Investigating future closed-loop virtual reality therapeutics in chronic musculoskeletal pain: Biomarkers, machine learning and immersive technologies

By

Dr Jordan Tsigarides, MBBS, MClinEd

A thesis submitted for the degree:

Doctor of Philosophy

Norwich Medical School

Faculty of Medicine and Health Sciences

University of East Anglia

Norwich, United Kingdom

Date of submission: 20th June 2026



This copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Declaration

I declare that the contents of this thesis and the work presented within it are my own, unless otherwise acknowledged, and have been generated as the result of my own original research. I confirm that the work within this thesis has not been submitted in an application for another degree or qualification in this or any other university or institution.

Word count: 64,829

Abstract word count: 300

In accordance with the University of East Anglia Generative AI Policy for Research and Innovation, the generative AI tools ChatGPT (OpenAI) and Claude (Anthropic) were used selectively in a supporting capacity during preparation of this thesis.

Their use was limited to editorial assistance with author-drafted text, including clarity, grammar, consistency and identification of repetition; occasional assistance in identifying potentially relevant literature, which was independently retrieved and verified by the author; and generation of schematic illustrations to inform the design of Figures 2.1 and 2.2, which were subsequently edited and checked by the author.

The studies, analyses, interpretation of results and conclusions were developed and finalised by the author. Generative AI was not used to collect or generate research data or study results, and no participant-level, personal, confidential or sensitive data were entered into generative AI tools. The author takes full responsibility for the content, accuracy and integrity of the thesis.

Acknowledgements

This PhD has been a huge privilege to complete, and with the guidance of my supervisors and wider collaborative network, I've been able to address research questions that I am both passionate about and feel are important for the future of healthcare. However, commencement of this programme was not trivial. As with many academic careers, the preamble was fraught with funding challenges and disappointment at times. This only made me more determined, and I will be forever thankful for the relentless support that my primary supervisor Alex provided throughout.

To the senior research team at Norwich Medical School who kindly part-funded this work, thank you for believing in it and resourcing a passionate clinical academic trainee for the task. To Hamish Mackenzie at Revolve Labs, thank you for supporting this work with funding and for backing me from start to finish; you made this PhD a reality.

Thanks to my secondary supervisor Dr Jennifer Bowler for her guidance, and to Prof Tony Bagnall for his direct support, teaching, and patience with the machine learning components. A huge thank you to Prof Frances Williams and the wider TwinsUK team for supporting the PainWaves study, to Dr Nick Fallon at the University of Liverpool, a long-term collaborator, and to Dr Maté Gyurkovics at the UEA for their EEG expertise. For the VR-FOCUS and VIPA-EEG studies, thanks to the team at the Norwich Epidemiology Centre and Norfolk Arthritis Register, including Jackie, Jack and Karen. A special acknowledgement to Dr Max Yates for his mentorship and support over the years.

None of this would have been possible without the fantastic participants who gave their time, provided a great deal of kindness, and for those living with chronic pain, often battled through difficult times to participate. A massive thank you all.

Finally, to my family including my wife Rebecca and perfect 9-month-old Rupert. Thank you for making me smile during the stressful times, for your own sacrifices during the journey, and for always being there. I love you all.

Abstract

Chronic musculoskeletal pain is a major cause of disability and health service burden, yet current treatments are largely ineffective or inaccessible. Immersive virtual reality (VR) has shown potential as a tool for pain management but is typically delivered as an open-loop, non-personalised intervention. Adaptive closed-loop VR therapeutics could deliver accessible, personalised digital pain interventions, but development requires mechanistically relevant biomarkers, embedded physiological sensing, and predictive models to enable state-responsive adaptation.

This work investigated whether sensor-derived measures, analysed using statistical and machine learning approaches, can inform future closed-loop VR therapeutics for chronic musculoskeletal pain. Three empirical studies were undertaken. PainWaves applied resting-state electroencephalography (rsEEG) within a discordant co-twin design, identifying reduced frontocentral beta band power as the most consistent rsEEG feature of chronic widespread pain after partial control for shared genetic and environmental background, with weaker alpha-related findings. VIPA-EEG examined within-session rsEEG change around attention-based VR exposure in fibromyalgia, showing short-term increases in alpha and beta band power, with alpha providing the clearest residual oscillatory effect after adjustment for the aperiodic component. VR-FOCUS evaluated headset-derived eye tracking during a graded VR N-back task in healthy controls and adults with chronic musculoskeletal pain. Time-series machine learning classifiers outperformed summary-feature baselines and supported subject-independent inference of task-defined cognitive load, although performance was lower and more variable in chronic pain.

Findings from this work support beta band power as a candidate mechanistic rsEEG feature, alpha as a more state-responsive rhythm around VR exposure, and embedded eye tracking as a tractable signal for cognitive load inference. They also highlight that predictive models may not generalise as reliably to clinically heterogeneous chronic pain cohorts. Overall, this work provides methodological and conceptual foundations for future personalised closed-loop VR therapeutics in chronic pain, while clarifying that rsEEG, eye tracking and machine learning serve distinct roles within that pathway.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Thesis Statement:

This thesis examines chronic musculoskeletal pain through a translational framework spanning biomarker identification, experimental modulation, and sensor-based cognitive state prediction. It shows that pain-relevant oscillatory abnormalities can be characterised beyond shared genetic and environmental influences, that immersive interventions may influence these signals, and that sensor-enabled immersive systems offer a plausible route toward personalised closed-loop therapeutics.

