low SNR in the infant fMRI data. This factor is addressed in Section 7.3.3.1.

2. Inappropriate choice of spatial ROIs. As described in Chapter 6, the adult study had some important limitations, including a low sample size, non-robust EEG-BOLD relationships and ambiguous interpretation of amplitude differences in adult pinprick-evoked potentials.

   In addition, the infant study confirmed that the adult and infant noxious-evoked EEG potential to the same pinprick stimulus are morphologically different. The adult PEP at Cz consists of an N-P complex between 0–300 ms post-stimulus, while the infant PEP consists of a larger-amplitude negative waveform around 400–700 ms post-stimulus. The infant PEP could be a precursor to the adult PEP, which the long latency likely arising through a lack of myelination. Alternatively, the late infant waveform could be a fundamentally different brain response, as hinted on in a previous comparative study of responses to a skin-breaking procedure in adults and infants [90].

   The Neurosynth signature includes multiple large brain areas, therefore correlations between EEG and BOLD-activity in smaller areas may be lost. A recent study showed that although adults and infants have similar noxious-evoked BOLD responses, variability is present at the level of single brain areas

[80]. The ROI analyses focused on smaller brain areas derived from the adult study.

Thus, there is uncertainty about the validity of the ROIs derived in the adult study because of the low quality of this study, and it is unclear if inter-individual differences in infant noxious-evoked BOLD and ERPs are comparable to inter-individual differences in adult noxious-evoked BOLD and PEPs. Both these factors could lead to an incorrect selection of spatial ROIs. This topic is addressed in Section 7.3.3.3

3. Large intra-individual variation in noxious-evoked EEG and BOLD activity obscures inter-individual variation. Because heel lances are necessarily single trials, it was not possible to investigate how much of the observed variability in EEG template magnitude was attributable to variability across infants as opposed to variability within infants arising through factors such as sleep state, developmental stage or illness, which may have varied between the EEG and the fMRI test occasion. The two test occasions were close together for most infants, but were sometimes several weeks apart in the adult study. Although there is a certain degree of stability in noxious-evoked BOLD responses across sessions [312, 129], the test-retest reliability of noxious-evoked BOLD activity is influenced by various external factors, such as interval duration and number of stimuli [129]. Thus, the long inter-session interval may have contributed to increased intra-individual variance in the adult BOLD responses and thus more uncertainty about the ROIs derived in this study. This topic is addressed in Section 7.3.3.2.

4. The absence of meaningful inter-individual differences in EEG template magnitude and/or BOLD responses. It is possible that inter-individual variation in either recording methods is not related to noxious stimulation, but is caused by factors such as differences in head size or shape, developmental status or recording artefact (e.g. electrode impedance or movement). As previous infant studies have shown intensity grading in noxious-evoked BOLD

responses [80] and EEG responses [132] using the same stimuli, this explanation is less likely if data quality is sufficient. Therefore data quality should be evaluated. This topic is addressed in Section 7.3.3.1.

In the following section I will outline several recommendations to overcome these issues.

7.3.3 Recommendations for follow-up studies

7.3.3.1 Sample size and data quality checks

For adult studies, simulation of fMRI data can be used to conduct sample size calculations [82]. Unfortunately simulations are notoriously difficult to conduct for infant fMRI studies, because existing simulation software does not take into account motion [82] - a major source of noise in infant data [20]. A previous study found significant correlations between resting state fMRI activity and noxious-evoked responses in a group of 18 infants [21]. As the proposed study will involve two modalities across two different test occasions, it seems reasonable to assume that a minimum sample size of 18–25 infants will be needed.

If data from either modality are of insufficient quality, any further analyses are unreliable. Quality control checks should therefore be carried out. For the fMRI data, the mean activation should be evaluated and compared to previous reports [116, 115, 20] to assess whether the data is likely to contain a large degree of noise. Outliers should be detected and additional nuisance variables, such as signal present in white matter and CSF, and that related to motion, should be regressed out at the group analysis stage [21]. For the EEG data, responses to noxious stimuli should be compared to responses from non-noxious stimuli as described in this chapter to confirm that noxious-evoked EEG activity is present.

7.3.3.2 Design

As described in Chapter 5, simultaneous EEG-fMRI recording during noxious stimulation would be the ideal way to investigate the relationship between BOLD activity and noxious-evoked EEG template responses, but it is highly challenging

to implement in infants. It is therefore likely that a follow-up study would still make use of separate EEG and fMRI sessions. However, efforts should be taken to minimise the intra-individual variability in EEG template magnitudes by making changes to the study design as described in the next paragraphs.

Type of stimuli The current study was centred around a clinically-required blood test, as this is a stimulus that is relevant to clinical practice. However, this limited the number of eligible infants because not all infants need blood tests during their stay in hospital, and those infants who do require a blood test soon after birth may also require continuous clinical monitoring (e.g. for serious infections) or treatment (e.g. for jaundice) on the ward. In addition, the fact that heel lances are necessarily single trials impacts data quality. In this chapter the between-infant variation in heel lance EEG template magnitude was larger than the between-infant variation for EEG template magnitude for the pinprick stimuli, suggesting that a large proportion of the variation in the heel lance responses is driven by within-infant variation rather than between-infant variation — thus adding noise to the EEG estimate. These challenges could be overcome by shifting the focus from clinically required heel lances to pinprick stimuli in the EEG environment. Not being reliant on the timing of blood tests could allow for more flexibility in planning the test occasions around the participating families' schedules and the availability of staff and scanner time, and would increase the number of infants who are eligible. The use of the pinprick stimuli would allow for multiple trials, which would increase the signal-to-noise ratio for the EEG test occasion and would allow for the stimulus to be repeated if movement artefact is expected. Using the same stimulus in the EEG and the fMRI session would also facilitate comparisons between the two imaging modalities.

Timing and number of test occasions To minimise intra-individual changes that occur over time, the time period between the EEG and the fMRI session should be kept to a minimum. It would be of interest to compare responses to pinprick stimuli across two EEG sessions which are several hours or days apart to obtain an estimate of the degree of intra-individual variation. In addition, the sleep state of

the infant during each stimulus should be recorded based on behavioural measures (eyes open or closed, presence of facial movements) and EEG recordings.

7.3.3.3 Analysis

Infant EEG and fMRI data should be preprocessed with equivalent pipelines as those described here, as these were appropriate in larger sample sizes [20, 131].

Relationships between EEG template magnitude and noxious-evoked BOLD activity should be quantified using both summary measures and a voxelwise approach. EEG template magnitude should be correlated to a signature that is relevant to pain in adults, such as the Neurosynth signature. Next, ROI-based analyses should be carried out. Potential ROIs are the ACC, the thalamus, the putamen, the postcentral gyrus and the supramarginal gyrus. The ACC is the putative generator of adult noxious-evoked laser- and contact-heat potentials [106] and was one of the putative ROIs identified in Chapter 6. Noxious-evoked BOLD activity in the thalamus showed the strongest relationship to the adult PEPs Chapter 6, and has been reported to co-vary with noxious-evoked ERPs in a previous adult study [224]. The putamen was identified in the voxelwise adult and infant analyses. Noxious-evoked BOLD activity in the postcentral gyrus and supramarginal gyrus showed weak positive associations with EEG template magnitude. These areas have also previously been reported to be related to noxious-evoked ERPs in adults [224, 210]. To identify more subtle changes, a voxelwise analysis should also be carried out, as described in this chapter.

Resting state fMRI and white matter structural integrity are more stable measures than task-related BOLD [119] and could provide additional information about inter-individual sensitivity to noxious stimuli. Infant noxious-evoked BOLD responses are linked to patterns of resting state fMRI activity and white matter integrity [21] and in adults, measurements of resting state fMRI and white matter microstructure have been found to relate to individual pain scores or thresholds [237, 342, 108]. In a follow-up study, diffusion-weighted data and resting state fMRI could be collected to identify any relationships between underlying brain activity

and structure and EEG responses. First, resting-state summary measures and white matter microstructure measures that are related to noxious-evoked BOLD responses (developed by a colleague in my research group in [21]) should be extracted and correlated with EEG template magnitudes. Second, whole brain analyses can be carried out to investigate whether there are any other white matter tracts which are related to variability in EEG template magnitudes.

7.3.4 Summary and next steps

In this part of my thesis I presented a pilot study in adults and infants that aimed to provide a framework for the investigation of the fMRI correlates of variation in infant noxious-evoked EEG responses. The studies highlighted that neuroimaging data contain a large degree of variance, therefore future studies should employ a larger sample size and multiple stimuli across multiple test occasions. fMRI-EEG relationships should be investigated at the level of multivariate spatial patterns as well as single voxels. The results from follow-up studies will be useful in interpreting changes in the magnitude of the template of noxious-evoked EEG activity in analgesic studies.

Part IV

Discussion

8 General discussion

8.1 Thesis summary

Pain assessment is of vital importance in neonatal care. This thesis focused on improving the understanding of infants' multimodal responses to noxious stimuli, with an emphasis on those derived from the brain, where experiences are generated. In the first part of this thesis, I showed that infants' responses to noxious and non-noxious stimuli can be discriminated based on multimodal measurements, and that many of these measurements change with age. Specifically, I identified a previously unknown discriminative noxious-evoked ERP in infants aged 30–33 weeks PMA, which has the potential to be used in a similar way to the previously derived template of noxious-evoked EEG activity in term infants in the future. Interpreting amplitude changes in infant noxious-evoked ERPs is challenging, partly because their neurological basis is poorly understood. In the second part of this thesis, I investigated the fMRI correlates of inter-individual variation in noxious-evoked EEG responses in a pilot study. The pilot study highlighted that it was feasible to record noxious-evoked fMRI and EEG in the same infant, and provided recommendations for larger follow-up investigations.

The following paragraphs outline ways in which the work in this thesis can be improved or built upon. The discussion is divided into topics relating to

the implementation of research findings into clinical practice, and topics relating to future research.

8.2 Towards clinical tools

In Chapter 3 I presented a classification model that predicts exposure to a noxious stimulus based on multimodal measurements, and in Chapter 4 I presented a novel noxious-evoked ERP template (PC 2 noxious) which can be projected to new data to quantify noxious-evoked brain activity in preterm infants. Before new pain assessment tools based on these findings can be implemented in clinical practice, they would need to undergo rigorous evaluation to ensure their conformity with COSMIN's recommendations for validity, reliability and responsiveness. Initial ideas for technical improvements, validation studies and potential applications of multimodal pain assessment tools (including brain-derived measurements) are discussed in Chapter 3. The following paragraphs will discuss practical considerations around the responsiveness and clinical significance of changes in measurement scores, and an outlook on how multimodal pain assessment could fit in clinical practice.

8.2.1 Responsiveness and clinical significance

As previously described, responsiveness is "the ability of a tool to detect change over time in the construct to be measured" [227]. This is different to detecting differences in single scores across participants, which is an aspect of construct validity [65]. Responsiveness is an important measurement property, but it has not been established for the majority of existing infant pain assessment tools [213]. Responsiveness studies require the construct of interest to change within participant during the study period. For the construct of infant pain, responsiveness could thus in theory be assessed by comparing measurements between graded noxious events, or by comparing scores before and after a pain-reducing intervention [213]. In practice, however, these studies are difficult to design, because we must assume with a certain confidence that the construct of pain is really changing over time.

In Chapter 3 I compared responses to a heel lance with responses to a control heel lance. Qualitatively, the difference between these two events is quite extreme: one is a vibrotactile, non-skin breaking stimulus, while the other is a skin-piercing stimulus which is verbally reported as painful by adults when applied to their finger [88]. While it impossible to know how infants experience these stimuli, and thus whether the construct of pain really changes between the two, this is a model with relatively good face validity.

The disadvantage of this model is that a tool that was validated in this extreme situation (such as the classifier developed in Chapter 3) could be insensitive to smaller, but potentially clinically important differences in the construct of pain. If such a tool were used to assess analgesic efficacy in a clinical trial, analgesics with small to moderate effects could be unjustly dismissed as ineffective. Thus, there is a need for responsiveness models in which more subtle changes in the construct of pain are expected. How should these be derived? One possibility is to compare two or more noxious events that are perceived differently by adults. For ethical reasons, these events should either form part of infants' clinical care or should involve the mild pinprick stimuli which do not pierce the skin or cause behavioural distress. For instance, a heel lance (in adults applied to the finger) could be compared to the 128 mN pinprick stimulus or adhesive tape removal. To get an idea of how much the construct of pain may change between these two stimuli, the relative change in NRS between the two stimuli could be assessed in adults. Still, the translation of adult NRS scores to infants would be ambiguous. A second approach is to compare different clinical procedures by their expected severity in infants. Laudiano-Dray and colleagues [190] conducted a large-scale analysis of RCTs of noxious clinical procedures in newborns to estimate each procedure's pain severity on a single "pain reactivity" scale ranging from 0 to 100. The pain reactivity scores were derived from individual studies by normalising pain scores within pain scale. Five clusters of procedures were found, ranging from mild to extremely severe [190]. As infants in the NICU often need to undergo multiple procedures over the course of their stay, scores for a new tool could be collected for each procedure. Responsiveness

could then be determined by the degree to which a tool's scores change between the different classes of procedures. The main disadvantage of this approach is that the severity scale is based on existing non-optimal pain tools, and that therefore it is not necessarily representative of the underlying construct.

Yet another possibility is to evaluate changes in scores in response to the administration of analgesics [213]. This approach suffers from the same problem as above. To assess responsiveness to an analgesic, there must be a reasonable expectation that the analgesic changes the construct of pain. As analgesic trials in infants utilise suboptimal behavioural and physiological outcome measures, it cannot be ascertained that this is the case for any analgesic in neonates. For example, although in this thesis paracetamol was used to investigate whether the classification model was sensitive to modulation by an analgesic, a recent meta-analysis evaluating the efficacy of paracetamol concluded that *"The paucity and low quality of existing data do not provide sufficient evidence to establish the role of paracetamol in reducing the effects of painful procedures in neonates"* [243]. This leads to a paradoxical situation in which the validity of pain assessment tools is determined by their responsiveness to certain interventions, and the efficacy of those same interventions is determined by the degree to which they change the scores of those same pain assessment tools. Unfortunately, this problem is inherent to infant pain assessment, in which no gold standard is available. A pragmatic solution could be to select pharmacological agents that have proven efficacy for acute pain in children, or to only use pharmacological agents with a known mechanism of action, such as opioids or local anaesthetic, for responsiveness studies.

A concept associated with responsiveness is that of interpretability, which relates to the qualitative meaning that can be assigned to changes in measurements [227, 63]. Interpretability is in turn related to clinical relevance and the degree of change that would be considered important in a clinical setting [63]. The three types of studies described above could help with establishing interpretability and determining which changes in scores are likely to be clinically significant. Changes in scores could be described in terms of the procedures that they relate to - e.g.

a change of x points corresponds to the difference between a heel lance and a control procedure in a term infant.

Further discussions among specialists should take place to evaluate each of these options.

8.2.2 Implementing multimodal pain assessment in clinical practice

Even after extensive validation studies, multimodal pain assessment tools will not immediately find their way to the NICU. The acquisition and analysis of multimodal measurements, including noxious-evoked ERPs, requires access to specialised equipment, knowledge and skills [22], as well as time – in a current clinical trial using multimodal outcome measurements, each test occasion is expected to last approximately an hour [44]. It is likely that multimodal pain assessment tools will influence clinical practice only indirectly at first, by improving the outcome measures of clinical trials of analgesics. However, there may be a place for routine multimodal pain assessment on the NICU in the future. ECG is already routinely recorded in infants in the NICU, and wireless EEG equipment has now become available [254], as well as approaches to automatically decode facial expressions from video recordings [143]. If these signals could be automatically processed in real time, it would be possible to follow up individual infants over their stay in the NICU to assess whether they may be receiving insufficient pain relief for clinical procedures, with relatively little extra training for clinical staff. As with other types of monitoring, the judgment of the clinical staff caring for the infant would remain critical. Two pain monitoring devices, one based on heart rate variability [35] and one based on skin conductance [334], are already being tested on NICUs [263, 335] and are deemed useful by clinicians [336].

This thesis focused on procedural pain. Studies into multimodal assessment of prolonged pain are another important step to improve clinical pain assessment. These studies are conceptually challenging to design, as the absence of a stimulus makes it difficult to determine when the construct of chronic pain is likely to occur or change.

Studies which involve clinical procedures of longer duration, such as retinopathy of prematurity screening, or in post-surgical patients, could be a starting point.

8.3 A better understanding of noxious-evoked brain activity in the context of brain development

In parallel to developing clinically usable multimodal pain assessment tools, noxious-evoked brain responses should be further investigated using complementary neuroimaging methods. The transition from delta brush responses to discriminative ERPs is of particular interest. In Chapter 4 several theories were formulated about the neuronal developments underlying the morphological changes in noxious-evoked ERPs, which will be addressed here.

8.3.1 The role of the subplate in noxious-evoked ERP transitions

As described in Chapter 4, the timing of the morphological changes in noxious-evoked ERPs loosely parallels the formation of thalamocortical connections through the subplate. Delta brush responses were observed mostly before 30 weeks PMA, a time at which subplate neurons form a relay between the thalamus and the cortex [179]. The occurrence of spontaneous delta brushes has been linked to signalling in the subplate [236, 11]. PC 2 noxious (30–33 weeks PMA) was mostly present during the developmental period in which direct thalamocortical connections are being established [179] and the subplate is maximal in thickness [327]. Subplate resolution occurs roughly during the same time period as the emergence of PC 3 noxious [179]. Therefore, it would be of interest to directly investigate whether subplate thickness in somatosensory areas relates to noxious-evoked EEG responses. The subplate can be visualised *in vivo* using structural MRI [327, 179], diffusion-weighted MRI [79] and ultrasound [255]. In a future study, subplate thickness and noxious-evoked ERPs could be recorded at several time points in the same infant to understand whether changes in EEG responses follow the same trajectory as subplate development.

If the subplate were the source of the morphological changes in noxious-evoked ERPs, a similar morphological development should also be apparent in non-noxious tactile responses, as these are also transmitted through thalamocortical connections. In Chapter 4 I noted that the response to the vibrotactile control heel lance underwent morphological changes, albeit less clearly than the noxious-evoked ERP. In line with this, previous studies investigating tactile responses to the foot have also shown a decrease in the amplitude of the tactile-evoked N2 peak with PMA (latency approximately 160 ms, non-significant in [345], significant in [275]) and an increase in the amplitude of the tactile-evoked P2 peak (latency approximately 300 ms, [345]). Therefore, the study should also include tactile EEG responses to the control heel lance and non-vibratory tactile stimuli (e.g. a gentle touch with a tendon hammer).