Table of contents

Acknowledgements *3*

Abstract *4*

Table of contents *6*

Publications and conference presentations *12*

List of abbreviations *14*

List of tables *18*

List of figures *19*

List of appendices *21*

1. Introduction *22*

1.1 The clinical and translational problem **22**

1.2 Thesis position and framework **22**

1.3 Thesis structure **24**

2. Background *25*

2.1 Chronic pain as a multidimensional condition **25**

2.1.1 Chronic pain as a major clinical and societal challenge 25

2.1.2 Limitations of current management approaches 26

2.1.3 Mechanistic and biopsychosocial models of chronic pain 28

2.1.4 Attention, cognition, and pain modulation 31

2.2 Immersive therapeutics for pain: evidence and constraints **32**

2.2.1 Why immersive technologies? 32

2.2.2 Evidence for VR in pain: acute and chronic 35

2.2.3 Mechanisms of immersive analgesia 38

2.2.4 Limitations, variability and the case for adaptation 41

2.3 Closed-loop personalisation: rationale, biomarkers, sensors and design requirements **46**

2.3.1 Closed-loop therapeutics: framework and rationale 46

2.3.2 Mechanistic and state-sensitive biomarker roles 49

2.3.3 Resting-state EEG as a mechanistic measure 53
2.3.4 Eye tracking as a pragmatic embedded signal 58
2.3.5 Cognitive load as an adaptive target 62
2.3.6 Machine learning for state inference 66
2.3.7 Alternative peripheral signals, briefly considered 67
2.3.8 Design requirements for future adaptive systems 70
2.3.9 Rationale for the empirical studies 71
3. Aims 72
4. Core methods: EEG, eye tracking, and machine learning 73
4.1 EEG framework 73
4.1.1 Acquisition 73
4.1.2 EEG preprocessing pipeline 74
4.1.3 Quality-control framework 76
4.1.4 Spectral analysis 76
4.1.5 Aperiodic-periodic decomposition 77
4.2 Eye tracking and VR design framework 78
4.2.1 Hardware and software choices 78
4.2.2 Cognitive task framework 79
4.2.3 Signal considerations 80
4.2.4 Preprocessing 80
4.3 Predictive modelling and machine learning framework 82
4.3.1 Rationale 82
4.3.2 Model classes 83
4.3.3 Evaluation 86
4.3.4 Dimensionality, over-fitting and validation 87
4.4 Reproducibility and data management 88
5. Study 1: The PainWaves Study: Resting state EEG alterations in twin pairs discordant for chronic widespread pain 90
5.1 Introduction 90

5.2 Study aims **91**

5.3 Methods **92**

5.3.1 Study design and participants 92

5.3.2 Ethics and governance 93

5.3.3 EEG acquisition 93

5.3.4 Self-reported measures and participant characteristics 94

5.3.5 Data preparation and preprocessing 96

5.3.6 Processed data quality 97

5.3.7 Sample size and power 97

5.3.8 Inferential framework 97

5.3.9 Sensitivity analyses 98

5.3.10 Primary spectral analysis 98

5.3.11 Secondary aperiodic-periodic spectral parameterisation analysis 99

5.3.12 Secondary alpha peak analysis 100

5.3.13 Exploratory connectivity analysis 100

5.3.14 Exploratory microstate analysis 101

5.3.15 Clinical descriptive statistics and missing-data handling 103

5.3.16 Exploratory symptom-association analyses 103

5.4 Results **104**

5.4.1 Primary spectral analysis 106

5.4.2 Secondary aperiodic-periodic spectral parameterisation 109

5.4.3 Secondary alpha peak analysis 111

5.4.4 Exploratory functional connectivity 112

5.4.5 Exploratory microstate analysis 113

5.4.6 Exploratory symptom-association analyses 114

5.5 Discussion **115**

6. Study 2: VIPA-EEG: Investigating resting-state EEG changes following virtual reality exposure in fibromyalgia ***123***

6.1 Introduction **123**

6.2 Study aims **124**

6.3 Methods 125

6.3.1 Study design and participants 125

6.3.2 Ethics and governance 126

6.3.3 VR intervention context 126

6.3.4 EEG acquisition 128

6.3.5 Self-reported measures 130

6.3.6 Data preparation and preprocessing 130

6.3.7 Processed data quality 131

6.3.8 Sample size and power 131

6.3.9 Inferential framework 132

6.3.10 Sensitivity analyses 132

6.3.11 Primary spectral analysis 133

6.3.12 Secondary aperiodic-periodic analysis 134

6.3.13 Secondary alpha peak analysis 135

6.3.14 Repeated-measures modelling of post-activity EEG effects 135

6.3.15 Clinical descriptives and missing-data handling 136

6.3.16 Models to evaluate EEG-clinical associations 136

6.4 Results 137

6.4.1 Primary spectral analysis 138

6.4.2 Secondary aperiodic-periodic spectral parameterisation 143

6.4.3 Secondary alpha peak analyses 145

6.4.4 EEG-clinical association analyses 146

6.5 Discussion 147

7. Study 3: The VR-FOCUS Study: Investigating eye tracking during a virtual reality N-back task as a predictor of cognitive load in chronic pain 153

7.1 Introduction 154

7.2 Study aims 156

7.3 Methods 156

7.3.1 Study design and participants 156

7.3.2 Ethics and governance 158

7.3.3 VR apparatus and implementation 159
7.3.4 VR N-back task design and virtual environment 159
7.3.5 Data logging 163
7.3.6 Experimental procedure 163
7.3.7 Measures 164
7.3.8 Sample size and power 167
7.3.9 Data preprocessing and feature extraction 167
7.3.10 Processed eye tracking data quality 168
7.3.11 Inferential framework 169
7.3.12 Statistical analysis 170
7.3.13 Predictive models 170
7.4 Results 172
7.4.1 Descriptive task-level outcomes across VR N-back levels 173
7.4.2 Pupil diameter and blink rate 176
7.4.3 Predictive models 177
7.4.4 Adverse effects and task acceptability 181
7.5 Discussion 181
8. Synthesis 188
8.1 Outcomes from the empirical studies 188
8.1.1 Beta band rsEEG activity 188
8.1.2 Alpha as a more state-responsive rhythm 190
8.1.3 VR-embedded eye tracking as a signal for state inference 190
8.1.4 Machine learning in the VR-FOCUS study 191
8.2 Implications for the design of future closed-loop VR therapeutics 193
8.2.1 Informing future system design requirements 193
8.2.2 Sensor selection in light of the present findings 193
8.2.3 Adaptation targets and decision logic 194
8.2.4 Personalisation and the limits of population-level inference 196
8.3 Methodological contributions 197
8.3.1 Automated reproducible EEG pipelines 197

8.3.2 Pragmatic acquisition approaches 198
8.3.3 Periodic-aperiodic separation as a routine standard 198
8.3.4 Subject-independent evaluation in pragmatic machine learning 199
8.4 Cumulative limitations of the empirical work 199
8.5 Priorities for future research 202
9. Conclusions 204
9.1 How the empirical findings addressed the aims of this work 204
9.2 How I would have approached the research differently 204
9.3 How the field has changed 205
9.4 The next phase of research: towards clinically deployable closed-loop VR therapeutics 207
9.5 Principal conclusions 207
10. References 209
11. Appendices 241
Appendix 1: Published parent feasibility study conducted alongside the VIPA-EEG study 241
Appendix 2: Description of study-specific questionnaires and outcome measures 268
Appendix 3: Additional rsEEG data and figures from PainWaves and VIPA-EEG 276
Appendix 4: Additional performance and self-reported data from VR-FOCUS 282
Appendix 5: Ethics Approval Forms 287
Appendix 6: Consent Forms 299
Appendix 7: Participant Information Sheets 305

Publications and conference presentations

Research published or under peer-review directly arising from this work

1. **Tsigarides J**, Bowler J, Dainty J, Bagnall A, Macgregor A. VR-FOCUS: Investigating eye tracking during a virtual reality N-back task as a predictor of cognitive load in chronic pain. Research Square [Preprint]. 2026. doi:10.21203/rs.3.rs-8680786/v1. Under review at Scientific Reports.
2. **Tsigarides J**, Kannuralil Sabu B, Al-Amri M, Hughes S, Fuggle N. Therapeutic Virtual Reality in Rheumatology: A Narrative Review. touchREVIEWS in RMD. 2026;5(1):26-34. doi:10.17925/rmd.2026.5.1.2.

Research publications in preparation

1. **Tsigarides J**, Gyurkovics M, Fallon N, Dainty J, Macgregor A, Williams F. Resting-state EEG alterations in monozygotic and dizygotic twin pairs discordant for chronic widespread pain. **(Chapter 5)**
2. **Tsigarides J**, Grove V, Shenker N, Sami S, MacGregor A. Effect of an Attention-based Virtual Reality Intervention on EEG signals in Fibromyalgia Syndrome. **(Chapter 6)**

Prior publications underpinning this work

1. **Tsigarides J**, Grove V, Chipping J, Dainty J, Miles S, Shenker N, Sami S, MacGregor A. Feasibility of attention-based virtual reality interventions in fibromyalgia syndrome: comparing systems, virtual environments and activities. British Journal of Pain. 2025;19(2):71-85. doi:10.1177/20494637241310696. (Copy included in Appendix 1)
2. **Tsigarides J**, Shenker NG, MacGregor A. Ready Patient One: the role of therapeutic virtual reality in the future management of chronic pain. Rheumatology. 2022;61(2):482-483. doi:10.1093/rheumatology/keab594.

Related publications and conference presentations

1. **Tsigarides J**, Rushbrooke A, Gyurkovics M, Bowyer V, Freydin M, Bagnall A, MacGregor A, Williams F. Neurophysiology of chronic widespread pain: time-series classification of resting-state EEG in pain discordant twin pairs from TwinsUK. Presented poster tour presentation at EULAR 2026; London, United Kingdom.

2. Ruiz-Mateos Serrano R, Troughton JG, Mirkhani N, Martinez N, Mariello M, **Tsigarides J**, Williamson S, Sapriza J, Susnoschi Luca I, Dominguez-Alfaro A, Cuttaz E, Thompson N, Swedick S, Almulla L, Guemes A. From Neural Sensing to Stimulation: An Interdisciplinary Roadmap for Neurotechnology. *arXiv* [Preprint]. 2025. doi:10.48550/arXiv.2510.07116. Accepted by *Science Advances*.
3. Scott IC, Babatunde O, Barker C, Beesley R, Beesley R, Birkinshaw H, Brooke M, Chaplin H, Chapman L, Ciurtin C, Dale J, Dockrell D, Dures E, Harrison K, Jani M, Lee C, McCarron M, Mallen CD, O'Connor A, Pidgeon C, Pincus T, Pratt D, Prior Y, Raza K, Rutter-Locher Z, Sharma S, Shaw K, Small S, Smith T, Tiffin L, **Tsigarides J**, Xenophontos M, Shenker NG. Pain management in people with inflammatory arthritis: British Society for Rheumatology guideline scope. *Rheumatology Advances in Practice*. 2024;8(4):rkae128. doi:10.1093/rap/rkae128.
4. Cooray S, Deng A, Dong T, Hammouche S A, Quansah M, **Tsigarides J**, Fuggle N. Artificial intelligence in musculoskeletal medicine: a scoping review. touchREVIEWS in RMD. 2024;4(1):15-23. doi:10.17925/rmd.2025.4.1.3
5. **Tsigarides J**, Rushbrooke A, Bagnall A. POS0257 PainWaves: the potential of machine learning to differentiate chronic pain cohorts using electroencephalography. *Annals of the Rheumatic Diseases*. 2023;82(Suppl 1):366.
6. Rushbrooke A, **Tsigarides J**, Sami S, Bagnall A. Time Series Classification of Electroencephalography Data. In: Rojas I, Joya G, Catala A, editors. *Advances in Computational Intelligence. IWANN 2023. Lecture Notes in Computer Science*. Vol 14134. Cham: Springer; 2023. p.601-613. doi:10.1007/978-3-031-43085-5_48.
7. **Tsigarides J**. E083 Creation of a National Co-Development Hub for digital health interventions in chronic pain. *Rheumatology*. 2026;65(Suppl 2):keag121.306. doi:10.1093/rheumatology/keag121.306.
8. **Tsigarides J**, James D, von Fragstein M, Cargill D. OA05 Virtual reality to educate and empower primary care practitioners in chronic pain management: the INSIGHT Programme. *Rheumatology*. 2025;64(Suppl 3):keaf142.005. doi:10.1093/rheumatology/keaf142.005.

List of abbreviations

AEC	Amplitude envelope correlation
ASR	Artifact Subspace Reconstruction
AUC	Area under the curve
BIDS	Brain Imaging Data Structure
CBT	Cognitive behavioural therapy
CI	Confidence interval
cLBP	Chronic low back pain
CSD	Current source density
CSV	Comma-separated values
CWP	Chronic widespread pain
DrCIF	Diverse Representation Canonical Interval Forest
DZ	Dizygotic
EC	Eyes-closed
EDA	Electrodermal activity
EEG	Electroencephalography
EMG	Electromyography
EO	Eyes-open
FDA	Food and Drug Administration
FIQR	Revised Fibromyalgia Impact Questionnaire
FIR	Finite impulse response
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near-infrared spectroscopy
FOOOF	Fitting Oscillations and One-Over-F

FS	Fibromyalgia Severity
GABA	Gamma-aminobutyric acid
GAD-7	Generalised Anxiety Disorder-7
GDPR	General Data Protection Regulation
GEV	Global explained variance
GFP	Global field power
HC	Healthy control
HIVE-COTE	Hierarchical Vote Collective of Transformation-based Ensembles
HMD	Head-mounted display
HRV	Heart rate variability
IAF	Individual alpha frequency
IPD	Interpupillary distance
IQR	Interquartile range
IRAS	Integrated Research Application System
LFESSQ	London Fibromyalgia Epidemiology Study Screening Questionnaire
LOSO	Leave-one-subject-out
MAE	Mean absolute error
MEG	Magnetoencephalography
MPQ-SF	Short-form McGill Pain Questionnaire
MRC	Medical Research Council
MRI	Magnetic resonance imaging
MZ	Monozygotic
NASA-TLX	NASA Task Load Index
NHS	National Health Service