A previous study in rats showed that ablation of subplate neurons within 36 hours after birth led to a disappearance of spontaneous spindle bursts [306], the rodent analogue of human delta brushes. Another study showed that noxious-evoked neural local field potentials change morphology in rat pups [42]. These studies could be further built upon by investigating whether noxious stimulation evokes spindle bursts in very young rat pups, and whether selective subplate ablation alters these and future noxious-evoked responses.

8.3.2 Anatomical source of developing noxious-evoked ERPs

As described in Chapter 1 and Part III, understanding which brain areas co-activate with noxious-evoked ERPs can aid the neuroscientific interpretation of changes in ERP amplitude. This is particularly interesting in the context of the gradual development from PC 2 noxious to PC 3 noxious-like EEG responses described in Chapter 4, as it would be of interest to understand whether very preterm infants recruit different brain areas in response to noxious stimulation compared to term infants.

The EEG-fMRI study proposed in Chapter 7 will provide information on the relationship between inter-individual differences in EEG responses and fMRI responses in term infants. This study does not however necessarily provide

information on the anatomical source of the noxious-evoked ERP. As previously described, simultaneous EEG-fMRI studies are difficult to conduct, and it is not possible to use fMRI to study responses to heel lances.

Functional ultrasound imaging (fUSI) is a new technique which could revolutionise the study of infant brain activity in a clinical setting [15, 70]. fUSI samples doppler ultrasound images at very high frequencies to create detailed images of cerebral blood flow [302, 67]. Similar to fMRI, blood flow is used as a proxy for neuronal activity [302, 67]. In animals, fUSI has been used to quantify cerebral responses to mechanical [313], visual [262, 202] and oculomotor tasks [74]. In human adults, fUSI has been used intra-operatively to map brain areas related to specific tasks [153]. The infant brain can be imaged through the fontanelles, and doing so has revealed that fUSI tracks sleep stages and seizure activity [69]. The major benefits of fUSI over fMRI is that it has higher temporal resolution and that it can be used at the cot-side, simultaneously with EEG [69]. Technical validation studies are still underway, but fUSI appears safe in term and preterm neonates [69]. When it has been further optimised to record cerebral activity in neonates, it could be used to record responses to clinical procedures in term and preterm neonates. This would allow a direct comparison of noxious-evoked ERP morphology and magnitude and the spatial extent of cerebral blood flow responses to clinical procedures.

8.3.3 The use of experimental noxious stimuli to study infant nociception

As described in Chapter 1, it is important to make a distinction between nociception (the encoding of noxious stimuli) and pain (the subjective multimodal experience). In adult studies of pain, experimental noxious stimuli are often of high intensity to ensure that they exceed an individual's subjective pain threshold (e.g. [312]). Neural responses can then be interpreted as pain-related signals. This practice would be unethical in infants. As a result, it is impossible to investigate signals representing the *experience* of pain using experimental noxious stimuli in infants. However, recording responses to mild experimental noxious stimuli which do not

cause the experience of pain (nor infant distress) can still yield valuable information about the developing pain *system*, as long as these stimuli probe nociceptive circuitry. Previous experiments are suggestive that the pinprick stimuli used in this thesis are effective at doing exactly this. For instance, Magerl and colleagues [203] used capsaicin pre-treatment of the skin in healthy adults to selectively block capsaicin-sensitive A- and C fibers on the hand, and found that a pricking sensation could still be elicited by 128 mN pinprick stimuli, albeit lower than before the blockade. After a complete A-fiber block, however, pinprick sensation was completely obliterated. These findings suggest that pinprick stimuli are transmitted through the class of capsaicin-insensitive A-fibers, which includes high-threshold mechanoreceptors and type I mechano-heat receptors [203]. Additionally, in a patient with a selective one-sided lesion of the spinothalamic tract (the main ascending nociceptive tract), 128 mN pinprick stimuli applied to the affected hand evoked a pinprick-evoked potential of approximately 50% lower amplitude than stimuli applied to the unaffected hand [147]. As stimulation of (tactile) Aβ-fibers evoked normal somatosensory-evoked potentials on both sides, these results suggest that pinprick-evoked signals are transmitted through nociceptive pathways [147]. Thus, while brain signals evoked by 128 mN pinprick stimuli should not be interpreted as pain signals, they are useful in quantifying the nociceptive input to the brain. Considering the negative effects of repeated noxious procedures on the developing brain [34] it is of great interest to investigate interventions that limit this input. Understanding nociceptive pathways through the use of mild, ethically acceptable experimental noxious stimuli may help to do so.

8.4 What does it feel like to be in pain?

In this thesis, I have focused on the construct of neonatal pain, its physiology and its measurement. I have provided evidence that even in very preterm infants, the nervous system may discriminate noxious from non-noxious events, and that noxious-evoked responses change from prematurity to term age to adulthood. Until now, I have explicitly steered clear of the question that is the most difficult to

answer, but perhaps also the most intriguing: for a newborn, what does it feel like to be in pain? Previously, I have argued that we will never be able to answer this question without verbal report. Yet, neuroimaging provides a new and unique way for us to make tentative inferences on infants' experiences, long before verbal communication is in place. As the field of neuroimaging develops, so will our understanding of the intricacies of noxious processing in the infant brain. In the meantime, we must focus on the fact that, regardless of our grasp on the exact infant pain experience, infants are worthy recipients of empathic and effective care [152], which necessitates robust pain assessment strategies. I hope that the work in this thesis has made a contribution to their development.

Part V

Appendix and bibliography

A

Appendix: Parent information leaflets and consent forms (infant studies)

DEPARTMENT OF PAEDIATRICS
Medical Sciences Division

Oxford University Hospitals NHS
NHS Foundation Trust



nipi
newborn infant pain investigations

Investigating pain in the developing human brain

PARENT INFORMATION LEAFLET

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 24 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.

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.

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. 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 clinical information about your child, for example their gestational age at birth and any medications they have received. 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 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, 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.

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
http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/

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 thestudy 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 Oxford Clinical Trials and Research Governance (CTRG) office (tel: 01865 (6)16480, email: ctrg@admin.ox.ac.uk).

12. Contact for further information

Prof Rebeccah Slater (Study Lead)
Professor of Paediatric Neuroscience
University of Oxford
01865 234537
rebeccah.slater@paediatrics.ox.ac.uk

Dr Eleri Adams (Clinical Lead)
Consultant Neonatologist
Oxford University Hospitals NHS Trust
01865 221356
eleri.adams@ouh.nhs.uk



Picture shows example of an EEG study.

Thank you for reading this information leaflet.

DEPARTMENT OF PAEDIATRICS
Medical Sciences Division

Oxford University Hospitals NHS
NHS Foundation Trust



nipi
newborn infant pain investigations

Investigating pain in the developing human brain

PARENT INFORMATION LEAFLET

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 and 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 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 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 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.

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 Dr 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 will 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 clinical information about your child, for example their gestational age at birth and any medications they have received. 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, 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.

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:
http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/

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 Clinical Trials and Research Governance (CTRG) office (01865 (6)16480, ctrg@admin.ox.ac.uk).

12. Contact for further information

Prof Rebeccah Slater (Study Lead)
Professor of Paediatric Neuroscience
University of Oxford
01865 234537
rebeccah.slater@paediatrics.ox.ac.uk

Dr Eleri Adams (Clinical Lead)
Consultant Neonatologist
Oxford University Hospitals NHS Trust
01865 221356
eleri.adams@ouh.nhs.uk



Picture shows example of an MRI study.

Thank you for reading this information leaflet.

Consent Form



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

Infant's name:

Study Title: Investigating pain in the developing human brain

Chief Investigator: Prof Rebeccah Slater (01865 234537, rebeccah.slater@paediatrics.ox.ac.uk)

Please initial each box

Please complete in black ballpoint pen.

1	I confirm that I have read and understood the information sheet (clinical procedures) (**v9.0, dated 31/05/2019)**, 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 consent to being approached in the future about other research studies that my child may be eligible for.	☐
9	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

Consent Form

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

Infant's name:

Study Title: Investigating pain in the developing human brain

Chief Investigator: Prof Rebeccah Slater (01865 234537, rebeccah.slater@paediatrics.ox.ac.uk)

Please initial each box

Please complete in black ballpoint pen.

1	I confirm that I have read and understood the information sheet (MRI) (**v9.0, dated 31/05/2019)**, 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 consent to being approached in the future about other research studies for which my child may be eligible.	☐
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

B

Appendix: Participant information leaflet and consent form (adult study)

Department of Paediatrics
Level 2, Children's Hospital
John Radcliffe
Headington, Oxford, OX3 9DU

Professor Rebeccah Slater
Direct Line: 01865 234 229
E-Mail: Rebeccah.slater@paediatrics.ox.ac.uk
neuroimaging.paediatrics.ox.ac.uk

Investigating pain in adults

PARTICIPANT INFORMATION SHEET

Ethics Approval Reference: R49359/RE002

We would like to invite you to take part in a research project. This sheet provides some information to help you decide whether to do so. Please take time to read this carefully and discuss it with friends, family or your GP if you wish. If there is anything that you do not understand, or if you would like more information, please ask us. Please take time to consider whether you wish to take part.

What is the purpose of the research?

The goal of pain research is to acquire new knowledge on the pathophysiology and treatment of acute and chronic pain. This requires research on humans and involves experimentally induced painful stimulation. In this study we are especially interested in how pain is processed in the brain. To do this, we measure brain activity using functional MRI (fMRI) and electroencephalography (EEG). We also look at how pain relates to the structure of the brain, using other MRI techniques, such as diffusion weighted imaging (DWI).

We want to compare the findings from this study with findings from another study investigating pain in infants. This will help us find out how the experience of pain changes during life.

Why have I been invited?

You have been invited to take part in this research because you are healthy, between 18 and 60 years of age, and speak fluent English. We will be recruiting up to 50 participants in this research. You cannot participate in this research if you are suffering from chronic pain or if you are pregnant. You can also not use pain medication in the 6 hours leading up to the study.

Do I have to take part?

No. It is up to you to decide if you want to take part in this research. We will describe the research, go through this information sheet with you, and answer any questions you may have. If you agree to take part, we will ask you to sign a consent form and will give you a copy for you to keep. However, you would still be free to withdraw at any time, without needing to give a reason. This would not affect legal rights you would receive. If you are a student at the University of Oxford or Oxford Brookes, there would be absolutely no academic penalty if you decide you do not want to take part, or if you decide to withdraw at any point.

What will happen to me if I take part?

The research involves two visits to the John Radcliffe Hospital in Oxford: the MRI visit and the EEG visit. Usually these visits will be on the same day, with a 30 minute break in between. In total, this will take a full morning or afternoon. Sometimes it may be possible to do the visits on separate days. In this case, each visit will take approximately 2 hours. It may also be possible to reverse the order of the visits. The researcher will discuss with you beforehand whether this applies to you. We would ask you to not take any pain relief medication in the 6 hours before the visit.

Before you can take part in the research, a researcher will contact you by phone or meet you to go over the information sheet, explain the procedures, and go through a Screening Form with you to check if it is safe for you to participate. If you agree, we would ask you to come to the John Radcliffe Hospital in Oxford for the first visit. At the start of your first visit one of our research team would meet you to answer any questions you may have. If you are happy to continue they will then ask you to sign a consent form. You will then proceed with the rest of the two visits, of which you can find a detailed description below.

<u>MRI visit</u>

The MRI visit takes place at the Wellcome Centre for Integrative Neuroimaging (WIN) at the John Radcliffe Hospital. During this visit you will undergo magnetic resonance imaging (MRI). Upon your arrival, we would answer any questions you may have and we would ask you to change into a "pyjama-style" top and trousers. A radiographer or scan operator would then go through a screening form with you once again, to make sure it's safe for you to participate. They would also provide you with you hearing protection. Afterwards, you would be asked to lie still on a table inside the MRI scanner. The research would involve having a series of magnetic resonance scans over a period of 60 minutes. During the scan, the researcher will stimulate your foot with a sharp touch device called a Pinprick. They will use up to 5 different Pinpricks with different forces and apply up to 30 stimulations with each probe. These probes induce the sensation of mechanical pain, but they do not pierce the skin. You will also be wearing a respiration belt to measure your breathing rate and a finger clip to monitor your blood flow. These are standard instruments to measure physiological parameters in the scanner. After the scan, you will change back to your own clothes. The research team can direct an accompanying person to an area where they can wait. Please let us know beforehand if you wear contact lenses or glasses.

<u>EEG visit</u>

The EEG visit takes place at the Women's Centre at the John Radcliffe Hospital. During the visit, you will undergo electroencephalography (EEG). The researcher will answer any questions you may have. They will then put a set of sensors made up of conductive material (electrodes) on your head. To establish electrical contact between the scalp and the sensors, paste containing salts that conduct electricity would be placed under each metal contact. In order to achieve a good connection, it is often necessary to prepare the area of the scalp under the sensor by cleaning it with rubbing alcohol and massaging an abrasive substance using a cotton swab. We would ask you to let the researcher know if at any time the procedure becomes uncomfortable. In such a case, we would terminate the study procedure, without this having any negative consequences for you. The paste used to make the electrical contact is water based and washes away easily. After applying the sensors, the researcher will stimulate your foot with sharp touches, as described above for the MRI visit. They will use up to 5 different probes with different forces and apply up to 30 stimulations with each probe. To determine the exact moment at which the probe touches your skin, the researcher will either video the probe and your foot or use probes that are directly connected to the EEG recording. The researcher will ask you to rate your pain and will ask you to fill out a questionnaire about your experience after the stimulations. Afterwards, the researcher will remove the electrodes and rinse your head with water and cotton wool. You can remove any residue with water.

Are there any risks in taking part in this research?

<u>Stimuli</u>

All stimuli that are used to experimentally induce pain have been safely used as part of our pain research programme in ethically approved projects. The stimulation intensity does not exceed the individual tolerance level. Participants are always able to terminate a painful stimulus at will.

MRI

MRI is safe and non-invasive and does not involve any ionizing radiation (x-rays). However, because it uses a large magnet to work, MRI scans are not suitable for everybody. Because of this, you will be asked pre-screening safety questions to help determine if you are able to take part. Normally, MRI scanning for research purposes would not be performed without further investigation if you have a heart pacemaker, mechanical heart valve, mechanical implant such as an aneurysm clip, hip replacement, or if you carry other pieces of metal that have accidentally entered your body.

While there is no evidence to suggest that MRI is harmful to unborn babies, as a precaution, the Department of Health advises against scanning pregnant women unless there is a clinical benefit. We do not test for pregnancy as routine so if you think you may be pregnant you should not take part in this research.

If you think you might be claustrophobic, please discuss this in advance with the researcher, or let the radiographer or operator know before your scan.

As some of the scans are noisy, we would give you earplugs to make this quieter for you. It is important that these are fitted correctly as they are designed to protect your hearing.

In preparation for your scan and for your comfort and safety we will ask you to change into pocketless and metal free "pyjama-style" top and trousers, which are available in a range of sizes. You may keep your underwear and socks on but we would ask ladies to remove underwired bras, if you have a suitable non-wired bra you may wear this instead. Please avoid any fabrics that contain metallic threads or have been silver impregnated (often marketed as anti-microbial/bacterial or anti-odour/stink). Metal jewellery including body piercing must also be removed. Eye shadow and mascara must also be avoided, since some types contain materials that can interact with the magnetic field. If you wish to wear eye makeup to your scan we can provide makeup removal wipes but you are advised to bring your own makeup to reapply. Lockers are provided to secure your personal belongings and clothing.

You will be introduced carefully to the scanner and allowed to leave at any stage. Whilst in the scanner you will have easy access to a call button should you wish to stop the scan or speak with the radiographer or operator.

It is important to note that we do not carry out scans for diagnostic purposes, only for research. Our scans are not routinely looked at by a doctor and are therefore not a substitute for a doctor's appointment. Occasionally, however, a possible abnormality may be detected. In this case, we would have the scan checked by a doctor. If the doctor felt that the abnormality was medically important, you would be contacted directly and recommended to have a hospital (NHS) diagnostic scan arranged. You would not be informed unless the doctor considers the finding has clear implications for your current or future health. All information about you is kept strictly confidential.

Electroencephalography (EEG)

EEG is a procedure for measuring brain waves. It is harmless and painless and carries no significant risk to participants, though some participants do report a brief itching sensation at the beginning of the recording. EEG recording has been used safely for many years, and we are not aware of any cases of adverse events. EEG equipment comes from certified suppliers, who are obliged by law to adhere to published guidelines on electrical and mechanical safety (IEC-601). If you feel any discomfort, then please let the researcher know and they will stop the procedure.

It is important to note that EEG is not carried out for diagnostic purposes, only for research. If you think you may have an undiagnosed medical condition affecting your brain then you should consult your GP. You should not take part in this study.

Are there any benefits from taking part in this research?

No. There will be no direct benefit to you from taking part in this research. It is hoped that the results from this research will help us to identify better measures for future studies in patients with acute and chronic pain, including young infants admitted to hospitals.

Will my time/travel costs be reimbursed?

Your time will be compensated with a £25 gift voucher per completed visit.

What will happen to the data provided?

The information you provide during the study is the **research data**. Any research data from which you can be identified (name, date of birth, contact details) is known as **personal data**.

Personal / sensitive data will be stored in a locked filing cabinet on University premises or on encrypted computers or encrypted hard drives. Your contact details for the current study will be stored for the duration of the research. With your consent, we will keep your contact details on a secure database or in a locked filing cabinet on University premises in order to contact you for future studies.

Other research data (including consent forms) will be stored for at least 3 years after publication or public release of the work of the research. All of your data will be given a unique number. The key linking this number to your personal details will be kept in a locked filing cabinet on site and will be stored for the duration of the research.

The researcher and the research team (including the researcher's supervisor) will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

We may use your anonymised data in future studies, and share data with other researchers (e.g. in online databases). All personal information that could identify you will be removed or changed before information is shared with other researchers or results are made public.

Will the research be published?

The research may be published in academic journals and disseminated in presentations.

The research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be openly accessible.

Who has reviewed this research?

All research studies are checked by an ethics committee to ensure the research is conducted safely and to the best standards. This research has been reviewed by, and received favourable opinion from, the University of Oxford Central University Research Ethics Committee.

Who is organising and funding the research?

This research is organised by the Department of Paediatrics at the University of Oxford and has been funded by The Wellcome Trust.

Who do I contact if I have a concern about the research or I wish to complain?

If you have a concern about any aspect of this research, please contact Marianne van der Vaart (01865 226824, marianne.vandervaart@paediatrics.ox.ac.uk) or Rebeccah Slater (01865 234229, rebeccah.slater@paediatrics.ox.ac.uk), and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Medical Sciences Interdivisional Research Ethics Committee (MS IDREC) at the University of Oxford who will seek to resolve the matter as soon as possible - Email: ethics@medsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD.

Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research.

The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest.

Further information about your rights with respect to your personal data is available from http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/.

Further Information and Contact Details

If you would like to discuss the research with someone beforehand, or if you have any questions afterwards, please contact Marianne van der Vaart on 01865 226824 or marianne.vandervaart@paediatrics.ox.ac.uk.

Department of Paediatrics
Level 2, Children's Hospital
John Radcliffe
Headington, Oxford, OX3 9DU

Professor Rebeccah Slater
Direct Line: 01865 234 229
E-Mail: Rebeccah.slater@paediatrics.ox.ac.uk
Marianne van der Vaart, DPhil Student
Direct line: 01865 226824
E-Mail: marianne.vandervaart@paediatrics.ox.ac.uk

PARTICIPANT CONSENT FORM

CUREC Approval Reference: R49359/RE002

Investigating pain in adults

Purpose of Study: to investigate how adults respond to a painful stimulus.

		Please initial each box
1	I confirm that I have read and understand the information sheet dated _____ (Version ____) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without penalty or affecting my legal rights.	☐
3	I understand that research data collected during the study, may be looked at by designated individuals from the University of Oxford where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data.	☐
4	I agree for research data collected in this study to be given to researchers, including those working outside of the EU, to be used in other research studies. I give permission for data from this study to be used in publication. I understand that any data and/or brain images of me that leave the Centre will be fully anonymised so that I cannot be identified.	☐
5	I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	☐
6	I understand how to raise a concern or make a complaint	☐
7	I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.	☐
8	I understand that the EEG scan is not useful for medical diagnosis	☐
9	I understand that the MRI scan is a research scan that 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 scan, I will be informed if a doctor thinks it is medically important. I understand that I would not be informed if the doctor does not think the finding has clear implications for my current or future health.	☐

10	I agree to take part in the above study.	☐
Optional:	I agree for my personal data to be kept in a secure database for the purpose of contacting me about future studies.	☐

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature

C

Appendix: Publications, funding and awards

C.1 Publications relevant to this thesis

Vaart, Marianne van der, Caroline Hartley, Luke Baxter, Gabriela Schmidt Mellado, Foteini Andritsou, Maria M Cobo, Ria Evans Fry, Eleri Adams, Sean Fitzgibbon, and Rebeccah Slater. 'Premature Infants Display Discriminable Behavioral, Physiological, and Brain Responses to Noxious and Nonnoxious Stimuli'. Cerebral Cortex, 2021, 1–17. *This published work was restructured and expanded for Chapter 3 and Chapter 4.*

C.2 Other publications

Vaart, Marianne van der, Eugene Duff, Nader Raafat, Richard Rogers, Caroline Hartley, and Rebeccah Slater. 'Multimodal Pain Assessment Improves Discrimination between Noxious and Non-noxious Stimuli in Infants'. Paediatric and Neonatal Pain 1, no. 1 (2019): 21–30.

Baxter, Luke, Ravi Poorun, Richard Rogers, **Marianne van der Vaart**, Alan Worley, and Caroline Hartley. 'Using Changes in Brain Activity to Assess Pain-

Relief in Infants: Methodological Considerations with Benoit et al. (2021)'. Early Human Development 157 (2021): 105361.

Cobo, Maria M., Caroline Hartley, Deniz Gursul, Foteini Andritsou, **Marianne van der Vaart**, Gabriela Schmidt Mellado, Luke Baxter, et al. 'Quantifying Noxious-Evoked Baseline Sensitivity in Neonates to Optimise Analgesic Trials'. ELife 10 (2021): e65266.

Cobo, Maria M, Gabrielle Green, Foteini Andritsou, Luke Baxter, Ria Evans Fry, Annika Grabbe, Deniz Gursul, Amy Hoskin, Gabriela Schmidt Mellado, **Marianne van der Vaart** et al. 'Early Life Inflammation Increases Spinal Cord Excitability and Nociceptive Sensitivity in Human Infants'. Nature Communications 13 (2022): 3943.

Duff, Eugene P., Fiona Moultrie, **Marianne van der Vaart**, Sezgi Goksan, Alexandra Abos, Sean P. Fitzgibbon, Luke Baxter, Tor D. Wager, and Rebeccah Slater. 'Inferring Pain Experience in Infants Using Quantitative Whole-Brain Functional MRI Signatures: A Cross-Sectional, Observational Study'. The Lancet Digital Health 2 (2020): e458–67.

Hartley, Caroline, Fiona Moultrie, Amy Hoskin, Gabrielle Green, Vaneesha Monk, Jenifer Bell, Andrew King, Miranda Buckle, **Marianne van der Vaart** et al. 'Analgesic Efficacy and Safety of Morphine in the Procedural Pain in Premature Infants (Poppi) Study: Randomised Placebo-Controlled Trial'. The Lancet 392 (2018): 2595–2605.

C.3 Funding and awards

Ketel1 Studiefonds Scholarship (€2.000, DPhil stipend) **2019**
Prins Bernhard Cultuurfondsbeurs Scholarship (€40.000, DPhil stipend) **2018**

University of Oxford Paediatric Department Academic Scholarship (≈ £22.000, covering DPhil tuition fees for 3 years) **2018**

Hendrik Muller Vaderlandsch fonds Scholarship (€5.000, DPhil stipend) **2018**

Bibliography

[1] Rodolfo Abreu, Alberto Leal, and Patrícia Figueiredo. EEG-Informed fMRI: A Review of Data Analysis Methods. *Frontiers in Human Neuroscience*, 12(February):1–23, 2018. 5.4

[2] Micah Allen, Davide Poggiali, Kirstie Whitaker, Tom Rhys Marshall, Jordy van Langen, and Rogier A. Kievit. Raincloud plots: A multi-platform tool for robust data visualization [version 2; peer review: 2 approved]. *Wellcome Open Research*, 4(63), 2021. 2.4

[3] Jesper L. R. Andersson, Mark S. Graham, Ivana Drobnjak, Hui Zhang, and Jon Campbell. Susceptibility-induced distortion that varies due to motion: Correction in diffusion MR without acquiring additional data. *NeuroImage*, 171:277–295, May 2018. 7.1.4.1

[4] Jesper L. R. Andersson, Mark S. Graham, Ivana Drobnjak, Hui Zhang, Nicola Filippini, and Matteo Bastiani. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *NeuroImage*, 152:450–466, May 2017. 7.1.4.1

[5] Jesper L. R Andersson, Mark S. Graham, Enikő Zsoldos, and Stamatios N. Sotiropoulos. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, 141:556–572, November 2016. 7.1.4.1

[6] Jesper L. R. Andersson and Stamatios N. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125:1063–1078, January 2016. 7.1.4.1

[7] M Andre, MD Lamblin, AM 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. *Neurophysiol Clin*, 40(2):59–124, 2010. 1.7.2, 1.9.1, 3, 4.1

[8] A. Vania Apkarian, M. Catherine Bushnell, Rolf-Detlef Treede, and Jon-Kar Zubieta. Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain*, 9(4):463–484, 2005. 1.6.1, 6.3

[9] Georgette Argiris, Yaakov Stern, and Christian Habeck. Quantifying Age-Related Changes in Brain and Behavior: A Longitudinal versus Cross-Sectional Approach. *eNeuro*, 8(4):0273–21, 2021. 4.4.5

[10] Tomoki Arichi, Serena J. Counsell, Alessandro Allievi, Marta Varela, Christian F. Beckmann, Etienne Burdet, A. David Edwards, Gianlorenzo Fagiolo, Nazakat Merchant, Nora Tusor, and Alejandro Melendez-Calderon. Development of BOLD signal hemodynamic responses in the human brain. *NeuroImage*, 63(2):663–673, 2012. 1.9.2.1, 7.1.4.2, 7.3.1

[11] Tomoki Arichi, Kimberley Whitehead, Giovanni Barone, Ronit Pressler, Francesco Padormo, A. David Edwards, and Lorenzo Fabrizi. Localization of spontaneous bursting neuronal activity in the preterm human brain with simultaneous EEG-fMRI. *eLife*, 6:e27814, 2017. 1.7.2, 4.4.1, 4.4.5, 5.4, 8.3.1

[12] Owen J. Arthurs and Simon Boniface. How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences*, 25(1):27–31, January 2002. 1.9.2.1

[13] Marianne Aspbury, Luke Baxter, Ravi Poorun, Marianne van der Vaart, Caroline Hartley, Maria M Cobo, Annalisa Hauck, Kirubin Pillay, Sean P Fitzgibbon, Aomesh Bhatt, and Rebeccah Slater. Establishing a standardised approach for the measurement of neonatal noxious-evoked brain activity (in preparation). *Submitted to Cortex as a pre-registered report*, 2022. 4.4.4, 5.1

[14] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, February 2008. 7.1.4.1

[15] Jerome Baranger, Charlie Demene, Alice Frerot, Flora Faure, Catherine Delanoë, Alexandre Houdouin, Jerome Mairesse, Valerie Biran, and Olivier Baud. Bedside functional monitoring of the dynamic brain connectivity in human neonates. *Nature Communications*, 12:1080, 2021. 8.3.2

[16] Marco Bartocci, Lena L. Bergqvist, Hugo Lagercrantz, and K. J.S. Anand. Pain activates cortical areas in the preterm newborn brain. *Pain*, 122:109–117, 2006. 1.9, 3.2.3

[17] Allan I. Basbaum, Diana M. Bautista, Grégory Scherrer, and David Julius. Cellular and Molecular Mechanisms of Pain. *Cell*, 139(2):267–284, October 2009. 1.6.1

[18] Hélène Bastuji, Maud Frot, Caroline Perchet, Michel Magnin, and Luis Garcia-Larrea. Pain networks from the inside: Spatiotemporal analysis of brain responses leading from nociception to conscious perception. *Human Brain Mapping*, 37(12):4301–4315, 2016. 1.6.4

[19] Luke Baxter. *Neuroimaging of Neonatal Nociception*. PhD thesis, University of Oxford, Oxford, January 2020. 1.4

[20] Luke Baxter, Sean Fitzgibbon, Fiona Moultrie, Sezgi Goksan, Mark Jenkinson, Stephen Smith, Jesper Andersson, Eugene Duff, and Rebeccah 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. 1.9.2, 7.1.4.1, 7.1.4.2, 7.3.1, 7.3.3.1, 7.3.3.3

[21] Luke Baxter, Fiona Moultrie, Sean Fitzgibbon, Marianne Aspbury, Roshni Mansfield, Matteo Bastiani, Richard Rogers, Saad Jbabdi, Eugene Duff, and Rebeccah Slater. Functional and diffusion MRI reveal the neurophysiological basis of neonates' noxious-stimulus evoked brain activity. *Nature Communications*, 12(1):2744, December 2021. 1.9.2.3, 3.1, 5.5, 7.1.4.1, 7.3.3.1, 7.3.3.3

[22] Luke Baxter, Ravi Poorun, Richard Rogers, Marianne van der Vaart, Alan Worley, and Caroline Hartley. Using changes in brain activity to assess pain-relief in infants: Methodological considerations with Benoit et al. (2021). *Early Human Development*, 157:105361, 2021. 8.2.2

[23] Christian F Beckmann and Stephen M Smith. Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*, 23(2):137–152, 2004. 2.3.2, 6.1.6, 7.1.4.1

[24] Timothy Behrens, Stephen Smith, Matthew Webster, and Thomas Nichols. Randomise/UserGuide - FslWiki. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise/UserGuide. 6.1.7

[25] Florian Beissner, Amadeus Brandau, Christian Henke, Lisa Felden, Ulf Baumgärtner, Rolf-Detlef Treede, Bruno G. Oertel, and Jörn Lötsch. Quick Discrimination of Adelta and C Fiber Mediated Pain Based on Three Verbal Descriptors. *PLOS ONE*, 5(9):e12944, September 2010. 1.6.2

[26] Yehezkel Ben-Ari, Jean-Luc Gaiarsa, Roman Tyzio, and Rustem Khazipov. GABA: A Pioneer Transmitter That Excites Immature Neurons and Generates Primitive Oscillations. *Physiological Reviews*, 87(4):1215–1284, October 2007. 4.4.1

[27] Britney Benoit, Aaron Newman, Ruth Martin-Misener, Margot Latimer, and Marsha 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, March 2021. 4.4.4

[28] Denny Borsboom, Gideon J. Mellenbergh, and Jaap van Heerden. The Concept of Validity. *Psychological Review*, 111(4):1061–1071, 2004. 1.5

[29] David Borsook, Jaymin Upadhyay, Eric H. Chudler, and Lino Becerra. A key role of the basal ganglia in pain and analgesia - insights gained through human functional imaging. *Molecular Pain*, 6(27), 2010. 6.3.1

[30] Megan A Boudewyn, Steven J Luck, Jaclyn L Farrens, and Emily S. Kappenman. How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, 55(6):e13049, 2018. 3.1

[31] Véronique Boudon-Millot. Must We Suffer in Order to Stay Healthy? Pleasure and Pain in Ancient Medical Literature. *Pain and Pleasure in Classical Times*, pages 36–54, August 2018. 1.2

[32] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. 3.2.6.1

[33] Leo Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Chapman&Hall/CRC, Boca Raton, Florida, 1984. 3.2.6.1

[34] Chelsie L. Brewer and Mark L. Baccei. The development of pain circuits and unique effects of neonatal injury. *Journal of Neural Transmission*, 127(4):467–479, 2020. 1.7.1, 8.3.3

[35] L. Butruille, J. De jonckheere, R. Marcilly, C. Boog, S. Bras da Costa, T. Rakza, L. Storme, and R. Logier. Development of a pain monitoring device focused on newborn infant applications: The NeoDoloris project. *IRBM*, 36(2):80–85, March 2015. 3.4.3, 8.2.2

[36] Rob Campbell. Raacampbell/shadedErrorBar. https://github.com/raacampbell/shadedErrorBar, 2021. 2.4

[37] Marsha Campbell-Yeo, Mats Eriksson, and Britney Benoit. Assessment and Management of Pain in Preterm Infants: A Practice Update. *Children*, 9(2):244, February 2022. 3.4.3

[38] 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 Breart. Epidemiology and treatment of painful procedures in neonates in intensive care units. *JAMA*, 300(1):60–70, 2008. 1.3, 3.1

[39] Ricardo Carbajal, Mats Eriksson, Emilie Courtois, Elaine Boyle, Alejandro Avila-Alvarez, Randi Dovland Andersen, Kosmas Sarafidis, Tarja Polkki, Cristina Matos, Paola Lago, Thalia Papadouri, Simon Attard Montalto, Mari-Liis Ilmoja, Sinno Simons, Rasa Tameliene, Bart van Overmeire, Angelika Berger, Anna Dobrzanska, Michael Schroth, Lena Bergqvist, Hugo Lagercrantz, Kanwaljeet J. S. Anand, and EUROPAIN Survey Working Group. Sedation and analgesia practices in neonatal intensive care units (EUROPAIN): Results from a prospective cohort study. *The Lancet. Respiratory Medicine*, 3(10):796–812, October 2015. 1.3

[40] Francesco Cerritelli, Martin G. Frasch, Marta C. Antonelli, Chiara Viglione, Stefano Vecchi, Marco Chiera, and Andrea Manzotti. A Review on the Vagus Nerve and Autonomic Nervous System During Fetal Development: Searching for Critical Windows. *Frontiers in Neuroscience*, 15:721605, 2021. 1.8.2

[41] Chris Chambers. Could a brain scan diagnose you as a psychopath? *The Guardian*, November 2013. 1.9.2.3

[42] Pishan Chang, Lorenzo Fabrizi, and Maria Fitzgerald. Distinct age-dependent C fiber-driven oscillatory activity in the rat somatosensory cortex. *eNeuro*, 7(5):1–11, 2020. 3.1, 8.3.1

[43] Christoph Christmann, Caroline Koeppe, Dieter F. Braus, Matthias Ruf, and Herta Flor. A simultaneous EEG-fMRI study of painful electric stimulation. *NeuroImage*, 34(4):1428–1437, 2007. 5.4

[44] Maria Cobo, Fiona Moultrie, Annalisa Hauck, Daniel Crankshaw, Vaneesha Monk, Caroline Hartley, Ria Evans Fry, Shellie Robinson, Marianne van der Vaart, Luke Baxter, Eleri Adams, Ravi Poorun, Aomesh Bhatt, and Rebeccah Slater. Study protocol: A multicentre, randomised controlled trial to investigate the effects of parental touch on relieving acute procedural pain in neonates (Petal) (in preparation). *Submitted to BMJ Open*, 2022. 1.9.1.3, 2.4, 3.1, 4.1, 4.4.2, 8.2.2

[45] Maria M Cobo, Gabrielle Green, Foteini Andritsou, Luke Baxter, Evans Fry, Annika Grabbe, Deniz Gursul, Amy Hoskin, Gabriela Mellado, Marianne van der Vaart, Eleri Adams, Aomesh Bhatt, Franziska Denk, Caroline Hartley, and Rebeccah Slater. Early life inflammation increases spinal cord excitability and nociceptive sensitivity in human infants. *Nature Communications*, 13:3943, 2022. 1.9.1.3, 2.1.1

[46] Maria M. Cobo, Caroline Hartley, Deniz Gursul, Foteini Andritsou, Marianne van der Vaart, Gabriela Schmidt Mellado, Luke Baxter, Eugene P. Duff, Miranda Buckle, Ria Evans Fry, Gabrielle Green, Amy Hoskin, Richard Rogers, Eleri Adams, Fiona Moultrie, Rebeccah Slater, Ria Evans Fry, Gabrielle Green, Amy Hoskin, Richard Rogers, Eleri Adams, Fiona Moultrie, and Rebeccah Slater. Quantifying noxious-evoked baseline sensitivity in neonates to optimise analgesic trials. *eLife*, 10:e65266, 2021. 1.9.1.3, 2.1.1, 2.1.2, 3.2.2, 5.1, 5.4, 5.5, 7.1.3.2, 7.3.1