NIR	Near-Infrared
NOAR	Norfolk Arthritis Register
NRS	Numerical rating scale
NSAIDs	Non-steroidal anti-inflammatory drugs
PAG	Periaqueductal grey
PAAS	PAAS Cognitive Load Rating Scale
PCS	Pain Catastrophizing Scale
PHQ-2	Patient Health Questionnaire-2
PHQ-9	Patient Health Questionnaire-9
PPG	Photoplethysmography
PROMIS	Patient-Reported Outcomes Measurement Information System
PSD	Power spectral density
REC	Research Ethics Committee
ROC	Receiver operating characteristic
ROI	Region of interest
ROCKET	Random Convolutional Kernel Transform
RVM	Rostroventromedial medulla
SD	Standard deviation
SDK	Software development kit
SNRIs	Serotonin-norepinephrine reuptake inhibitors
SSRIs	Selective serotonin reuptake inhibitors
SSS	Symptom Severity Scale
STC	Shapelet Transform Classifier

TDE	Temporal Dictionary Ensemble
TEPRs	Task-evoked pupillary responses
UEA	University of East Anglia
UK	United Kingdom
VAS	Visual analogue scale
VR	Virtual reality
VRSQ	Virtual Reality Sickness Questionnaire
WPI	Widespread Pain Index
XR	Extended reality

List of tables

Table 5.1. Participant characteristics for the PainWaves study 104

Table 5.2. Within-pair spectral effect sizes and cluster-level p-values across primary and sensitivity quality-control tiers 107

Table 6.1. Comparison of the VR activities investigated in the parent feasibility study[8] 127

Table 6.2. Baseline participant characteristics for the VIPA-EEG cohort..... 138

Table 6.3. Pre versus post-intervention spectral effect sizes and cluster-level p-values across primary and sensitivity quality-control tiers 141

Table 7.1. Baseline participant characteristics for the VR-FOCUS cohort..... 173

Table 7.2. Classifier accuracy metrics for 1-Back and 4-Back class labels with left pupil size as signal input 179

Table A2.1. Example eye tracking CSV output from a single VR N-back task 274

Table A3.1. PainWaves quality-control tier composition 276

Table A3.2. PainWaves zygosity-stratified spectral results 277

Table A3.3. PainWaves exploratory connectivity summary 278

Table A3.6. VIPA-EEG quality-control characteristics by recording state..... 281

Table A4.1. Exploratory between-group comparisons for task performance indices 282

Table A4.2. Participant-reported questionnaire data by group and task..... 283

Table A4.3. VRSQ tolerability scores by group 285

Table A4.4. Exploratory linear regression associations between clinical factors and ocular outcomes in the chronic pain group 285

List of figures

Figure 2.1. The biopsychosocial model of pain 29

Figure 2.2. Schematic architecture of an untethered VR headset 33

Figure 2.3. Closed-loop framework for adaptive VR therapeutics 49

Figure 2.4. Aperiodic separation in EEG spectral analysis 57

Figure 4.1. EEG acquisition hardware used across the rsEEG studies 74

Figure 4.2. AEON time-series classifier families evaluated for VR eye tracking cognitive load classification 85

Figure 5.1. Group-level resting-state power spectral density across quality-control tiers 108

Figure 5.2. Band power topographies for alpha and beta bands 108

Figure 5.3. Aperiodic-adjusted frontocentral broad beta power in co-twins discordant for chronic widespread pain.....110

Figure 5.4. Within-pair difference in aperiodic-adjusted broad beta power in co-twins discordant for chronic widespread pain 110

Figure 6.1. Pre and post-intervention resting-state power spectral density across quality-control tiers..... 141

Figure 6.2. Pre-post intervention differences in alpha and beta band power topographies 142

Figure 6.3. Repeated post-activity alpha, broad beta and low beta effect sizes across quality-control tiers..... 142

Figure 6.4. Alpha and beta band power topographies across sequential activity contrasts..... 143

Figure 6.5. Pre-post increase in aperiodic-adjusted alpha power following VR intervention. 144

Figure 6.6. Aperiodic-adjusted frontocentral alpha power post-intervention. 145

Figure 7.1. Overview of VR-FOCUS experimental setup, in-headset task display, and N-back procedure..... 163

Figure 7.2. Performance indices across tasks and between groups..... 174

Figure 7.3. Participant-reported post-task measures across tasks and between groups..... 175

Figure 7.4. Trends in pupil diameter and blink rate across N-back levels and between groups 176

Figure 7.5: ROC curves for 1-Back vs 4-Back across HIVE-COTE 2.0, Rotation Forest and DrCIF classifiers (left pupil size only)..... 180

Figure 7.6: Normalised confusion matrices for 1-Back vs 4-Back ROCKET classification (left pupil size only)..... 180

Figure A3.1. PainWaves supplementary spectral topographies across all frequency bands 279

List of appendices

Appendix 1: Published parent feasibility study conducted alongside the VIPA-EEG study 241

Appendix 2: Description of study-specific questionnaires and outcome measures 268

Appendix 3: Additional rsEEG data and figures from PainWaves and VIPA-EEG 276

Appendix 4: Additional performance and self-reported data from VR-FOCUS 282

Appendix 5: Ethics Approval Forms 287

Appendix 6: Consent Forms 299

Appendix 7: Participant Information Sheets 305

1. Introduction

1.1 The clinical and translational problem

Chronic musculoskeletal pain affects a substantial proportion of the adult population and remains a leading cause of disability worldwide[1-3]. Despite the availability of pharmacological and non-pharmacological treatment options, outcomes remain variable, and no current treatment pathway offers durable and scalable benefit across diverse patient populations[4]. Recommended non-pharmacological approaches such as graded exercise and cognitive behavioural therapy have additional challenges, notably limited access and the need for sustained compliance. The result is a long-term condition characterised by high prevalence, substantial disability and recurrent healthcare demand, for which current treatment models are poorly aligned with the clinical reality of the patients' experience.

Virtual reality (VR) has emerged as a promising technology with therapeutic potential within healthcare[5, 6]. Immersive systems can deliver controlled, multimodal experiences that influence attention, sensory context, affective state, embodiment and behavioural engagement, and can support delivery outside specialist clinical environments. Existing evidence in chronic pain demonstrates feasibility with encouraging signals of efficacy across paradigms[7, 8]. The clinical and conceptual context for this evidence, and the limitations that constrain its translation, are developed in Chapter 2.

A consistent limitation across the existing literature is that current VR systems for chronic pain operate as essentially static interventions. Therapeutic content is selected in advance, delivered through pre-specified sessions, and evaluated through outcomes measured before and after exposure[5]. This model is poorly matched to chronic musculoskeletal pain, which is characterised by heterogeneous symptom expression between individuals and substantial fluctuation in pain, fatigue, attention, affect and cognitive capacity within individuals over time[9-11]. A static intervention is unlikely to remain optimally calibrated across this variation.