[47] Robert C. Coghill. The Distributed Nociceptive System: A Framework for Understanding Pain. *Trends in Neurosciences*, 43(10):780–794, 2020. 1.6.4

[48] Robert C Coghill, John G Mchaffie, and Y Yen. Neural correlates of interindividual differences in the subjective experience of pain. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14):8538–8542, 2003. 5.5

[49] Matthew Colonnese and Rustem Khazipov. Spontaneous activity in developing sensory circuits: Implications for resting state fMRI. *NeuroImage*, 62(4):2212–2221, 2012. 1.7.2

[50] Committee on Fetus and Newborn. Age Terminology During the Perinatal Period. *Pediatrics*, 114(5):1362–1364, November 2004. (document)

[51] Xiaomei Cong, Jacqueline M McGrath, Regina M Cusson, and Di Zhang. Pain assessment and measurement in neonates. *Advances in neonatal care*, 13(6):379–95, 2013. 1.4, 1.8.1, 3.1

[52] Anni Copeland, Eero Silver, Riikka Korja, Satu J. Lehtola, Harri Merisaari, Ekaterina Saukko, Susanne Sinisalo, Jani Saunavaara, Tuire Lähdesmäki, Riitta Parkkola, Saara Nolvi, Linnea Karlsson, Hasse Karlsson, and Jetro J. Tuulari. Infant and Child MRI: A Review of Scanning Procedures. *Frontiers in Neuroscience*, 15:666020, 2021. 1.9.2

[53] Laura Cornelissen, Lorenzo Fabrizi, Deborah Patten, Alan Worley, Judith Meek, Stewart Boyd, Rebeccah Slater, and Maria Fitzgerald. Postnatal Temporal, Spatial and Modality Tuning of Nociceptive Cutaneous Flexion Reflexes in Human Infants. *PLoS ONE*, 8(10):e76470, 2013. 1.7.1, 1.8.1, 3.1, 3.2.3, 4.1, 4.4.3

[54] P. Cortelli, G. Giannini, V. Favoni, S. Cevoli, and G. Pierangeli. Nociception and autonomic nervous system. *Neurological Sciences*, 34(Suppl 1):S41–S46, 2013. 1.8.2

[55] Haline Tereza Matias de Lima Costa, Tatiana Xavier Costa, Rand Randall Martins, and Antônio Gouveia Oliveira. Use of off-label and unlicensed medicines in neonatal intensive care. *PLoS ONE*, 13(9):e0204427, September 2018. 1.3

[56] Kenneth D. Craig, Michael F. Whitfield, Ruth V.E. Grunau, Julie Linton, and Heather D. Hadjistavropoulos. Pain in the preterm neonate: Behavioural and physiological indices. *Pain*, 52(3):287–299, 1993. 1.4, 3.1, 3.2.3, 3.4.3, 4.1

[57] S Dall'Orso, J Steinweg, A G Allievi, A D Edwards, E Burdet, and T Arichi. Somatotopic Mapping of the Developing Sensorimotor Cortex in the Preterm Human Brain. *Cerebral Cortex*, 28(7):2507–2515, 2018. 1.7.3

[58] Steve Davidson, Hai Truong, and Glenn J Giesler Jr. A quantitative analysis of spinothalamic tract neurons in adult and developing mouse. *Journal of Comparative Neurology*, 518(16):3193–3204, 2010. 1.7.2

[59] Karen D. Davis, Herta Flor, Henry T. Greely, Gian Domenico Iannetti, Sean MacKey, Markus Ploner, Amanda Pustilnik, Irene Tracey, Rolf Detlef Treede, and Tor D. Wager. Brain imaging tests for chronic pain: Medical, legal and ethical issues and recommendations. *Nature Reviews Neurology*, 13(10):624–638, 2017. 5.2

[60] Josepheen De Asis-Cruz, Kushal Kapse, Sudeepta K. Basu, Mariam Said, Dustin Scheinost, Jonathan Murnick, Taeun Chang, Adre du Plessis, and Catherine Limperopoulos. Functional brain connectivity in ex utero premature infants compared to in utero fetuses. *NeuroImage*, 219:117043, October 2020. 4.4.5

[61] Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. Concepts, theories and models, and types of measurements. In *Measurement in Medicine: A Practical Guide*, pages 7–29. Cambridge University Press, Cambridge, 2011. 1.5

[62] Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. Development of a measurement instrument. In *Measurement in Medicine: A Practical Guide*, pages 30–64. Cambridge University Press, Cambridge, 2011. 1.5

[63] Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. Interpretability. In *Measurement in Medicine: A Practical Guide*, pages 227–274. Cambridge University Press, Cambridge, 2011. 3.4.4.2, 5.2, 8.2.1

[64] Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. Reliability. In *Measurement in Medicine: A Practical Guide*, pages 96–149. Cambridge University Press, Cambridge, 2011. 3.4.4.2, 3.4.4.2

[65] Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. Responsiveness. In *Measurement in Medicine: A Practical Guide*, pages 202–226. Cambridge University Press, Cambridge, 2011. 3.4.4.2, 3.4.4.2, 8.2.1

[66] Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. Validity. In *Measurement in Medicine: A Practical Guide*, pages 150–201. Cambridge University Press, Cambridge, 2011. 1.5, 3.4.4.2, 5.2

[67] Thomas Deffieux, Charlie Demené, and Mickael Tanter. Functional Ultrasound Imaging: A New Imaging Modality for Neuroscience. *Neuroscience*, 474:110–121, October 2021. 8.3.2

[68] Arnaud Delorme and Scott Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21, 2004. 1.9.1.4, 2.3.1, 2.4, 3.2.5.1, 3.2.5.4, 3.2.5.5, 3.2.5.6, 6.1.4, 6.1.5.1

[69] Charlie Demené, Jérome Baranger, Miguel Bernal, Catherine Delanoe, Stéphane Auvin, Valérie Biran, Marianne Alison, Jérome Mairesse, Elisabeth Harribaud, Mathieu Pernot, Mickael Tanter, and Olivier Baud. Functional ultrasound imaging of brain activity in human newborns. *Science Translation Medicine*, 9:eaah6756, 2017. 8.3.2

[70] Charlie Demené, Jérôme Mairesse, Jérôme Baranger, Mickael Tanter, and Olivier Baud. Ultrafast Doppler for neonatal brain imaging. *NeuroImage*, 185:851–856, 2018. 8.3.2

[71] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, July 2006. 6.1.7.2, 7.1.6.3

[72] Bianca U. Devsam and Sharon Kinney. The clinical utility of the pain assessment tool in ventilated, sedated, and muscle-relaxed neonates. *Australian Critical Care*, 34(4):333–339, July 2021. 3.4.3

[73] Miranda G DiLorenzo, Rebecca Pillai Riddell, David B Flora, and Kenneth D Craig. Infant Clinical Pain Assessment: Core Behavioral Cues. *Journal of Pain*, 19(9):1024–1032, 2018. 3.1

[74] Alexandre Dizeux, Marc Gesnik, Harry Ahnine, Kevin Blaize, Fabrice Arcizet, Serge Picaud, José Alain Sahel, Thomas Deffieux, Pierre Pouget, and Mickael Tanter. Functional ultrasound imaging of the brain reveals propagation of task-related brain activity in behaving primates. *Nature Communications*, 10:1400, 2019. 8.3.2

[75] D'Mello, R and Dickenson, A. H. Spinal cord mechanisms of pain. *British Journal of Anaesthesia*, 101(1):8–16, July 2008. 1.6.3

[76] Sam M. Doesburg, Cecil M. Chau, Teresa P.L. Cheung, Alexander Moiseev, Urs Ribary, Anthony T. Herdman, Steven P. Miller, Ivan L. Cepeda, Anne Synnes, and Ruth E. Grunau. Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. *Pain*, 154(10):1946–1952, 2013. 1.3

[77] 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. 1.7.2, 3.1, 4.1

[78] Adrienne E. Dubin and Ardem Patapoutian. Nociceptors: The sensors of the pain pathway. *Journal of Clinical Investigation*, 120(11):3760–3772, 2010. 1.6.2, 1.6.3

[79] Jeroen Dudink, Jan Buijs, Paul Govaert, Arjen L. Van Zwol, Nikk Conneman, Johannes B. Van Goudoever, and Maarten Lequin. Diffusion tensor imaging of the cortical plate and subplate in very-low-birth-weight infants. *Pediatric Radiology*, 40(8):1397–1404, 2010. 8.3.1

[80] Eugene P. Duff, Fiona Moultrie, Marianne van der Vaart, Sezgi Goksan, Alexandra Abos, Sean P. Fitzgibbon, Luke Baxter, Tor D. Wager, and Rebeccah Slater. Inferring pain experience in infants using quantitative whole-brain functional MRI signatures: A cross-sectional, observational study. *The Lancet Digital Health*, 2:e458–e467, 2020. 1.7.2, 1.9.2.3, 4.4.5, 5.6, 2, 4

[81] Richard P Dum, David J Levinthal, and Peter L Strick. The spinothalamic system targets motor and sensory areas in the cerebral cortex of monkeys. *The Journal of Neuroscience*, 29(45):14223–14235, 2009. 1.6.4

[82] Cameron T. Ellis, Christopher Baldassano, Anna C. Schapiro, Ming Bo Cai, and Jonathan D. Cohen. Facilitating open-science with realistic fMRI simulation: Validation and application. *PeerJ*, 8:e8564, 2020. 7.3.3.1

[83] Sigrid Elsenbruch, Julia Schmid, Jennifer S. Kullmann, Joswin Kattoor, Nina Theysohn, Michael Forsting, and Vassilios Kotsis. Visceral sensitivity correlates with decreased regional gray matter volume in healthy volunteers: A voxel-based morphometry study. *Pain*, 155(2):244–249, 2014. 6.3.1

[84] Nichole Emerson, Fadel Zeidan, Oleg V Lobanov, Morten S Hadsel, Katherine T Martucci, Alexandre S Quevedo, Christopher J Starr, Hadas Nahman-Averbuch, Irit Weissman-Fogel, Yelena Granovsky, David Yarnitsky, and

Robert Coghill. Pain Sensitivity is Inversely Related to Regional Grey Matter Density in the Brain. *Pain*, 155(3):566–573, 2015. 5.5

[85] Mats Eriksson and Marsha Campbell-Yeo. Assessment of pain in newborn infants. *Seminars in Fetal and Neonatal Medicine*, 24(4):101003, August 2019. 3.4.3

[86] Mats Eriksson, Hanne Storm, Asbjörn Fremming, and Jens Schollin. Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. *Acta Paediatrica*, 97:27–30, 2008. 3.1, 3.4.3

[87] Nathalie Erpelding, Massieh Moayedi, and Karen D. Davis. Cortical thickness correlates of pain and temperature sensitivity. *Pain*, 153(8):1602–1609, 2012. 5.5

[88] L. Fabrizi, G. Williams, A. Lee, J. Meek, R. Slater, S. Olhede, and M. Fitzgerald. Cortical activity evoked by an acute painful tissue-damaging stimulus in healthy adult volunteers. *Journal of Neurophysiology*, 109(9):2393–2403, 2013. 8.2.1

[89] Lorenzo Fabrizi, Rebeccah Slater, Alan Worley, Judith Meek, Stewart Boyd, Sofia Olhede, and Maria Fitzgerald. A shift in sensory processing that enables the developing human brain to discriminate touch from pain. *Current Biology*, 21(18):1552–1558, 2011. 1.7.2, 1.7.3, 1.9.1.3, 3.1, 3.2.5.3, 4.1, 4.3.1.2, 4.3.2, 4.4.1, 4.4.2

[90] Lorenzo Fabrizi, Madeleine Verriotis, Gemma Williams, Amy Lee, Judith Meek, Sofia Olhede, and Maria Fitzgerald. Encoding of mechanical nociception differs in the adult and infant brain. *Scientific Reports*, 6:28642, 2016. 1.9.1.4, 3.2.3, 3.4.2, 4.4.3, 5.6, 2

[91] Yingtao Fang, Jiazhou Wang, Xiaomin Ou, Hongmei Ying, Chaosu Hu, Zhen Zhang, and Weigang Hu. The impact of training sample size on deep

learning-based organ auto-segmentation for head-and-neck patients. *Physics in Medicine & Biology*, 66(18):185012, 2021. 3.4.3

[92] J. Feito, O. García-Suárez, J. García-Piqueras, Y. García-Mesa, A. Pérez-Sánchez, I. Suazo, R. Cabo, J. Suárez-Quintanilla, J. Cobo, and J. A. Vega. The development of human digital Meissner's and Pacinian corpuscles. *Annals of Anatomy*, 219:8–24, September 2018. 1.7.3, 4.4.1

[93] D. G. Ferrington and Mark J. Rowe. Functional capacities of tactile afferent fibres in neonatal kittens. *The Journal of Physiology*, 307(1):335–353, 1980. 1.7.3

[94] M. Fitzgerald and S. Gibson. The postnatal physiological and neurochemical development of peripheral sensory C fibres. *Neuroscience*, 13(3):933–944, November 1984. 1.7.1

[95] Maria Fitzgerald. The development of nociceptive circuits. *Nature Reviews Neuroscience*, 6:507–520, 2005. 1.7.1, 1.7.3, 1.8.1, 4.1

[96] Maria Fitzgerald. What do we really know about newborn infant pain? *Experimental Physiology*, 100(12):1451–1457, 2015. 1.8.1, 3.4.1

[97] Maria Fitzgerald and Ernest Jennings. The postnatal development of spinal sensory processing. *Proceedings of the National Academy of Sciences*, 96(14):7719–7722, July 1999. 3.1

[98] Sean Fitzgibbon. dHCP augmented volumetric atlas-extended. https://git.fmrib.ox.ac.uk/seanf/dhcp-resources/-/blob/master/docs/dhcp-augmented-volumetric-atlas-extended.md. 7.1.5, 7.1.6.1, 7.1.6.2

[99] Sean P. Fitzgibbon, Samuel J. Harrison, Mark Jenkinson, Luke Baxter, Emma C. Robinson, Matteo Bastiani, Jelena Bozek, Vyacheslav Karolis, Lucilio Cordero Grande, Anthony N. Price, Emer Hughes, Antonios Makropoulos, Jonathan Passerat-Palmbach, Andreas Schuh, Jianliang Gao, Seyedeh-Rezvan Farahibozorg, Jonathan O'Muircheartaigh, Judit Ciarrusta, Camilla

O'Keeffe, Jakki Brandon, Tomoki Arichi, Daniel Rueckert, Joseph V. Hajnal, A. David Edwards, Stephen M. Smith, Eugene Duff, and Jesper Andersson. The developing Human Connectome Project (dHCP) automated resting-state functional processing framework for newborn infants. *NeuroImage*, 223:117303, December 2020. 1.9.2, 7.1.4.1

[100] Robert B. Flint, Floor van Beek, Peter Andriessen, Luc J. Zimmermann, Kian D. Liem, Irwin K.M. Reiss, Ronald de Groot, Dick Tibboel, David M. Burger, and Sinno H.P. Simons. Large differences in neonatal drug use between NICUs are common practice: Time for consensus? *British Journal of Clinical Pharmacology*, 84(6):1313–1323, June 2018. 1.3

[101] L. S. Franck, S. Cox, A. Allen, and I. Winter. Parental concern and distress about infant pain. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 89(1):F71–F75, January 2004. 1.3

[102] Marcel Franz, Moritz M. Nickel, Alexander Ritter, Wolfgang H.R. Miltner, and Thomas Weiss. Somatosensory spatial attention modulates amplitudes, latencies, and latency jitter of laser-evoked brain potentials. *Journal of Neurophysiology*, 113(7):2760–2768, 2015. 6.3.2

[103] Jean A. Frazier, Sufen Chiu, Janis L. Breeze, Nikos Makris, Nicholas Lange, David N. Kennedy, Martha R. Herbert, Eileen K. Bent, Vamsi K. Koneru, Megan E. Dieterich, Steven M. Hodge, Scott L. Rauch, P. Ellen Grant, Bruce M. Cohen, Larry J. Seidman, Verne S. Caviness, and Joseph Biederman. Structural Brain Magnetic Resonance Imaging of Limbic and Thalamic Volumes in Pediatric Bipolar Disorder. *American Journal of Psychiatry*, 162(7):1256–1265, July 2005. 6.1.7.2, 7.1.6.3

[104] Luis Garcí-Larrea, Roland Peyron, Bernard Laurent, and François Mauguière. Association and dissociation between laser-evoked potentials and pain perception:. *NeuroReport*, 8(17):3785–3789, December 1997. 1.9.1.2

[105] Pedro J García-Laencina, José-Luis Sancho-Gómez, and Aníbal R Figueiras-Vidal. Pattern classification with missing data: A review. *Neural Comput & Applic*, 19:263–282, 2010. 3.2.6.1

[106] L. Garcia-Larrea, M. Frot, and M. Valeriani. Brain generators of laser-evoked potentials: From dipoles to functional significance. *Neurophysiologie Clinique*, 33(6):279–292, 2003. 1.9.1.2, 5.4, 7.3.3.3

[107] Gauthier, C.J and Fan, A.P. BOLD signal physiology: Models and applications. *NeuroImage*, 187:116–127, February 2019. 1.9.2.1

[108] Maria Geisler, Elizabeth Rizzoni, Nikolaos Makris, Ofer Pasternak, Yogesh Rathi, Sylvain Bouix, Marco Herbsleb, Karl Jürgen Bär, Thomas Weiss, and Zora Kikinis. Microstructural alterations in medial forebrain bundle are associated with interindividual pain sensitivity. *Human Brain Mapping*, 42(4):1130–1137, 2021. 5.5, 7.3.3.3

[109] Carmen Geißler, Christopher Schulze, Sebastian Botzenhardt, Wolfgang Rascher, and Antje Neubert. Drug Utilisation and Off-Label Use on a German Neonatal Intensive Care Unit: A Retrospective Cohort Study and 10-Year Comparison. *Pharmacy (Basel, Switzerland)*, 8(3):E173, September 2020. 1.3