1.2 Thesis position and framework

This thesis argues that immersive therapeutics for chronic musculoskeletal pain are constrained by current static delivery models, and that further progress requires a

deliberate transition toward closed-loop adaptive systems informed by sensor-derived biomarkers.

Credible adaptive immersive therapeutics require three separable competencies. Each can be developed independently, although their eventual integration remains the translational goal. The first is mechanistic characterisation, which identifies what may be biologically altered in chronic pain and is therefore worth targeting. The second is state-sensitive measurement, which captures the user's current functional or cognitive state during intervention delivery in a way that can inform adaptation. The third is predictive inference, which converts rich, temporally structured sensor data into decision-relevant estimates of user state. The three competencies are governed by different selection criteria, served by different signal types, and developed through different methodological strategies. They have often been conflated in the digital therapeutics literature, where a single signal is sometimes expected to perform all three roles at once[12-15]. The framework advanced here treats them as distinct translational tasks.

The framework therefore implies that adaptive immersive therapeutics will require several sensing or analytical modalities, each chosen for a specific translational role, and that expecting one modality to serve all three roles will continue to produce evidence that is neither mechanistically interpretable nor practically deployable. The criteria distinguishing mechanistic from embedded sensing are developed in Chapter 2.

The empirical work is structured around this framework. Because chronic musculoskeletal pain spans regional presentations through to widespread, centrally mediated phenotypes, the population studied is matched to the role of each study rather than held constant. Resting-state electroencephalography (rsEEG) is used to investigate mechanistic neurophysiology in chronic widespread pain (CWP) and fibromyalgia, the phenotypes in which brain-based oscillatory abnormality is most plausible. It is applied first in a co-twin control study that accounts for the genetic and environmental factors that influence EEG, and then in a repeated-measures study that tests whether candidate features change around immersive VR exposure. Headset-derived eye tracking is used to test whether dynamic ocular signals acquired during a controlled VR cognitive task support subject-independent inference of task-defined state in healthy controls and a broader chronic

musculoskeletal pain sample, reflecting the clinical heterogeneity any deployable signal must tolerate. Time-series machine learning serves as the analytical tool that takes temporal signals and provides a prediction of current user state to guide adaptation. The work does not attempt to deliver a fully realised closed-loop VR therapeutic; it instead establishes the foundations on which such systems can credibly be built.

1.3 Thesis structure

The thesis comprises nine chapters organised to move from clinical and conceptual context, through aims and methodological framework, to empirical investigation and integrative synthesis.

Chapter 2 develops the background and theoretical context for the work. It examines chronic pain as a heterogeneous and dynamic condition, reviews the evidence base and limitations of current immersive therapeutics, and sets out the rationale for closed-loop personalisation. It introduces the conceptual distinction between mechanistic and state-sensitive biomarkers in detail and considers the design requirements of pragmatic adaptive systems.

Chapter 3 sets out the aims of the thesis in light of the framework introduced here and the context developed in Chapter 2. Study-specific aims are stated at the beginning of the corresponding empirical chapters.

Chapter 4 outlines the methodological framework that links the empirical work, with emphasis on the technologies, software platforms, analytical pipelines and statistical approaches across the studies. This includes general details related to the EEG preprocessing pipeline, periodic-aperiodic spectral decomposition, eye tracking acquisition framework and the time-series machine learning approach. Study-specific implementation is reported in the corresponding empirical chapters.

Chapter 5 (PainWaves) investigates rsEEG differences in monozygotic (MZ) and dizygotic (DZ) twin pairs discordant for CWP. The discordant co-twin control design partly accounts for the genetic and environmental influences on rsEEG traits and supports clearer interpretation of neurophysiological features in CWP.

Chapter 6 (VIPA-EEG) examines whether candidate rsEEG signals in fibromyalgia change around immersive VR exposure. The chapter is presented as a secondary analysis of a parent feasibility study[8] and investigates whether relevant rsEEG

signals identified from PainWaves and the wider literature change with an attention-based VR intervention.

Chapter 7 (VR-FOCUS) evaluates whether headset-derived eye tracking acquired during a graded VR N-back task supports subject-independent inference of task-defined cognitive state in chronic pain and healthy control participants, comparing conventional and time-series machine learning approaches.

Chapter 8 synthesises findings across studies, considers points of convergence and uncertainty, and develops the implications for biomarker interpretation, sensor selection, state inference, candidate adaptation targets and methodological practice.

Chapter 9 concludes the thesis, reflecting on the decisions made, progress in the field, and the follow-on work required.

2. Background

2.1 Chronic pain as a multidimensional condition

2.1.1 Chronic pain as a major clinical and societal challenge

Chronic pain, commonly defined as pain that persists or recurs for longer than 3 months, is a highly prevalent long-term condition and a leading contributor to global disability[16, 17]. In the United Kingdom (UK), chronic pain affects a substantial proportion of the adult population. Estimates from Arthritis UK suggest that approximately 34% of adults experience chronic pain, with around 12% reporting high-impact pain that substantially impacts daily functioning[1]. Chronic musculoskeletal pain represents the largest and most relevant subgroup, accounting for a significant proportion of disability, healthcare utilisation, and long-term work absence[17].

The impact of chronic pain extends beyond symptom persistence to encompass wide-ranging effects on physical, psychological, and social functioning. Individuals with chronic pain frequently report limitations in mobility, reduced participation in employment and social roles, and diminished quality of life[2, 18]. At a population level, musculoskeletal conditions such as low back pain are consistently identified as leading causes of years lived with disability worldwide, reflecting both their high prevalence and chronicity[3]. These functional consequences are often compounded by comorbid fatigue, mood disturbance, and poor sleep, further amplifying the

burden experienced by individuals and contributing to complex patterns of healthcare need.

From a health-system perspective, chronic pain places considerable and sustained demand on services. It is one of the most common reasons for primary care consultation, and individuals with chronic pain have higher rates of healthcare utilisation, including increased general practitioner visits, specialist referrals, and hospital admissions[19]. Within the UK, musculoskeletal pain accounts for a substantial proportion of general practice workload, with millions of consultations annually attributed to conditions such as back pain, osteoarthritis, and widespread pain syndromes. This demand contributes to pressure on already constrained services, including specialist pain clinics, where waiting times frequently exceed recommended standards and access to multidisciplinary care remains variable[1, 20].