[110] Stanimira Georgieva, Suzannah Lester, Valdas Noreika, Meryem Nazli Yilmaz, Sam Wass, and Victoria Leong. Toward the Understanding of Topographical and Spectral Signatures of Infant Movement Artifacts in Naturalistic EEG. *Frontiers in Neuroscience*, 14:352, April 2020. 3.4.3

[111] Stephan Geuter, Elizabeth A. Reynolds Losin, Mathieu Roy, Lauren Y. Atlas, Liane Schmidt, Anjali Krishnan, Leonie Koban, Tor D. Wager, and Martin A. Lindquist. Multiple Brain Networks Mediating Stimulus-Pain Relationships in Humans. *Cerebral cortex*, 30(7):4204–4219, 2020. 6.3.1

[112] Sharyn Gibbins and Bonnie Stevens. The influence of gestational age on the efficacy and short-term safety of sucrose for procedural pain relief. *Adv Neonatal Care*, 3(5):241–249, 2003. 3.1

[113] Sharyn Gibbins, Bonnie Stevens, Patrick J. McGrath, Janet Yamada, Joseph Beyene, Lynn Breau, Carol Camfield, Allen Finley, Linda Franck, Celeste Johnston, Alixe Howlett, Patricia McKeever, Karel O'Brien, and Arne Ohlsson. Comparison of pain responses in infants of different gestational ages. *Neonatology*, 93(1):10–18, 2007. 3.1

[114] Vito Giordano, Joy Edobor, Philipp Deindl, Brigitte Wildner, Katharina Goeral, Philipp Steinbauer, Tobias Werther, Angelika Berger, and Monika 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, December 2019. 1.4

[115] Sezgi Goksan, Luke Baxter, Fiona Moultrie, Eugene Duff, Gareth Hathway, Caroline Hartley, Irene Tracey, and Rebeccah Slater. The influence of the descending pain modulatory system on infant pain-related brain activity. *eLife*, 7:e37125, 2018. 5.5, 7.3.1, 7.3.3.1

[116] Sezgi Goksan, Caroline Hartley, Faith Emery, Naomi Cockrill, Ravi Poorun, Fiona Moultrie, Richard Rogers, Jon Campbell, Michael Sanders, Eleri Adams, Stuart Clare, Mark Jenkinson, Irene Tracey, and Rebeccah Slater. fMRI reveals neural activity overlap between adult and infant pain. *eLife*, 4:e06356, 2015. 1.7.2, 1.9.2.3, 2.2.2.3, 4.4.5, 5.6, 7.1.1, 7.3.1, 7.3.3.1

[117] Sezgi Goksan, Caroline Hartley, Samuel A. Hurley, Anderson M. Winkler, Eugene P. Duff, Mark Jenkinson, Richard Rogers, Stuart Clare, and Rebeccah 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. 1.9.2, 2.2.2.3

[118] Romain Grandchamp and Arnaud Delorme. Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Frontiers in Psychology*, 2:236, 2011. 1.9.1.4, 3.2.5.4

[119] Caterina Gratton, Timothy O. Laumann, Ashley N. Nielsen, Deanna J. Greene, Evan M. Gordon, Adrian W. Gilmore, Steven M. Nelson, Rebecca S. Coalson, Abraham Z. Snyder, Bradley L. Schlaggar, Nico U. F. Dosenbach, and Steven E. Petersen. Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, 98(2):439–452.e5, April 2018. 7.3.3.3

[120] Gabrielle Green, Caroline Hartley, Amy Hoskin, Eugene Duff, Adam Shriver, Dominic Wilkinson, Eleri Adams, Richard Rogers, Fiona Moultrie, and Rebeccah Slater. Behavioural discrimination of noxious stimuli in infants is dependent on brain maturation. *Pain*, 160(2):493–500, 2019. 1.8.1, 2.1.1, 3.1, 3.2.3, 4.1, 4.2.3, 4.3.1.1, 4.4.2

[121] Ludovica Griffanti, Gwenaëlle Douaud, Janine Bijsterbosch, Stefania Evangelisti, Fidel Alfaro-Almagro, Matthew F. Glasser, Eugene P. Duff, Sean Fitzgibbon, Robert Westphal, Davide Carone, Christian F. Beckmann, and Stephen M. Smith. Hand classification of fMRI ICA noise components. *NeuroImage*, 154:188–205, 2017. 2.3.2

[122] Ludovica Griffanti, Gholamreza Salimi-Khorshidi, Christian F. Beckmann, Edward J. Auerbach, Gwenaëlle Douaud, Claire E. Sexton, Enikő Zsoldos, Klaus P. Ebmeier, Nicola Filippini, Clare E. Mackay, Steen Moeller, Junqian Xu, Essa Yacoub, Giuseppe Baselli, Kamil Ugurbil, Karla L. Miller, and Stephen M. Smith. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, 95:232–247, July 2014. 2.3.2, 6.1.6, 7.1.4.1

[123] Ruth V E Grunau, C Celeste Johnston, and Kenneth D Craig. Neonatal facial and cry responses to invasive and non-invasive procedures. *Pain*, 42(3):295–305, 1990. 1.4, 3.1, 3.2.3

[124] Ruth V.E. Grunau and Kenneth D. Craig. Pain expression in neonates: Facial action and cry. *Pain*, 28(3):395–410, 1987. 1.4

[125] Deniz Gursul, Sezgi Goksan, Caroline Hartley, Gabriela Schmidt Mellado, Fiona Moultrie, Amy Hoskin, Eleri Adams, Gareth Hathway, Susannah Walker, Francis McGlone, and Rebeccah Slater. Stroking modulates noxious-evoked brain activity in human infants. *Current Biology*, 28:R1380–R1381, 2018. 1.9.1.3, 2.1.1, 3, 5.1, 7.3.1

[126] Sven Haase. Weighted median. https://www.mathworks.com/matlabcentral/fileexchange/230 weighted-median, 2021. 4.2.3

[127] Mijna Hadders-Algra. Early human brain development: Starring the subplate. *Neuroscience and Biobehavioral Reviews*, 92:276–290, 2018. 4.1, 4.4.1

[128] Jenny Haefeli, Patrick Freund, John L K Kramer, Julia Blum, Roger Luechinger, and Armin Curt. Differences in cortical coding of heat evoked pain beyond the perceived intensity: An fMRI and EEG study. *Human Brain Mapping*, 35(4):1379–1389, 2014. 5.5

[129] Xiaochun Han, Yoni K. Ashar, Philip Kragel, Bogdan Petre, Victoria Schelkun, Lauren Y. Atlas, Luke J. Chang, Marieke Jepma, Leonie Koban, Elizabeth A. Reynolds Losin, Mathieu Roy, Choong-Wan Woo, and Tor D. Wager. Effect sizes and test-retest reliability of the fMRI-based neurologic pain signature. *NeuroImage*, 247:118844, February 2022. 6.3.3, 3

[130] Julia Hartkopf, Julia Moser, Franziska Schleger, Hubert Preissl, and Jana Keune. Changes in event-related brain responses and habituation during child development – A systematic literature review. *Clinical Neurophysiology*, 130(12):2238–2254, 2019. 4.4.1

[131] Caroline Hartley, Eugene P Duff, Gabrielle Green, Gabriela Schmidt Mellado, Alan Worley, Richard Rogers, and Rebeccah Slater. Nociceptive brain activity as a measure of analgesic efficacy in infants. *Science Translational Medicine*, 9:eaah6122, 2017. 1.9.1.3, 2.1.1.1, 2.1.1.3, 2.4, 3.1, 3.2.3, **??**, 3.2.5.2, 3.2.5.3, 3.4.2, 4.1, 4.2.3, 4.3.1.2, 4.5, 4.4.2, 5.1, 5.2, 5.3, 6.1.5, 6.1.5.1, 7.1.3.1, 7.1.3.2, 7.3.1, 7.3.3.3

[132] Caroline Hartley, Sezgi Goksan, Ravi Poorun, Kelly Brotherhood, Gabriela Schmidt Mellado, Fiona Moultrie, Richard Rogers, Eleri Adams, and Rebeccah Slater. The relationship between nociceptive brain activity, spinal reflex withdrawal and behaviour in newborn infants. *Scientific Reports*, 5(1):12519, 2015. 1.8.1, 1.9.1.3, 1.9.2.3, 2.2.1, 3.1, 3.2.3, 4.4.2, 4.4.3, 6.3.2, 7.1.2, 4

[133] Caroline Hartley, Fiona Moultrie, Deniz Gursul, Amy Hoskin, Eleri Adams, Richard Rogers, and Rebeccah Slater. Changing Balance of Spinal Cord Excitability and Nociceptive Brain Activity in Early Human Development. *Current Biology*, 26:1998–2002, 2016. 1.7.1, 1.8.1, 1.9.1.3, 2.1.1, 3.2.3, 4.1, 4.4.3

[134] Caroline Hartley, Fiona Moultrie, Amy Hoskin, Gabrielle Green, Vaneesha Monk, Jenifer Bell, Andrew King, Miranda Buckle, Marianne van der Vaart, Deniz Gursul, Sezgi Goksan, Ed Jusczak, Jane Norman, Richard Rogers, Chetan Patel, Eleri Adams, and Rebeccah Slater. Analgesic efficacy and safety of morphine in the Procedural Pain in Premature Infants (Poppi) study: Randomised placebo-controlled trial. *The Lancet*, 392:2595–605, 2018. 1.9.1.3, 3, 3.1, 3.4.4.1, 4.1, 4.4.2

[135] Linda A Hatfield and Elizabeth A Ely. Measurement of Acute Pain in Infants: A Review of Behavioral and Physiological Variables. *Biological Research for Nursing*, 17(1):100–111, 2015. 1.4

[136] G. J. Hathway, S. Koch, L. Low, and M. Fitzgerald. The changing balance of brainstem–spinal cord modulation of pain processing over the first weeks of rat postnatal life. *The Journal of Physiology*, 587(12):2927–2935, 2009. 1.7.1

[137] Cassandra L. Hendrix and Moriah E. Thomason. A survey of protocols from 54 infant and toddler neuroimaging research labs. *Developmental Cognitive Neuroscience*, 54:101060, April 2022. 1.9.2

[138] Christiane Hermann, Johanna Hohmeister, Sueha Demirakça, Katrin Zohsel, and Herta Flor. Long-term alteration of pain sensitivity in school-aged children with early pain experiences. *Pain*, 125(3):278–285, 2006. 1.3

[139] Stefanie Hoehl and Sebastian Wahl. Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, 37(3):187–209, 2012. 3.1, 3.4.2

[140] Sture Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. 4.2.3

[141] Liisa Holsti, Ruth E Grunau, Tim F Oberlander, Michael F Whitfield, and Joanne Weinberg. Body movements: An important additional factor in discriminating pain from stress in preterm infants. *Clinical Journal of Pain*, 21(6):491–498, 2005. 1.4, 3.2.3

[142] Björn Horing, Christian Sprenger, and Christian Büchel. The parietal operculum preferentially encodes heat pain and not salience. *PLoS Biology*, 17(8):e3000205, 2019. 6.3.3

[143] Kreshnik Hoti, Paola Teresa Chivers, and Jeffery David Hughes. Assessing procedural pain in infants: A feasibility study evaluating a point-of-care mobile solution based on automated facial analysis. *The Lancet. Digital health*, 3(10):e623–e634, 2021. 3.4.4.2, 8.2.2

[144] L. Hu and G. D. Iannetti. Neural indicators of perceptual variability of pain across species. *Proceedings of the National Academy of Sciences of the United States of America*, 116(5):1782–1791, 2019. 1.9.1.2, 1.9.1.4, 3.4.2

[145] Laurence T. Hunt. The life-changing magic of sharing your data. *Nature Human Behaviour*, 3(4):312–315, 2019. 3.4.1

[146] Florian Hutzler. Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage*, 84:1061–1069, January 2014. 1.9.2.3

[147] G. D. Iannetti, U. Baumgärtner, I. Tracey, R. D. Treede, and W. Magerl. Pinprick-evoked brain potentials: A novel tool to assess central sensitization of nociceptive pathways in humans. *Journal of Neurophysiology*, 110(5):1107–1116, 2013. 1.6.2, 1.9.1.2, 2.2.1, 6.1.5, 6.1.5.1, 6.2.2, 6.3, 6.3.2, 8.3.3

[148] G. D. Iannetti, N. P. Hughes, M. C. Lee, and A. Mouraux. Determinants of Laser-Evoked EEG Responses: Pain Perception or Stimulus Saliency? *Journal of Neurophysiology*, 100(2):815–828, August 2008. 1.9.1.2

[149] G. D. Iannetti, L. Zambreanu, G. Cruccu, and I. Tracey. Operculoinsular cortex encodes pain intensity at the earliest stages of cortical processing as indicated by amplitude of laser-evoked potentials in humans. *Neuroscience*, 131(1):199–208, 2005. 1.9.1.2

[150] IASP Subcommittee on Taxonomy. Pain terms: A list with definitions and notes on usage. Recommended by the IASP Subcommittee on Taxonomy. *Pain*, 6(3):249, June 1979. 1.2

[151] IASP Task Force on Taxonomy. IASP Terminology. https://www.iasp-pain.org/resources/terminology/, 2011. 1.2

[152] Emre Ilhan, Verity Pacey, Laura Brown, Kaye Spence, Kelly Gray, Jennifer E. Rowland, Karolyn White, and Julia M. Hush. Neonates as intrinsically worthy recipients of pain management in neonatal intensive care. *Medicine, Health Care and Philosophy*, 24(1):65–72, March 2021. 8.4

[153] Marion Imbault, Dorian Chauvet, Jean Luc Gennisson, Laurent Capelle, and Mickael Tanter. Intraoperative Functional Ultrasound Imaging of Human Brain Activity. *Scientific Reports*, 7:7304, 2017. 8.3.2

[154] Giorgio M. Innocenti and David J. Price. Exuberance in the development of cortical networks. *Nature Reviews Neuroscience*, 6(12):955–965, 2005. 1.7.2

[155] Jean Isnard, Michel Magnin, Julien Jung, Franois Mauguière, and Luis Garcia-Larrea. Does the insula tell our brain that we are in pain? *Pain*, 152:946–951, 2011. 1.6.4

[156] Alana Jackman and Maria Fitzgerald. Development of peripheral hindlimb and central spinal cord innervation by subpopulations of dorsal root ganglion cells in the embryonic rat. *Journal of Comparative Neurology*, 418(3):281–298, March 2000. 1.7.1

[157] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, June 2001. 6.1.6, 7.1.4.1

[158] Mark Jenkinson. Atlasquery - FslWiki. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlasquery. 6.1.7.2, 7.1.6.3

[159] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 17(2):825–841, 2002. 6.1.6, 7.1.4.1

[160] Mark Jenkinson, Christian F. Beckmann, Timothy E. J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782–790, August 2012. 6.1.6, 6.1.6, 7.1.6.1

[161] Mark Jenkinson, Janine Bijsterbosch, Michael Chappell, and Anderson Winkler. *Short Introduction to the General Linear Model for Neuroimaging*. Primer Appendix. 2.3.2

[162] Mark Jenkinson and Michael Chappell. MRI Modalities for Neuroimaging. In *Introduction to Neuroimaging Analysis*, Oxford Neuroimaging Primers, pages 23–84. Oxford University Press, Oxford, United Kingdom, 2018. 1.9.2.1

[163] Mark Jenkinson and Michael Chappell. Overview of MRI analysis. In *Introduction to Neuroimaging Analysis*, Oxford Neuroimaging Primers, pages 85–143. Oxford University Press, Oxford, United Kingdom, 2018. 2.3.2

[164] C. Céleste Johnston, Bonnie Stevens, Fang Yang, and Linda Horton. Developmental changes in response to heelstick in preterm infants: A prospective cohort study. *Developmental Medicine & Child Neurology*, 38(5):438–445, May 1996. 3.1, 3.2.3, 4.1

[165] Celeste Johnston, Bonnie J. Stevens, Fang Yang, and Linda Horton. Differential response to pain by very premature neonates. *Pain*, 61(3):471–479, 1995. 3.2.3

[166] L. Jones, MP Laudiano-Dray, K. Whitehead, M. Verriotis, J. Meek, M. Fitzgerald, and L. Fabrizi. EEG, behavioural, and physiological responses to a painful procedure in human neonates with relevant medical history. *UK Data Service. SN: 853311. http://doi.org/10.5255/UKDA-SN-853204*, 2018. 2.1.3, 3, 3.2.4, 3.4.1

[167] Laura Jones, Lorenzo Fabrizi, Maria Laudiano-Dray, Kimberley Whitehead, Judith Meek, Madeleine Verriotis, and Maria Fitzgerald. Nociceptive Cortical Activity Is Dissociated from Nociceptive Behavior in Newborn Human Infants under Stress. *Current Biology*, pages 1–6, 2017. 1.8.1, 3.2.5.1, 4.4.2

[168] Laura Jones, Maria Pureza Laudiano-Dray, Kimberley Whitehead, Judith Meek, Maria Fitzgerald, Lorenzo Fabrizi, and Rebecca Pillai Riddell. The impact of parental contact upon cortical noxious-related activity in human neonates. *European Journal of Pain (United Kingdom)*, 25(1):149–159, 2021. 1.9.1.3

[169] Laura Jones, Maria Pureza Laudiano-Dray, Kimberley Whitehead, Madeleine Verriotis, Judith Meek, Maria Fitzgerald, and Lorenzo Fabrizi. EEG, behavioural and physiological recordings following a painful procedure in human neonates. *Scientific Data*, 5:180248, 2018. 2.1.3, 3, 3.2.4, 3.2.5.1, 4.4.4

[170] Laura Jones, Madeleine Verriotis, Robert J Cooper, Maria Pureza Laudiano-Dray, Mohammed Rupawala, Judith Meek, Lorenzo Fabrizi, and Maria Fitzgerald. Widespread nociceptive maps in the human neonatal somatosensory cortex. *eLife*, 11:e71655, April 2022. 4.4.1

[171] A. Kaminska, V. Delattre, J. Laschet, J. Dubois, M. Labidurie, A. Duval, A. Manresa, J. F. Magny, S. Hovhannisyan, M. Mokhtari, L. Ouss, A. Boissel, L. Hertz-Pannier, M. Sintsov, M. Minlebaev, R. Khazipov, and C. Chiron. Cortical auditory-evoked responses in preterm neonates: Revisited by spectral and temporal analyses. *Cerebral Cortex*, 28(10):3429–3444, 2018. 1.7.2

[172] Patrick Kanold. Subplate neurons: Crucial regulators of cortical development and plasticity. *Frontiers in Neuroanatomy*, 3, 2009. 1.7.2

[173] Severin Kasser, Caroline Hartley, Hanna Rickenbacher, Noemi Klarer, Antoinette Depoorter, Alexandre N. Datta, Maria M. Cobo, Sezgi Goksan, Amy Hoskin, Walter Magerl, Evelyn A Huhn, Gabrielle Green, Rebeccah Slater, and Sven Wellmann. Birth experience in newborn infants is associated with changes in nociceptive sensitivity. *Scientific Reports*, 9(4117):1–8, 2019. 1.9.1.3, 4.4.4

[174] Rustem Khazipov, Anton Sirota, Xavier Leinekugel, Gregory L. Holmes, Yehezkel Ben-Ari, and György Buzsáki. Early motor activity drives spindle bursts in the developing somatosensory cortex. *Nature*, 432(7018):758–761, December 2004. 1.7.2

[175] Hiroyuki Kidokoro. Delta brushes are not just a hallmark of EEG in human preterm infants. *Pediatrics International*, 63(2):130–136, February 2021. 1.7.2

[176] Werner Kilb, Sergei Kirischuk, and Heiko J. Luhmann. Electrical activity patterns and the functional maturation of the neocortex: Shaping of developing cortical circuits by electrical activity. *European Journal of Neuroscience*, 34(10):1677–1686, November 2011. 1.7.2

[177] Timo Kirschstein and Rüdiger Köhling. What is the source of the EEG? *Clinical EEG and neuroscience: official journal of the EEG and Clinical Neuroscience Society (ENCS)*, 40(3):146–149, 2009. 1.9.1.1, 4.4.1

[178] Marta Korom, M. Catalina Camacho, Courtney A. Filippi, Roxane Licandro, Lucille A. Moore, Alexander Dufford, Lilla Zöllei, Alice M. Graham, Marisa Spann, Brittany Howell, Sarah Shultz, and Dustin Scheinost. Dear reviewers: Responses to common reviewer critiques about infant neuroimaging studies. *Developmental Cognitive Neuroscience*, 53:101055, February 2022. 1.9.2

[179] Ivica Kostović, Nataša Jovanov-Milošević, Milan Radoš, Goran Sedmak, Vesna Benjak, Mirna Kostović-Srzentić, Lana Vasung, Marko Čuljat, Marko Radoš, Petra Hüppi, and Miloš Judaš. Perinatal and early postnatal reorganization of the subplate and related cellular compartments in the human cerebral wall as revealed by histological and MRI approaches. *Brain Structure and Function*, 219(1):231–253, 2014. 1.7.2, 8.3.1

[180] Ivica Kostović and Miloš Judaš. The development of the subplate and thalamocortical connections in the human foetal brain. *Acta Paediatrica*, 99(8):1119–1127, 2010. 1.7.2, 3.1, 4.4.1

[181] Ivica Kostovic and Pasko Rakic. Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *Journal of Comparative Neurology*, 297(3):441–470, 1990. 1.7.2

[182] M. Kozberg and E. Hillman. Neurovascular coupling and energy metabolism in the developing brain. *Progress in brain research*, 225:213–242, 2016. 1.9.2.1

[183] Željka Krsnik, Visnja Majić, Lana Vasung, Hao Huang, and Ivica Kostović. Growth of thalamocortical fibers to the somatosensory cortex in the human fetal brain. *Frontiers in Neuroscience*, 11(APR):233, 2017. 4.4.1

[184] Eric Kugelberg, Kerstin Eklund, and Lennart Grjmby. Electromyographic Study of the Nociceptive Reflexes of the Lower Limb. Mechanism of the Plantar Responses. *Brain*, 83:394–410, 1960. 1.6.3.1

[185] Miriam Kunz, Jen I. Chen, and Pierre Rainville. Keeping an eye on pain expression in primary somatosensory cortex. *NeuroImage*, 217:116885, 2020. 1.8.1

[186] Miriam Kunz, Stefan Lautenbacher, Nadine Leblanc, and Pierre Rainville. Are both the sensory and the affective dimensions of pain encoded in the face? *Pain*, 153(2):350–358, 2012. 1.8.1

[187] Brandon N. Kyle and Daniel W. McNeil. Autonomic Arousal And Experimentally Induced Pain: A Critical Review of the Literature. *Pain Research and Management*, 19(3):159–167, 2014. 1.8.2

[188] O'Brien M. Kyololo, Bonnie J. Stevens, and Julia Songok. Mothers' Perceptions about Pain in Hospitalized Newborn Infants in Kenya. *Journal of Pediatric Nursing*, 47:51–57, July 2019. 1.3

[189] Landy, Mark A, Goyal, Megan, and Lai, Helen C. Nociceptor subtypes are born continuously over DRG development. *Developmental Biology*, 479:91–98, November 2021. 1.7.1

[190] Maria Pureza Laudiano-Dray, Rebecca R Pillai Ridell, Laura Jones, Rajeshwari Iyer, Kimberley Whitehead, Maria Fitzgerald, Lorenzo Fabrizi, and Judith Meek. Quantification of neonatal procedural pain severity: A platform for estimating total pain burden in individual infants. *Pain*, 161(6):1270–1277, 2020. 8.2.1

[191] J. Lawrence, D. Alcock, P. McGrath, J. Kay, S. B. MacMurray, and C. Dulberg. The development of a tool to assess neonatal pain. *Neonatal network*, 12(6):59–66, September 1993. 1.4

[192] Lechner, Stefan G, Frenzel, Henning, Wang, Rui, and Lewin, Gary R. Developmental waves of mechanosensitivity acquisition in sensory neuron subtypes during embryonic development. *The EMBO Journal*, 28(10):1479–1491, 2009. 1.7.1

[193] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. 3.4.3, 3.4.4.2

[194] Jean Pascal Lefaucheur. Clinical neurophysiology of pain. In *Handbook of Clinical Neurology*, volume 161, pages 121–148. Elsevier B.V., 2019. 1.9.1.2

[195] Susanna Leikos, Anton Tokariev, Ninah Koolen, Päivi Nevalainen, and Sampsa Vanhatalo. Cortical responses to tactile stimuli in preterm infants. *European Journal of Neuroscience*, 51(4):1059–1073, February 2020. 4.4.1, 4.4.2

[196] M. Liang, Q. Su, A. Mouraux, and G. D. Iannetti. Spatial Patterns of Brain Activity Preferentially Reflecting Transient Pain and Stimulus Intensity. *Cerebral Cortex*, 29(5):2211–2227, 2019. 1.6.4

[197] Marco L. Loggia, Mylne Juneau, and M. Catherine Bushnell. Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain*, 152(3):592–598, 2011. 1.8.2

[198] Luck, SJ. Artifact Rejection and Correction. In *An Introduction to the Event-Related Potential Technique.*, pages 219–248. MIT Press, 2014. 2.3.1

[199] Luck, SJ. Basics of Fourier Analysis and Filtering. In *An Introduction to the Event-Related Potential Technique.*, pages 219–248. MIT Press, 2014. 1.9.1.4, 2.3.1

[200] Luck, SJ. A Broad Overview of the Event-Related Potential Technique. In *An Introduction to the Event-Related Potential Technique.*, pages 1–34. MIT Press, 2014. 1.9.1

[201] Luck, SJ. A Closer Look at ERPs and ERP Components. In *An Introduction to the Event-Related Potential Technique.*, pages 35–70. MIT Press, 2014. 1.9.1.1, 4.4.1

[202] Émilie Macé, Gabriel Montaldo, Stuart Trenholm, Cameron Cowan, Alexandra Brignall, Alan Urban, and Botond Roska. Whole-Brain Functional Ultrasound Imaging Reveals Brain Modules for Visuomotor Integration. *Neuron*, 100(5):1241–1251.e7, 2018. 8.3.2

[203] Walter Magerl, Perry N. Fuchs, Richard A. Meyer, and Rolf-Detlef Treede. Roles of capsaicin-insensitive nociceptors in cutaneous pain and secondary hyperalgesia. *Brain*, 124(9):1754–1764, September 2001. 1.6.2, 1.9.1.2, 1.9.2.3, 2.2.1, 8.3.3

[204] Nathalie L. Maitre, Alexandra P. Key, Olena D. Chorna, James C. Slaughter, Pawel J. Matusz, Mark T. Wallace, and Micah M. Murray. The Dual Nature of Early-Life Experience on Somatosensory Processing in the Human Infant Brain. *Current Biology*, 27:1048–1054, 2017. 4.4.5

[205] Nathalie L. Maitre, Ann R. Stark, Carrie C. McCoy Menser, Olena D. Chorna, Daniel J. France, Alexandra F. Key, Ken Wilkens, Melissa Moore-Clingenpeel, Don M. Wilkes, and Stephen 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. 4.4.4

[206] Scott Makeig. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, 86:283–293, 1993. 1.9.1.4

[207] Nikos Makris, Jill M. Goldstein, David Kennedy, Steven M. Hodge, Verne S. Caviness, Stephen V. Faraone, Ming T. Tsuang, and Larry J. Seidman. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, 83(2):155–171, April 2006. 6.1.7.2, 7.1.6.3

[208] Antonios Makropoulos, Emma C. Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J. Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, Gregor Lenz, Filippo Mortari, Tencho Tenev, Eugene P. Duff, Matteo Bastiani, Lucilio Cordero-Grande, Emer Hughes, Nora Tusor, Jacques-Donald Tournier, Jana Hutter, Anthony N. Price, Rui Pedro A. G. Teixeira, Maria Murgasova, Suresh Victor, Christopher Kelly, Mary A. Rutherford, Stephen M. Smith, A. David Edwards, Joseph V. Hajnal, Mark Jenkinson, and Daniel Rueckert. The Developing Human Connectome Project: A Minimal Processing Pipeline for Neonatal Cortical Surface Reconstruction. *NeuroImage*, 173:88–112, June 2018. 7.1.4.1

[209] Audrey Marchal, Meggane Melchior, André Dufour, Pierrick Poisbeau, Claire Zores, and Pierre Kuhn. Pain Behavioural Response to Acoustic and Light Environmental Changes in Very Preterm Infants. *Children*, 8(12):1081, November 2021. 3.4.4.2

[210] Stephen D. Mayhew, Nicholas Hylands-White, Camillo Porcaro, Stuart W.G. Derbyshire, and Andrew P. Bagshaw. Intrinsic variability in the human response to pain is assembled from multiple, dynamic brain processes. *NeuroImage*, 75:68–78, 2013. 5.4, 5.5, 7.3.3.3

[211] L. Mazzola, J. Isnard, and F. Mauguière. Somatosensory and Pain Responses to Stimulation of the Second Somatosensory Area (SII) in Humans. A Comparison with SI and Insular Responses. *Cerebral Cortex*, 16(7):960–968, July 2006. 1.6.4

[212] Ronald McCarthy, Carmel Martin-Fairey, Dorothy K Sojka, Erik D Herzog, Emily S Jungheim, Molly J Stout, Justin C Fay, Mala Mahendroo, Jeff

Reese, Jennifer L Herington, Erin J Plosa, Elaine L Shelton, and Sarah K England. Mouse models of preterm birth: Suggested assessment and reporting guidelines. *Biology of Reproduction*, 99(5):922–937, November 2018. 1.7.1

[213] Naomi Meesters, Tinne Dilles, Sinno Simons, and Monique van Dijk. Do Pain Measurement Instruments Detect the Effect of Pain-Reducing Interventions in Neonates? A Systematic Review on Responsiveness. *Journal of Pain*, 20(7):760–770, 2019. 1.4, 3.4.4.2, 8.2.1

[214] Ronald Melzack. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain*, 1(3):277–299, September 1975. 6.1.2.1

[215] Ronald Melzack. Phantom limbs and the concept of a neuromatrix. *Trends in Neurosciences*, 13(3):88–92, March 1990. 1.6.1

[216] Ronald Melzack. From the gate to the neuromatrix. *PAIN*, 82:S121–S126, August 1999. 1.6.1

[217] Melzack, Ronald and Wall, Patrick D. Pain Mechanisms: A New Theory. *Science*, 150(3699):971–979, 1965. 1.6.1

[218] Lorne M Mendell. Constructing and Deconstructing the Gate Theory of Pain. *Pain*, 155(2):210–216, 2014. 1.6.1

[219] Mathieu Milh, Anna Kaminska, Catherine Huon, Alexandre Lapillonne, Yehezkel Ben-Ari, and Rustem Khazipov. Rapid cortical oscillations and early motor activity in premature human neonate. *Cerebral Cortex*, 17(7):1582–1594, 2007. 1.7.2

[220] Dominik Mischkowski, Esther E. Palacios-Barrios, Lauren Banker, Troy C. Dildine, and Lauren Y. Atlas. Pain or nociception? Subjective experience mediates the effects of acute noxious heat on autonomic responses - Corrected and republished. *Pain*, 160(6):1469–1481, 2019. 1.8.2

[221] Gaurav Misra, Wei En Wang, Derek B. Archer, Arnab Roy, and Stephen A. Coombes. Automated classification of pain perception using high-density electroencephalography data. *Journal of Neurophysiology*, 117(2):786–795, 2017. 6.3.1

[222] AJ Mitchell, A Green, DA Jeffs, and PK Roberson. Physiologic effects of retinopathy of prematurity screening examinations. *Advanced Neonatal Care*, 11(4):291–297, 2011. 1.3, 3.4.4.1

[223] Massieh Moayedi and Karen D. Davis. Theories of pain: From specificity to gate control. *Journal of Neurophysiology*, 109(1):5–12, 2013. 1.6.1

[224] A. Mobascher, J. Brinkmeyer, T. Warbrick, F. Musso, H. J. Wittsack, A. Saleh, A. Schnitzler, and G. Winterer. Laser-evoked potential P2 single-trial amplitudes covary with the fMRI BOLD response in the medial pain system and interconnected subcortical structures. *NeuroImage*, 45(3):917–926, 2009. 4.4.1, 5.4, 7.3.2, 7.3.3.3

[225] A. Mobascher, J. Brinkmeyer, T. Warbrick, F. Musso, H. J. Wittsack, R. Stoermer, A. Saleh, A. Schnitzler, and G. Winterer. Fluctuations in electrodermal activity reveal variations in single trial brain responses to painful laser stimuli - A fMRI/EEG study. *NeuroImage*, 44(3):1081–1092, 2009. 5.4

[226] Moehring, Francie, Halter, Priyabarata, Seal, Rebecca, and Stucky, Cheryl L. Uncovering the Cells and Circuits of Touch in Normal and Pathological Settings. *Neuron*, 100(2):349–360, October 2018. 1.6.5

[227] Mokkink, L.B., Terwee, C.B., Patrick, D.L., Stratford, P.W., Knol, D.W., Bouter, L.M., and de Vet, H. C. W. 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, July 2010. 1.5, 3.4.4.2, 3.4.4.2, 5.2, 8.2.1

[228] A. R. Moore, W.-L. Zhou, I. Jakovcevski, N. Zecevic, and S. D. Antic. Spontaneous Electrical Activity in the Human Fetal Cortex In Vitro. *Journal of Neuroscience*, 31(7):2391–2398, February 2011. 1.7.2

[229] Carsten Dahl Mørch, Ole Kæseler Andersen, Thomas Graven-Nielsen, and Lars Arendt-Nielsen. Nociceptive withdrawal reflexes evoked by uniform-temperature laser heat stimulation of large skin areas in humans. *Journal of Neuroscience Methods*, 160(1):85–92, 2007. 1.8.1

[230] Evalotte Mörelius, Lena Hellström-Westas, Catarina Carlén, Elisabeth Norman, and Nina Nelson. Is a nappy change stressful to neonates? *Early human development*, 82(10):669–676, October 2006. 1.4

[231] F Moultrie, A Shriver, C Hartley, D Wilkinson, A Ewer, R Rogers, E Adams, and R Slater. A universal right to pain relief - balancing the risks in a vulnerable patient population. *Lancet Child and Adolescent Health*, 3(2):62–64, 2018. 3.4.4.1

[232] Fiona Moultrie, Rebeccah Slater, and Caroline Hartley. Improving the treatment of infant pain. *Current Opinion in Supportive and Palliative Care*, 11(2):112–117, 2017. 1.10, 3.1

[233] A. Mouraux and G. D. Iannetti. Nociceptive laser-evoked brain potentials do not reflect nociceptive-specific neural activity. *Journal of Neurophysiology*, 101(6):3258–3269, 2009. 3.4.4.2, 5.3, 6.3.2

[234] André Mouraux and Gian Domenico Iannetti. The search for pain biomarkers in the human brain. *Brain*, 141:3290–3307, 2018. 6.3.2

[235] Sarah B. Mulkey, Rathinaswamy B. Govindan, Laura Hitchings, Tareq Al-Shargabi, Nicole Herrera, Christopher B. Swisher, Augustine Eze, Stephanie Russo, Sarah D. Schlatterer, Marni B. Jacobs, Robert McCarter, Alex Kline, G. Larry Maxwell, Robin Baker, and Adre J. du Plessis. Autonomic nervous

system maturation in the premature extrauterine milieu. *Pediatric Research*, 89(4):863–868, March 2021. 4.1, 4.4.3

[236] M. M. Myers, P. G. Grieve, A. Izraelit, W. P. Fifer, J. R. Isler, R. A. Darnall, and R. I. Stark. Developmental profiles of infant EEG: Overlap with transient cortical circuits. *Clinical Neurophysiology*, 123(8):1502–1511, August 2012. 4.1, 4.4.1, 8.3.1

[237] Lynn Neumann, Niklas Wulms, Vanessa Witte, Tamas Spisak, Matthias Zunhammer, Ulrike Bingel, and Tobias Schmidt-Wilcke. Network properties and regional brain morphology of the insular cortex correlate with individual pain thresholds. *Human Brain Mapping*, 42(15):4896–4908, 2021. 5.5, 7.3.3.3

[238] Moritz M. Nickel, Elisabeth S. May, Laura Tiemann, Martina Postorino, Son Ta Dinh, and Markus Ploner. Autonomic responses to tonic pain are more closely related to stimulus intensity than to pain intensity. *Pain*, 158(11):2129–2136, November 2017. 1.8.2