The burden of chronic pain also extends beyond the individual to affect informal caregivers and wider social networks. Many individuals with chronic pain require support with activities of daily living, particularly where pain is accompanied by disability or multimorbidity. Caregivers often report adverse effects on their own physical and mental health, including increased stress, reduced wellbeing, and limitations in their own daily activities[21-23]. These broader impacts reinforce the understanding of chronic pain as a condition with significant societal implications, rather than one confined to individual symptom experience.

Chronic pain should therefore be understood as a prevalent, disabling long-term condition that generates sustained demand on healthcare services, particularly in musculoskeletal populations. This burden highlights the need for interventions that are scalable and accessible beyond specialist clinical settings, with meaningful therapeutic impact across multiple domains important to people living with pain.

2.1.2 Limitations of current management approaches

Management of chronic musculoskeletal pain remains challenging, with most available treatments commonly providing modest or variable positive effects. Clinical guidance recommends a multimodal approach that may include pharmacological and non-pharmacological strategies, yet across major intervention classes, effect sizes are often small and no single treatment has demonstrated consistently strong and durable benefit across diverse patient populations[4, 24-26]. As a result, treatment

pathways are frequently characterised by partial symptom relief, ongoing functional impairment, and repeated healthcare utilisation.

Pharmacological therapies are widely used but are associated with important limitations. Analgesics such as paracetamol and non-steroidal anti-inflammatory drugs (NSAIDs) may offer short-term symptomatic relief in certain musculoskeletal pain conditions, but evidence of benefit in chronic pain is poor. In these populations, long-term efficacy has not been established and potential benefit must be weighed against recognised harms, particularly the gastrointestinal, renal, and cardiovascular risks associated with NSAIDs[27, 28]. Opioids, despite historically widespread use in chronic non-cancer pain, have been shown to confer only small improvements in pain and function, with increasing evidence of dose-dependent harms including dependence, tolerance, and increased risk of overdose[29, 30]. These concerns have led to a shift in guideline recommendations away from routine opioid use in chronic pain[4, 28].

Non-pharmacological interventions, including structured exercise programmes and psychological therapies such as cognitive behavioural therapy (CBT), are recommended as first-line management options for chronic pain[4, 28]. These approaches address biopsychosocial dimensions of pain-related disability, including physical deconditioning, fear of movement and activity avoidance, maladaptive pain-related psychology, and reduced self-efficacy. Despite demonstrating small-to-moderate improvements in pain and function in clinical trials[25, 26], their implementation at scale is limited by practical and service-related constraints. Many interventions require sustained engagement with healthcare professionals, repeated in-person sessions, and individual tailoring, all of which contribute to variability in access, adherence, and effectiveness in real-world settings[31, 32]. In practice, this leads to continued reliance on pharmacological management, with UK observational data demonstrating substantial ongoing opioid prescribing across long-term musculoskeletal conditions despite increasing recognition of the limited efficacy and potential harms of these treatments[33, 34].

A further challenge lies in the heterogeneity of treatment response. Chronic musculoskeletal pain encompasses a broad spectrum of conditions and presentations, with substantial variability in symptom profile, comorbidity, and psychosocial context. More broadly, pain itself is characterised by marked inter-

individual differences in susceptibility, severity, and response to treatment[9]. Evidence from musculoskeletal pain trials indicates that average treatment effects can obscure clinically meaningful inter-individual variation, with some patients deriving substantial benefit and others little or no improvement[35]. This variability is further compounded by within-person symptom fluctuation over time. Longitudinal studies in musculoskeletal pain populations have shown that pain varies meaningfully from week-to-week and that these fluctuations are associated with changes in fatigue, mood, sleep, stiffness, daily functioning, and healthcare use[10]. Similar patterns of intra-individual variability have also been demonstrated in fibromyalgia, including fluctuations in cognitive function alongside pain, fatigue, and affective symptoms[11, 36].

These limitations highlight a fundamental mismatch between current treatment models and the clinical reality of chronic musculoskeletal pain. Existing approaches are often insufficiently effective, difficult to deliver at scale, and poorly adapted to individual variability and temporal fluctuation. These constraints support the exploration of alternative therapeutic strategies that are both scalable and capable of responding to the dynamic and multidimensional nature of chronic pain.

2.1.3 Mechanistic and biopsychosocial models of chronic pain

Chronic musculoskeletal pain cannot be explained solely by peripheral nociceptive input or structural pathology. Contemporary models emphasise the interaction between sensory, cognitive, affective, and social processes in shaping the experience of pain, with considerable variability both between individuals and within individuals over time[18, 37]. This perspective reflects a shift away from reductionist biomedical frameworks toward integrative models that better account for the complexity and persistence of chronic pain.

The biopsychosocial model provides a widely adopted conceptual framework for understanding this complexity[38]. Within this model, biological factors such as tissue pathology and neurophysiological processes interact with psychological factors including attention, beliefs, and affect, and with social and environmental factors such as relationships, work demands, healthcare access, and cultural expectations (Figure 2.1). These domains interact dynamically rather than being independent, influencing both the intensity of pain and its impact on function and quality of life[9, 18]. Pain-related fear and maladaptive beliefs can amplify perceived threat and promote

avoidance behaviours, while social context and prior experience may shape expectations and coping strategies[39].

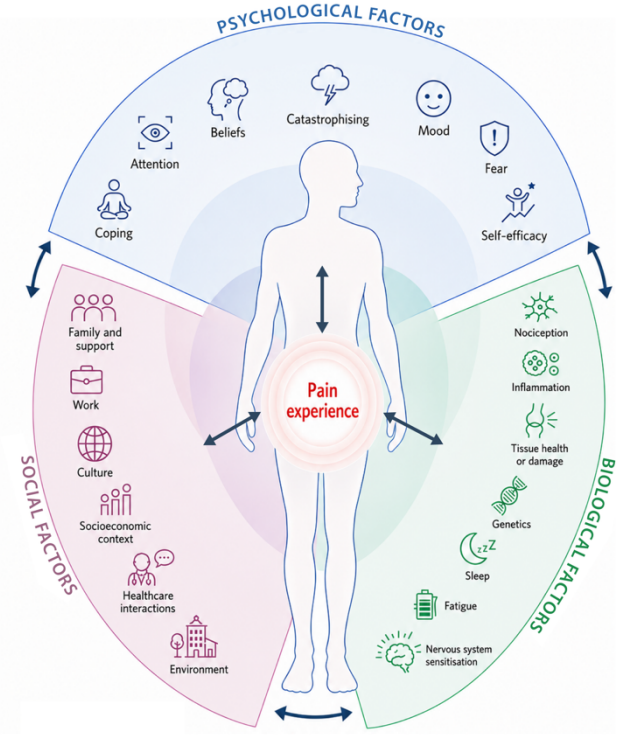


Figure 2.1. The biopsychosocial model of pain

Pain is represented as a multidimensional experience arising from interacting biological, psychological and social influences. The model emphasises that pain perception, persistence and impact are shaped by dynamic interactions across these domains rather than by peripheral tissue pathology alone.