[239] Moritz M. Nickel, Laura Tiemann, Vanessa D. Hohn, Elisabeth S. May, Cristina Gil Avila, Falk Eippert, and Markus Ploner. Temporal–spectral signaling of sensory information and expectations in the cerebral processing of pain. *Proceedings of the National Academy of Sciences of the United States of America*, 119:e2116616119, 2022. 1.9.1.2

[240] Hendrik J. Niemarkt, Ward Jennekens, Jaco W. Pasman, Titia Katgert, Carola Van Pul, Antonio W.D. Gavilanes, Boris W. Kramer, Luc J. Zimmermann, Sidarto Bambang Oetomo, and Peter Andriessen. Maturational changes in automated EEG spectral power analysis in preterm infants. *Pediatric Research*, 70(5):529–534, 2011. 1.7.2, 4.1

[241] Elisabeth Norman, Ingmar Rosén, Sampsa Vanhatalo, Karin Stjernqvist, Ove Ökland, Vineta Fellman, and Lena Hellström-Westas. Electroencephalographic response to procedural pain in healthy term newborn infants. *Pediatric Research*, 64(4):429–434, 2008. 1.9.1.4

[242] Tim F. Oberlander, Ruth E. Grunau, Colleen Fitzgerald, and Michael F. Whitfield. Does parenchymal brain injury affect biobehavioral pain responses in very low birth weight infants at 32 weeks' postconceptional age? *Pediatrics*, 110(3):570–576, 2002. 1.8.1

[243] Arne Ohlsson and Prakeshkumar S. Shah. Paracetamol (acetaminophen) for prevention or treatment of pain in newborns. *Cochrane Database of Systematic Reviews*, (1):CD011219, 2020. 8.2.1

[244] Weiwei Peng and Dandan Tang. Pain related cortical oscillations: Methodological advances and potential applications. *Frontiers in Computational Neuroscience*, 10:9, 2016. 1.9.1.4

[245] Luis Pereira-da-Silva, Daniel Virella, Ivete Monteiro, Sandra Gomes, Patrícia Rodrigues, Micaela Serelha, and Hanne Storm. Skin conductance indices discriminate nociceptive responses to acute stimuli from different heel prick procedures in infants. *The Journal of Maternal-Fetal & Neonatal Medicine*, 25(6):796–801, June 2012. 3.4.3

[246] Edward R. Perl. Ideas about pain, a historical view. *Nature Reviews Neuroscience*, 8(1):71–80, January 2007. 1.6.1

[247] Kristine Phillips and Daniel J Clauw. Central pain mechanisms in chronic pain states – Maybe it is all in their head. *Best Practice & Research Clinical Rheumatology*, 25(2):141–154, April 2011. 1.2

[248] Kirubin Pillay, Anneleen Dereymaeker, Katrien Jansen, Gunnar Naulaers, and Maarten 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, April 2020. 4.4.2

[249] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, September 2019. 6.1.4

[250] R Poldrack. Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2):59–63, February 2006. 1.9.2.3

[251] Russell A. Poldrack. Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, 72(5):692–697, December 2011. 1.9.2.3

[252] Russell A. Poldrack, Grace Huckins, and Gael Varoquaux. Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, pages 1–7, 2019. 3.4.3

[253] F. L. Porter, C. M. Wolf, and J. P. Miller. Procedural pain in newborn infants: The influence of intensity and development. *Pediatrics*, 104(1):e13, 1999. 3.1, 4.1

[254] Lihui Pu, Katarzyna Malgorzata Lion, Michael Todorovic, and Wendy Moyle. Portable EEG monitoring for older adults with dementia and chronic pain - A feasibility study. *Geriatric Nursing*, 42(1):124–128, January 2021. 8.2.2

[255] D. Pugash, G. Hendson, C. P. Dunham, K. Dewar, D. M. Money, and D. Prayer. Sonographic assessment of normal and abnormal patterns of fetal cerebral lamination. *Ultrasound in Obstetrics and Gynecology*, 40(6):642–651, 2012. 8.3.1

[256] Kathleen A. Puntillo, Adeline Max, Jean-Francois Timsit, Lucile Vignoud, Gerald Chanques, Gemma Robleda, Ferran Roche-Campo, Jordi Mancebo, Jigeeshu V. Divatia, Marcio Soares, Daniela C. Ionescu, Ioana M. Grintescu, Irena L. Vasiliu, Salvatore Maurizio Maggiore, Katerina Rusinova, Radoslaw Owczuk, Ingrid Egerod, Elizabeth D. E. Papathanassoglou, Maria Kyranou, Gavin M. Joynt, Gastón Burghi, Ross C. Freebairn, Kwok M. Ho, Anne Kaarlola, Rik T. Gerritsen, Jozef Kesecioglu, Miroslav M. S. Sulaj, Michelle Norrenberg, Dominique D. Benoit, Myriam S. G. Seha, Akram Hennein, Fernando J. Periera, Julie S. Benbenishty, Fekri Abroug, Andrew Aquilina,

Júlia R. C. Monte, Youzhong An, and Elie Azoulay. Determinants of Procedural Pain Intensity in the Intensive Care Unit. The Europain® Study. *American Journal of Respiratory and Critical Care Medicine*, 189(1):39–47, January 2014. 1.3

[257] Gari Purcell-Jones, Frances Dormon, and Edward Sumner. Paediatric anaesthetists' perceptions of neonatal and infant pain. *Pain*, 33(2):181–187, May 1988. 1.1

[258] Srinivasa N. Raja, Daniel B. Carr, Milton Cohen, Nanna B. Finnerup, Herta Flor, Stephen Gibson, Francis J. Keefe, Jeffrey S. Mogil, Matthias Ringkamp, Kathleen A. Sluka, Xue-Jun Song, Bonnie Stevens, Mark D. Sullivan, Perri R. Tutelman, Takahiro Ushida, and Kyle Vader. The revised International Association for the Study of Pain definition of pain: Concepts, challenges, and compromises. *PAIN*, 161(9):1976–1982, September 2020. 1.2, 3.1

[259] Manon Ranger, Cecil M.Y. Chau, Amanmeet Garg, Todd S. Woodward, Mirza Faisal Beg, Bruce Bjornson, Kenneth Poskitt, Kevin Fitzpatrick, Anne R. Synnes, Steven P. Miller, and Ruth E. Grunau. Neonatal Pain-Related Stress Predicts Cortical Thickness at Age 7 Years in Children Born Very Preterm. *PLoS ONE*, 8(10):e76702, 2013. 1.3

[260] Manon Ranger, C. Céleste Johnston, and K. J S Anand. Current Controversies Regarding Pain Assessment in Neonates. *Seminars in Perinatology*, 31:283–288, 2007. 3.1

[261] Manon Ranger, Jill G. Zwicker, Cecil M.Y. Chau, Min Tae M. Park, M. Mallar Chakravarthy, Kenneth Poskitt, Steven P. Miller, Bruce H. Bjornson, Emily W.Y. Tam, Vann Chau, Anne R. Synnes, and Ruth E. Grunau. Neonatal Pain and Infection Relate to Smaller Cerebellum in Very Preterm Children at School Age. *Journal of Pediatrics*, 167(2):292–298, 2015. 1.3

[262] Richard Rau, Pieter Kruizinga, Frits Mastik, Markus Belau, Nico de Jong, Johannes G. Bosch, Wolfgang Scheffer, and Georg Maret. 3D functional ultrasound imaging of pigeons. *NeuroImage*, 183:469–477, 2018. 8.3.2

[263] Morgan Recher, Mohamed Riadh Boukhris, Mathieu Jeanne, Laurent Storme, Stéphane Leteurtre, Nada Sabourdin, and Julien De jonckheere. The newborn infant parasympathetic evaluation in pediatric and neonatology: A literature review. *Journal of Clinical Monitoring and Computing*, 35(5):959–966, October 2021. 8.2.2

[264] Daniëlla W.E. Roofthooft, Sinno H.P. Simons, Kanwaljeet J.S. Anand, Dick Tibboel, and Monique Van Dijk. Eight years later, are we still hurting newborn infants? *Neonatology*, 105(3):218–226, 2014. 1.3

[265] Jan Rosner, Paulina Simonne Scheuren, Stephanie Anja Stalder, Armin Curt, and Michèle Hubli. Pinprick Evoked Potentials-Reliable Acquisition in Healthy Human Volunteers. *Pain medicine*, 21(4):736–746, 2020. 6.1.5, 6.1.5.1, 6.2.2, 6.3, 6.3.2

[266] Jean Michel Roué, Iris Morag, Wassim M. Haddad, Behnood Gholami, and Kanwaljeet J.S. Anand. Using sensor-fusion and machine-learning algorithms to assess acute pain in non-verbal infants: A study protocol. *BMJ Open*, 11(1):1–7, 2021. 3.4.1

[267] Jean Michel Roué, Stéphane Rioualen, Julie Gendras, Laurent Misery, Maëlenn Gouillou, and Jacques Sizun. Multi-modal pain assessment: Are near-infrared spectroscopy, skin conductance, salivary cortisol, physiologic parameters, and Neonatal Facial Coding System interrelated during venepuncture in healthy, term neonates? *Journal of Pain Research*, 11:2257–2267, 2018. 1.8.2, 3.4.3

[268] Joni N Saby and Peter J Marshall. The Utility of EEG Band Power Analysis in the Study of Infancy and Early Childhood. *Dev Neuropsychol*, 37(3):253–273, 2012. 3.4.2

[269] Md Sirajus Salekin, Peter R. Mouton, Ghada Zamzmi, Raj Patel, Dmitry Goldgof, Marcia Kneusel, Sammie L. Elkins, Eileen Murray, Mary E. Coughlin, Denise Maguire, Thao Ho, and Yu Sun. Future roles of artificial intelligence in early pain management of newborns. *Paediatric and Neonatal Pain*, 3:134–145, 2021. 3.4.3

[270] Md Sirajus Salekin, Ghada Zamzmi, Dmitry Goldgof, Rangachar Kasturi, Thao Ho, and Yu Sun. Multimodal spatio-temporal deep learning approach for neonatal postoperative pain assessment. *Computers in Biology and Medicine*, 129:104150, 2021. 3.4.3, 3.4.4.2

[271] Md Sirajus Salekin, Ghada Zamzmi, Jacqueline Hausmann, Dmitry Goldgof, Rangachar Kasturi, Marcia Kneusel, Terri Ashmeade, Thao Ho, and Yu Sun. Multimodal neonatal procedural and postoperative pain assessment dataset. *Data in Brief*, 35:106796, April 2021. 3.4.1

[272] Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F. Beckmann, Matthew F. Glasser, Ludovica Griffanti, and Stephen M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–468, April 2014. 2.3.2, 6.1.6, 7.1.4.1

[273] Giorgio Sandrini, Mariano Serrao, Paolo Rossi, Antonietta Romaniello, Giorgio Cruccu, and Jean Claude Willer. The lower limb flexion reflex in humans. *Progress in Neurobiology*, 77(6):353–395, 2005. 1.6.3.1, 1.8.1

[274] Karin Schenk, Liliane Stoffel, Reto Bürgin, Bonnie Stevens, Dirk Bassler, Sven Schulzke, Mathias Nelle, and Eva Cignacco. The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonates. *BMC Pediatrics*, 19:1–21, 2019. 3.4.3, 4.1, 4.4.3

[275] Gabriela Schmidt Mellado, Kirubin Pillay, Eleri Adams, Ana Alarcon, Foteini Andritsou, Maria M Cobo, Ria Evans Fry, Sean Fitzgibbon, Fiona Moultrie,

Luke Baxter, and Rebeccah Slater. The impact of premature extrauterine exposure on infants' stimulus-evoked brain activity across multiple sensory systems. *NeuroImage: Clinical*, 33:102914, 2021. 2.1.1, 3, 4.1, 4.4.5, 8.3.1

[276] Andreas Schuh, Antonios Makropoulos, Emma C. Robinson, Lucilio Cordero-Grande, Emer Hughes, Jana Hutter, Anthony N Price, Maria Murgasova, Rui Pedro A. G. Teixeira, Nora Tusor, Johannes K. Steinweg, Suresh Victor, Mary A. Rutherford, Joseph V. Hajnal, A. David Edwards, and Daniel Rueckert. Unbiased construction of a temporally consistent morphological atlas of neonatal brain development. *bioRxiv : the preprint server for biology*, 2018. 7.1.4.1, 7.1.5, 7.1.6.1, 7.1, 7.1.6.2, 7.7

[277] Enrico Schulz, Andrew Zherdin, Laura Tiemann, Claudia Plant, and Markus Ploner. Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. *Cerebral Cortex*, 22(5):1118–1123, 2012. 1.9.1.4, 5.5

[278] David J. Scott, Mary M. Heitzeg, Robert A. Koeppe, Christian S. Stohler, and Jon Kar Zubieta. Variations in the human pain stress experience mediated by ventral and dorsal basal ganglia dopamine activity. *Journal of Neuroscience*, 26(42):10789–10795, 2006. 6.3.1

[279] Andrew R. Segerdahl, Melvin Mezue, Thomas W. Okell, John T. Farrar, and Irene Tracey. The dorsal posterior insula subserves a fundamental role in human pain. *Nature Neuroscience*, 18(4):499–500, 2015. 1.6.4, 6.3.3

[280] Gila Sellam, Eva L. Cignacco, Kenneth D. Craig, and Sandra Engberg. 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. 3.1

[281] Alexander J. Shackman, Tim V. Salomons, Heleen A. Slagter, Andrew S. Fox, Jameel J. Winter, and Richard J. Davidson. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, 12(3):154–167, March 2011. 1.6.4

[282] Allan Siegel and Hreday N Sapru. Somatosensory System. In *Essential Neuroscience*, pages 258–266. Lippincott Williams & Wilkins, a Wolters Kluwers Business, Baltimore, 2011. 1.6.2, 1.6.3, 1.6.3.2, 1.6.4, 1.6.5

[283] Allan Siegel and Hreday N Sapru. The Spinal Cord. In *Essential Neuroscience*, pages 146–151. Lippincott Williams & Wilkins, a Wolters Kluwers Business, Baltimore, 2011. 1.6.3, 1.6.3.2, 1.6.4, 1.6.5

[284] Sinno H. P. Simons, Monique van Dijk, Kanwaljeet S. Anand, Daniella Roofthooft, Richard A. van Lingen, and Dick Tibboel. Do We Still Hurt Newborn Babies? A Prospective Study of Procedural Pain and Analgesia in Neonates. *Archives of Pediatrics & Adolescent Medicine*, 157(11):1058–1064, November 2003. 1.3

[285] 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. *Lancet*, 376(9748):1225–1232, 2010. 3, 3.1, 3.2.5.3

[286] 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. 1.9.1.3, 3.2.5.3

[287] Rebeccah Slater. The challenge of distinguishing pain from distress in young children. *The Lancet. Child & Adolescent Health*, 3(6):367–368, June 2019. 1.4

[288] Rebeccah Slater, Anne Cantarella, Shiromi Gallella, Alan Worley, Stewart Boyd, Judith Meek, and Maria Fitzgerald. Cortical Pain Responses in Human Infants. *Journal of Neuroscience*, 26(14):3662–3666, 2006. 1.7.2, 1.9, 4.1

[289] Rebeccah Slater, Fiona Moultrie, Ralph Bax, John van den Anker, and Aomesh Bhatt. Preterm health: Time to bridge the evidence gap. *The Lancet*, 396(10255):872–873, 2020. 1.3

[290] Rebeccah Slater, Alan Worley, Lorenzo Fabrizi, Siân Roberts, Judith Meek, Stewart Boyd, and Maria Fitzgerald. Evoked potentials generated by noxious stimulation in the human infant brain. *Eur J Pain*, 14(3):321–326, March 2010. 1.9.1.3, 3.1, 3.2.5.3, 4.3.1.2, 4.3.2, 4.4.5, 7.3.1

[291] Stephen M Smith and J Michael Brady. SUSAN—A New Approach to Low Level Image Processing. *International Journal of Computer Vision*, 23(1):45–78, 1997. 6.1.6, 7.1.4.1

[292] Stephen M. Smith and Thomas E. Nichols. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1):83–98, January 2009. 6.1.7.1

[293] Tamas Spisak, Balint Kincses, Frederik Schlitt, Matthias Zunhammer, Tobias Schmidt-Wilcke, Zsigmond T. Kincses, and Ulrike Bingel. Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. *Nature Communications*, 11:187, 2020. 5.5

[294] Christopher J. Starr, Lumy Sawaki, George F. Wittenberg, Jonathan H. Burdette, Yoshitetsu Oshiro, Alexandre S. Quevedo, John G. McHaffie, and Robert C. Coghill. The contribution of the putamen to sensory aspects of pain: Insights from structural connectivity and brain lesions. *Brain*, 134(7):1987–2004, 2011. 6.3.1

[295] Stephen. ColorBrewer: Attractive and Distinctive Colormaps. https://github.com/DrosteEffect/BrewerMap, 2021. 2.4

[296] B Stevens, C Johnston, P Petryshen, and A Taddio. Premature Infant Pain Profile: Development and initial validation. *Clin J Pain*, 12(1):13–22, 1996. 1.4, 1.5, 3.1

[297] Bonnie Stevens, Sharyn Gibbins, Janet Yamada, Kimberley Dionne, Grace Lee, Celeste Johnston, and Anna Taddio. The Premature Infant Pain Profile-Revised (PIPP-R). *The Clinical journal of pain*, 30(3):238–243, 2014. 3.1

[298] Joseph Stromberg. The Neuroscientist Who Discovered He Was a Psychopath. *Smithsonian Magazine*, November 2013. 1.9.2.3

[299] Q. Su, W. Qin, Q. Q. Yang, C. S. Yu, T. Y. Qian, A. Mouraux, G. D. Iannetti, and M. Liang. Brain regions preferentially responding to transient and iso-intense painful or tactile stimuli. *NeuroImage*, 192:52–65, 2019. 1.6.4

[300] Franois Tadel, Sylvain Baillet, John C. Mosher, Dimitrios Pantazis, and Richard M. Leahy. Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011:879716, 2011. 2.4, 3.2.5.1

[301] Linette Liqi Tan and Rohini Kuner. Neocortical circuits in pain and pain relief. *Nature Reviews Neuroscience*, 22(8):458–471, 2021. 1.6.3.2, 1.6.4, 1.9.1.2

[302] Mickael Tanter, Emilie Macé, Bruno-Felix Osmanski, Gabriel Montaldo, Ivan Cohen, and Mathias Fink. Functional ultrasound imaging of the brain: Theory and basic principles. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 60(3):492–506, 2013. 8.3.2