Advances in pain classification have led to the recognition of ‘nociplastic pain’ as a distinct mechanistic descriptor. Nociplastic pain refers to pain arising from altered nociceptive processing despite no clear evidence of ongoing tissue damage or disease sufficient to explain symptom severity, and in the absence of defining features of neuropathic pain[40]. Within this framework, altered nociceptive processing may involve peripheral or central sensitisation. Peripheral sensitisation refers to increased responsiveness of nociceptive afferents, whereas central sensitisation refers to amplification of signalling within the central nervous system. In fibromyalgia and chronic widespread pain (CWP), symptom expression is often understood to be dominated by alterations in central pain processing, including enhanced facilitation and impaired descending inhibitory control within pain-modulatory systems, while recognising that peripheral sensitisation may also contribute and interact with central

mechanisms[41]. Population-level neuroimaging work has linked Fibromyalgia Index scores to altered functional and structural connectivity within the descending pain modulatory system, supporting the neurobiological relevance of nociplastic symptom burden in a large UK cohort[42]. More broadly, neuroimaging findings reinforce the role of distributed brain and brainstem networks that integrate nociceptive, cognitive, and affective influences in shaping the pain experience[43]. Complementary neurophysiological approaches, including electroencephalography (EEG), have been used to characterise altered resting-state neural activity in chronic pain, supporting the view that persistent pain is associated with measurable changes in central nervous system function beyond structural pathology alone[44]. Nociplastic pain therefore draws attention to the limitations of purely peripheral or structural explanations and emphasises the importance of central and systemic contributors to chronic pain.

These mechanistic insights support the view that pain is not simply transmitted, but actively evaluated, inferred, and behaviourally regulated. Neurocomputational perspectives conceptualise pain as an active inferential and learning process rather than a passive reflection of nociception. Pain is embedded within broader reinforcement learning systems that encode prediction, uncertainty, and value, with aversive prediction errors guiding behavioural adaptation[45]. Within this framework, chronic pain may reflect persistent or maladaptive updating of threat-related predictions, resulting in sustained pain experience even in the absence of ongoing tissue damage[46]. This perspective offers a link between sensory, cognitive, and behavioural domains, and an explanation for why pain is both highly individual and dynamically modulated over time.

Cognitive and affective processes are central to multidimensional frameworks of chronic pain. Pain-related psychology, including catastrophising, threat appraisal, and expectations about pain, influences pain intensity and disability[47]. Affective states such as anxiety, depression, and distress are closely linked to pain outcomes within a bidirectional relationship[48, 49]. These processes are supported by overlapping neural systems involving regions such as the anterior cingulate cortex, insula, and prefrontal cortex, which are implicated in pain processing and broader cognitive-affective regulation[50].

Chronic musculoskeletal pain is therefore best conceptualised as a multidimensional and dynamic condition in which biological, psychological, and social processes interact to shape symptom experience and functional outcomes. Treatments that target a single domain, or assume a stable and uniform patient state, are unlikely to capture the complexity of chronic pain or its variability across individuals and over time. Domain-specific measures remain essential for identifying underlying mechanisms, but effective therapeutic strategies may require integration with approaches capable of capturing and responding to dynamic, state-dependent aspects of the pain experience.

2.1.4 Attention, cognition, and pain modulation

Attention and broader cognitive processes shape the perception and impact of pain. Rather than reflecting a fixed output of nociception, pain is dynamically modulated by the allocation of cognitive resources, expectations, and ongoing task demands[43, 51, 52]. This modulation is relevant in chronic musculoskeletal pain, where central processes are a major contributor to the experience of pain and its associated disability.

Early theoretical models, most notably the gate control theory of pain, proposed that cognitive and affective factors could influence the transmission of nociceptive signals at the spinal level[53]. Subsequent work has refined this framework, but the core insight that pain is modifiable by higher-order processes has remained fundamental. Contemporary models emphasise distributed brain networks that integrate nociceptive input with contextual, emotional, and cognitive information to shape the pain experience[43, 54].

Empirical evidence shows that attentional allocation has a measurable impact on pain perception. Directing attention toward pain can amplify its perceived intensity, whereas engaging in cognitively demanding tasks can attenuate pain perception, a phenomenon commonly described as attentional analgesia[55-57]. Experimental paradigms have shown that increasing cognitive demand can reduce pain ratings under conditions where task engagement is sufficient, while also altering neural responses within pain-related regions including the anterior cingulate cortex, insula, and prefrontal cortex[58]. Neuroimaging evidence also indicates that pain-related and cognition-related networks can remain concurrently active, with the degree of interaction dependent on the relative demands imposed by pain intensity and

cognitive task load[59, 60]. This is consistent with broader evidence that pain interferes with cognitive function across attention, working memory, and executive control, while cognitive engagement modulates the salience and processing of pain[61].

At a neural level, attentional modulation of pain is mediated by interactions between cortical and brainstem systems, including descending modulatory pathways[43, 62]. The periaqueductal grey (PAG) and rostroventromedial medulla (RVM) form part of a descending control system that can either inhibit or facilitate nociceptive transmission at the spinal level[54, 63]. Functional neuroimaging studies suggest that attentional engagement can influence these systems, contributing to reduced pain perception through enhanced descending inhibition[55, 57]. This provides a mechanistic link between cognitive processes and physiological modulation of pain.

Beyond attention, broader cognitive constructs such as expectation, prediction, and cognitive control also influence pain processing. Placebo and nocebo effects illustrate how expectations modulate pain through top-down mechanisms, altering both subjective experience and neural activity[64]. The same inferential framework recasts attention as a mechanism for prioritising particular sensory inputs or predictions, thereby influencing the pain that results[45].

Attention and cognition are therefore key determinants of how pain is experienced and modulated over time. Pain interacts with competing cognitive demands, is shaped by expectations and predictions, and is influenced by descending neural systems. For interventions that rely on attention, engagement, or cognitive task performance, therapeutic benefit is therefore likely to depend on matching task demands to the user's current capacity and state.

2.2 Immersive therapeutics for pain: evidence and constraints

2.2.1 Why immersive technologies?

Virtual reality (VR) is a novel technology that uses a head-mounted display (HMD) to provide an immersive, 360° audiovisual experience that may be passive or interactive depending on application design. Over the last decade, VR technology has matured rapidly through advances in computing and substantial industry investment. Modern standalone systems combine high-fidelity audiovisual stimulation with real-time head and body tracking (Figure 2.2). This enables movement within and interaction with virtual environments, allowing users to

perceive and navigate virtual spaces in ways that can approximate, augment, or deliberately reshape real-world perception and action[65, 66]. Increasingly, VR systems also incorporate embedded sensors, including eye tracking, allowing behavioural and physiological signals to be captured during the same experience used for therapeutic delivery[67]. Consumer pricing and availability have grown alongside this technical maturity, supporting interest in VR as a platform for home-based pain therapeutics[6].