[303] A. Taylor and G. McLeod. Basic pharmacology of local anaesthetics. *BJA Education*, 20(2):34–41, February 2020. 5.3

[304] Laura Tiemann, Vanessa D. Hohn, Son Ta Dinh, Elisabeth S. May, Moritz M. Nickel, Joachim Gross, and Markus Ploner. Distinct patterns of brain activity mediate perceptual and motor and autonomic responses to noxious stimuli. *Nature Communications*, 9(1):4487, 2018. 1.9.1.2, 5.3

[305] Andrew J. Todd. Neuronal circuitry for pain processing in the dorsal horn. *Nature Reviews Neuroscience*, 11:823–836, 2010. 1.6.3

[306] E. A. Tolner, A. Sheikh, A. Y. Yukin, K. Kaila, and P. O. Kanold. Subplate Neurons Promote Spindle Bursts and Thalamocortical Patterning in the Neonatal Rat Somatosensory Cortex. *Journal of Neuroscience*, 32(2):692–702, January 2012. 1.7.2, 8.3.1

[307] David Tomé, Fernando Barbosa, Kamila Nowak, and João Marques-Teixeira. The development of the N1 and N2 components in auditory oddball paradigms: A systematic review with narrative analysis and suggested normative values. *Journal of Neural Transmission*, 122(3):375–391, 2015. 4.4.1

[308] Irene Tracey and Patrick W. Mantyh. The Cerebral Signature for Pain Perception and Its Modulation. *Neuron*, 55:377–391, 2007. 1.6.1, 1.6.4

[309] Irene Tracey, Clifford J. Woolf, and Nick A. Andrews. Composite Pain Biomarker Signatures for Objective Assessment and Effective Treatment. *Neuron*, 101(5):783–800, 2019. 1.9.2.2, 3.1, 5.2

[310] Yiheng Tu, Ao Tan, Yanru Bai, Yeung Sam Hung, and Zhiguo Zhang. Decoding subjective intensity of nociceptive pain from pre-stimulus and post-stimulus brain activities. *Frontiers in Computational Neuroscience*, 10:32, 2016. 1.9.1.4

[311] Yiheng Tu, Binlong Zhang, Jin Cao, Georgia Wilson, Zhiguo Zhang, and Jian Kong. Identifying inter-individual differences in pain threshold using brain connectome: A test-retest reproducible study. *Neuroimage*, 202:116049, 2019. 5.5

[312] Jaymin Upadhyay, Jordan Lemme, Julie Anderson, David Bleakman, Thomas Large, Jeffrey L. Evelhoch, Richard Hargreaves, David Borsook, and Lino Becerra. Test–retest reliability of evoked heat stimulation BOLD fMRI. *Journal of Neuroscience Methods*, 253:38–46, September 2015. 3, 8.3.3

[313] Alan Urban, Clara Dussaux, Guillaume Martel, Clément Brunner, Emilie Mace, and Gabriel Montaldo. Real-time imaging of brain activity in freely

moving rats using functional ultrasound. *Nature Methods*, 12(9):873–878, September 2015. 8.3.2

[314] Geert J. M. Van Boxtel. Computational and statistical methods for analyzing event-related potential data. *Behavior Research Methods, Instruments, & Computers*, 30(1):87–102, March 1998. 3.2.5.3

[315] E. N. van den Broeke, P. de Hemptinne, M. Mercken, D. M. Torta, J. Lambert, and A. Mouraux. Central sensitization of nociceptive pathways demonstrated by robot-controlled pinprick-evoked brain potentials. *Clinical Neurophysiology*, 131(10):2491–2498, 2020. 6.1.5, 6.1.5.1, 6.3

[316] Emanuel N. van den Broeke, Bart de Vries, Julien Lambert, Diana M. Torta, and André Mouraux. Phase-locked and non-phase-locked EEG responses to pinprick stimulation before and after experimentally-induced secondary hyperalgesia. *Clinical Neurophysiology*, 128(8):1445–1456, 2017. 6.1.5, 6.1.5.1, 6.3

[317] Emanuel N. Van Den Broeke, Julien Lambert, Gan Huang, and André Mouraux. Central sensitization of mechanical nociceptive pathways is associated with a long-lasting increase of pinprick-evoked brain potentials. *Frontiers in Human Neuroscience*, 10:531, 2016. 6.1.5, 6.3

[318] Emanuel N. van den Broeke, André Mouraux, Antonia H. Groneberg, Doreen B. Pfau, Rolf Detlef Treede, and Thomas Klein. Characterizing pinprick-evoked brain potentials before and after experimentally induced secondary hyperalgesia. *Journal of Neurophysiology*, 114(5):2672–2681, 2015. 6.1.5, 6.2.2, 6.3, 6.3.2

[319] Cees van der Pluijm and Hein Polzer. *Zeslettergreperigheid*. Nijgh & Van Ditmar, Amsterdam, 2009. (document)

[320] Marianne van der Vaart, Eugene Duff, Nader Raafat, Richard Rogers, Caroline Hartley, and Rebeccah Slater. Multimodal pain assessment improves

discrimination between noxious and non-noxious stimuli in infants. *Paediatric and Neonatal Pain*, 1(1):21–30, 2019. 2.1.1, 3.1, 3.2.3, 3.2.5.6, 3.2.5.7, 3.4.3, 5.1

[321] Marianne van der Vaart, Caroline Hartley, Luke Baxter, Gabriela Schmidt Mellado, Foteini Andritsou, Maria M Cobo, Ria Evans Fry, Eleri Adams, Sean Fitzgibbon, and Rebeccah Slater. Premature Infants Display Discriminable Behavioral, Physiological, and Brain Responses to Noxious and Nonnoxious Stimuli. *Cerebral Cortex*, page bhab449, 2021. 2.1, 3, 3.2, 3.2, 3.3, 3.6, 3.4, 3.7, 4, 4.3, 4.3.1.2, 4.5, 4.7, 4.9, 4.10

[322] Marianne van der Vaart, Caroline Hartley, Luke Baxter, Gabriela Schmidt Mellado, Maria M. Cobo, Ria Evans Fry, Eleri Adams, Sean Fitzgibbon, and Rebeccah Slater. Premature infants display discriminable behavioural, physiological and brain responses to noxious and non-noxious stimuli. *medRxiv*, page https://doi.org/10.1101/2021.08.18.21262106, 2021. 2.1, 3, 4

[323] Sampsa Vanhatalo, Aiman Alnajjar, Vinh T Nguyen, Paul Colditz, and Peter Fransson. Safety of EEG–fMRI recordings in newborn infants at 3T: A study using a baby-size phantom. *Clinical Neurophysiology*, 125(5):941–946, May 2014. 5.4

[324] Sampsa Vanhatalo, Veikko Jousmäki, Sture Andersson, and Marjo Metsäranta. An Easy and Practical Method for Routine, Bedside Testing of Somatosensory Systems in Extremely Low Birth Weight Infants. *Pediatric Research*, 66(6):710–713, December 2009. 4.4.2

[325] Sampsa Vanhatalo and Kai Kaila. Development of neonatal EEG activity: From phenomenology to physiology. *Seminars in Fetal and Neonatal Medicine*, 11(6):471–478, December 2006. 4.4.1

[326] Sampsa Vanhatalo, J. Matias Palva, Sture Andersson, Claudio Rivera, Juha Voipio, and Kai Kaila. Slow endogenous activity transients and developmental

expression of K+–Cl- cotransporter 2 in the immature human cortex. *European Journal of Neuroscience*, 22(11):2799–2804, December 2005. 4.4.1

[327] Lana Vasung, Claude Lepage, Milan Radoš, Mihovil Pletikos, Jennifer S. Goldman, Jonas Richiardi, Marina Raguž, Elda Fischi-Gómez, Sherif Karama, Petra S. Huppi, Alan C. Evans, and Ivica Kostovic. Quantitative and Qualitative Analysis of Transient Fetal Compartments during Prenatal Human Brain Development. *Frontiers in Neuroanatomy*, 10:11, 2016. 1.7.2, 8.3.1

[328] M. Verriotis, L. Fabrizi, A. Lee, R. J. Cooper, M. Fitzgerald, and J. Meek. Mapping Cortical Responses to Somatosensory Stimuli in Human Infants with Simultaneous Near-Infrared Spectroscopy and Event-Related Potential Recording. *eNeuro*, 3(2):1–15, 2016. 1.9.1.3, 3.2.5.1

[329] Madeleine Verriotis, Pishan Chang, Maria Fitzgerald, and Lorenzo Fabrizi. The development of the nociceptive brain. *Neuroscience*, 338:207–219, 2016. 3.1

[330] Madeleine Verriotis, Lorenzo Fabrizi, Amy Lee, Sheryl Ledwidge, Judith Meek, and Maria Fitzgerald. Cortical activity evoked by inoculation needle prick in infants up to one-year old. *Pain*, 156(2):222–230, February 2015. 1.9.1.3

[331] Madeleine Verriotis, Laura Jones, Kimberley Whitehead, Maria Laudiano-Dray, Ismini Panayotidis, Hemani Patel, Judith Meek, Lorenzo Fabrizi, and Maria Fitzgerald. The distribution of pain activity across the human neonatal brain is sex dependent. *NeuroImage*, 178(April):69–77, 2018. 4.2.5, 4.4.5

[332] 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. 1.3

[333] Tor D. Wager, Lauren Y. Atlas, Martin A. Lindquist, Mathieu Roy, Choong-Wan Woo, and Ethan Kross. An fMRI-Based Neurologic Signature of Physical Pain. *N Engl J Med*, 365(15):1388–1397, 2013. 1.7.2, 1.9.2.2

[334] Wojciech Walas, Zenon Halaba, Agata Kubiaczyk, Andrzej Piotrowski, Julita latka-Grot, Tomasz Szczapa, Monika Romul, Iwona Maroszyńska, Ewelina Malinowska, Magdalena Rutkowska, Michał Skrzypek, and Robert Śmigiel. Skin conductance measurement for the assessment of analgosedation adequacy in infants treated with mechanical ventilation: A multicenter pilot study. *Advances in Clinical and Experimental Medicine*, 29(9):1117–1121, September 2020. 8.2.2

[335] Wojciech Walas, Zenon P. Halaba, Tomasz Szczapa, Julita Latka-Grot, Iwona Maroszyńska, Ewelina Malinowska, Magdalena Rutkowska, Agata Kubiaczyk, Monika Wrońska, Michał Skrzypek, Julien De Jonckheere, Mickael Jean-Noel, and Andrzej Piotrowski. Procedural Pain Assessment in Infants Without Analgosedation: Comparison of Newborn Infant Parasympathetic Evaluation and Skin Conductance Activity - A Pilot Study. *Frontiers in Pediatrics*, 9:746504, 2022. 3.4.3, 8.2.2

[336] Wojciech Walas, Julita Latka-Grot, Tomasz Szczapa, Iwona Maroszyńska, Magdalena Rutkowska, Alicja Bartkowska-Śniatkowska, and Andrzej Piotrowski. Usefulness of Two Types of Pain Monitors in Newborns Treated in NICU, in The Opinion of Experts: Results of The Survey. *Journal of Mother and Child*, 25(2):72–76, April 2022. 8.2.2

[337] Marlene Walden, Joy Hinson Penticuff, Bonnie Stevens, Marilyn J. Lotas, Claudia A. Kozinetz, Angela Clark, and Kay C Avant. Maturational Changes in Physiologic and Behavioral Responses of Preterm Neonates To Pain. *Advances in Neonatal Care*, 1(2):94–106, 2001. 3.2.3

[338] S. M. 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. 1.3, 4.2.5

[339] Fabrice Wallois, Laura Routier, Claire Heberlé, Mahdi Mahmoudzadeh, Emilie Bourel-Ponchel, and Sahar Moghimi. Back to basics: The neuronal substrates and mechanisms that underlie the electroencephalogram in premature neonates. *Neurophysiologie Clinique*, 51(1):5–33, 2021. 1.7.2, 4.4.1

[340] E Walls-Esquivel, M.F. Vecchierini, C Héberlé, and F Wallois. Electroencephalography (EEG) recording techniques and artefact detection in early premature babies. *Neurophysiologie Clinique/Clinical Neurophysiology*, 37(5):299–309, October 2007. 3, 3.4.4.2

[341] V. Wanigasekera, K. Wartolowska, J. P. Huggins, E. P. Duff, W. Vennart, M. Whitlock, N. Massat, L. Pauer, P. Rogers, B. Hoggart, and I. Tracey. Disambiguating pharmacological mechanisms from placebo in neuropathic pain using functional neuroimaging. *British Journal of Anaesthesia*, 120(2):299–307, February 2018. 5.2

[342] Tracy Warbrick, Vera Fegers-Stollenwerk, Ivan I. Maximov, Farida Grinberg, and N. Jon Shah. Using Structural and Functional Brain Imaging to Investigate Responses to Acute Thermal Pain. *Journal of Pain*, 17(7):836–844, 2016. 5.5, 7.3.3.3

[343] Jordana A. Waxman, Rebecca R. Pillai Riddell, Paula Tablon, Louis A. Schmidt, and Angelina Pinhasov. Development of Cardiovascular indices of acute pain responding in infants: A systematic review. *Pain Research and Management*, 2016:8458696, 2016. 3.1, 3.4.2

[344] Lachlan Webb, Minna Kauppila, James A. Roberts, Sampsa Vanhatalo, and Nathan J. Stevenson. Automated detection of artefacts in neonatal EEG with residual neural networks. *Computer Methods and Programs in Biomedicine*, 208:106194, September 2021. 3.4.4.2

[345] K. Whitehead, C. Papadelis, M. P. Laudiano-Dray, J. Meek, and L. Fabrizi. The Emergence of Hierarchical Somatosensory Processing in Late Prematurity. *Cerebral cortex*, 29(5):2245–2260, 2019. 4.4.1, 8.3.1

[346] Kimberley Whitehead, Ronit Pressler, and Lorenzo Fabrizi. Characteristics and clinical significance of delta brushes in the EEG of premature infants. *Clinical Neurophysiology Practice*, 2:12–18, 2017. 1.7.2, 4.4.3

[347] Jean Claude Willer. Comparative study of perceived pain and nociceptive flexion reflex in man. *Pain*, 3(1):69–80, 1977. 1.8.1

[348] Amber L. Williams, Asif Z. Khattak, Christen N. Garza, and Robert E. Lasky. The behavioral pain response to heelstick in preterm neonates studied longitudinally: Description, development, determinants, and components. *Early Human Development*, 85(6):369–374, 2009. 4.1

[349] Gemma Williams, Lorenzo Fabrizi, Judith Meek, Deborah Jackson, Irene Tracey, Nicola Robertson, Rebeccah Slater, and Maria Fitzgerald. Functional magnetic resonance imaging can be used to explore tactile and nociceptive processing in the infant brain. *Acta Paediatrica, International Journal of Paediatrics*, 104(2):158–166, 2015. 1.7.2, 1.9.2.3

[350] Anderson M. Winkler, Gerard R. Ridgway, Matthew A. Webster, Stephen M. Smith, and Thomas E. Nichols. Permutation inference for the general linear model. *NeuroImage*, 92:381–397, 2014. 3.2.7.1, 3.2.7.2, 4.2.3, 4.2.4.1, 4.2.4.3, 6.1.3, 6.1.5.1, 6.1.5.2, 6.1.6, 6.1.7, 6.1.7.1, 6.1.7.1, 6.1.7.2, 7.1.3.2, 7.1.5, 7.1.6.3

[351] Choong Wan Woo, Luke J. Chang, Martin A. Lindquist, and Tor D. Wager. Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, 20(3):365–377, 2017. 1.9.2.2, 5.2

[352] Choong Wan Woo, Liane Schmidt, Anjali Krishnan, Marieke Jepma, Mathieu Roy, Martin A. Lindquist, Lauren Y. Atlas, and Tor D. Wager. Quantifying

cerebral contributions to pain beyond nociception. *Nature Communications*, 8:1–14, 2017. 1.7.2

[353] Charles D Woody. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and biological engineering volume*, 5:539–554, 1967. 3.2.5.2

[354] Mark W Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M Smith. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45:S173–S186, 2009. 6.1.6, 6.1.6, 7.1.6.1

[355] Mark W. Woolrich, Brian D. Ripley, Michael Brady, and Stephen M. Smith. Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, 14(6):1370–1386, 2001. 6.1.6, 7.1.4.2

[356] A Worley, L Fabrizi, S Boyd, and R Slater. Multi-modal pain measurements in infants. *Journal of Neuroscience Methods*, 205:252–257, 2012. 4, 2.1.1.2, 3.2.4

[357] Anna Xu, Bart Larsen, Erica B. Baller, J. Cobb Scott, Vaishnavi Sharma, Azeez Adebimpe, Allan I. Basbaum, Robert H. Dworkin, Robert R. Edwards, Clifford J. Woolf, Simon B. Eickhoff, Claudia R. Eickhoff, and Theodore D. Satterthwaite. Convergent neural representations of experimentally-induced acute pain in healthy volunteers: A large-scale fMRI meta-analysis. *Neuroscience and Biobehavioral Reviews*, 112:300–323, 2020. 1.6.4, 1.9.1.2, 1.9.2.2, 6.3, 6.3.1

[358] Tal Yarkoni, Russell A. Poldrack, Thomas E. Nichols, David C. Van Essen, and Tor D. Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665–670, August 2011. 1.9.2.2, 6.1.6

[359] Ghada Zamzmi, Rangachar Kasturi, and Dmitry Goldgof. A Review of Automated Pain Assessment in Infants: Features, Classification Tasks, and Databases. *IEEE Reviews in Biomedical Engineering*, 11:77–96, 2018. 3.1

[360] Ghada Zamzmi, Chih-Yun Pai, Dmitry Goldgof, Rangachar Kasturi, Yu Sun, and Terri Ashmeade. Automated pain assessment in neonates. In *Proc. 20th Scand. Conf. Image Anal.*, pages 350–361, 2017. 3.1

[361] Rushi Zou, Linling Li, Li Zhang, Gan Huang, Zhen Liang, and Zhiguo Zhang. Predicting Individual Pain Thresholds From Morphological Connectivity Using Structural MRI: A Multivariate Analysis Study. *Frontiers in Neuroscience*, 15:615944, 2021. 5.5