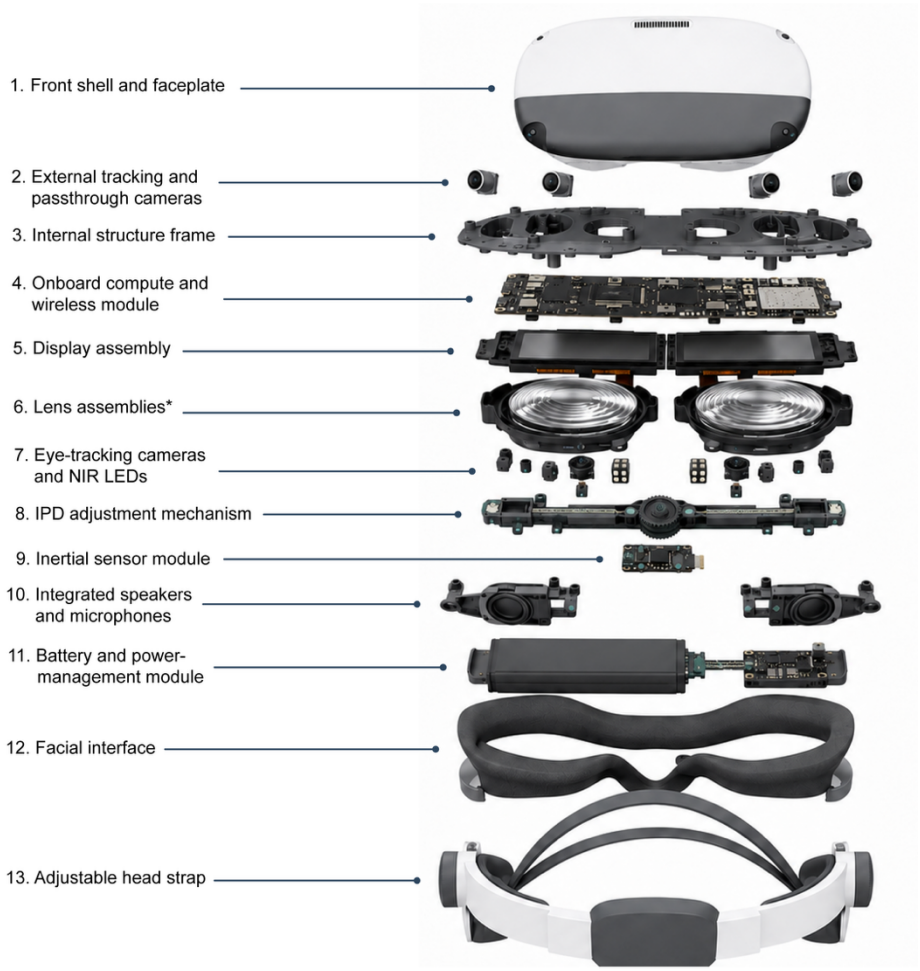


Figure 2.2. Schematic architecture of an untethered VR headset

The exploded schematic illustrates major component groups commonly integrated within contemporary VR headsets, not including interaction peripherals such as wireless controllers. *The depicted optical assemblies resemble Fresnel lenses; other headsets may use alternative optical designs, including pancake lenses. Component placement is illustrative and does not represent a specific manufacturer. NIR, near-infrared; IPD, interpupillary distance.

The relevance of immersive technologies to pain lies less in hardware specifications than in their capacity to deliver tightly controlled, multisensory, and highly immersive experiences. Unlike conventional screen-based media, immersive VR encapsulates the user within a contained, three-dimensional audiovisual environment, reducing the salience of competing sensory inputs and supporting sustained attentional engagement[6]. This aligns with established models of attentional modulation of pain, in which allocation of cognitive resources influences the perception and processing of nociceptive signals[68]. Immersive environments can also be designed to manipulate task demands, sensory context, and behavioural contingencies, enabling systematic engagement of cognitive, affective, and behavioural processes relevant to pain.

A further distinguishing feature of immersive technologies is their capacity to integrate multiple therapeutic components within a single delivery platform. VR interventions can incorporate attentional engagement, graded movement or exposure, emotional regulation strategies and educational content, supporting multidimensional targeting of pain-related processes[6, 68-70]. This is relevant in chronic musculoskeletal pain, where symptom experience reflects the interaction of multiple factors rather than a single underlying mechanism. Delivery of multimodal interventions within a controlled and repeatable environment offers advantages over traditional approaches, which are often fragmented across different therapeutic modalities.

Immersive systems also provide methodological advantages for intervention development and mechanistic investigation. Virtual environments can be specified, reproduced, and modified with precision, allowing systematic variation of stimulus characteristics and task parameters[71]. This creates opportunities for therapeutic delivery, and for investigating the mechanisms underlying pain modulation and intervention response. The interactive and engaging nature of VR may also support adherence and user engagement compared with more passive or didactic interventions.

The same features that make immersive technologies powerful also introduce complexity. The effects of VR are likely to depend on the interaction between user characteristics, task demands, and environmental design. Immersive interventions should not therefore be considered as uniform or inherently effective, but as

configurable systems whose impact depends on how they engage underlying cognitive and affective processes. This distinction provides the basis for evaluating the current evidence and constraints of immersive therapeutics.

2.2.2 Evidence for VR in pain: acute and chronic

The evidence base for immersive VR in pain spans both acute and chronic pain contexts, but the maturity and interpretability of that evidence differ substantially between them. In acute and procedural pain, the literature is relatively mature and provides strong proof-of-principle that immersive VR can reduce pain-related outcomes. In chronic pain, the literature is more heterogeneous, with greater variation in patient populations, intervention content, delivery models, comparators, and outcomes. Acute pain studies primarily establish that immersive engagement can modulate the acute pain experience, whereas chronic pain studies must address whether VR can support clinically meaningful and durable improvements in more complex, multidimensional conditions.

Early experimental and clinical studies provided the initial mechanistic and therapeutic rationale for VR analgesia. Hoffman and colleagues demonstrated that immersive VR reduced both subjective pain ratings and pain-related brain activity during experimentally induced thermal pain, providing early neuroimaging evidence that immersive attentional engagement can modulate brain responses associated with acute pain[72]. Subsequent electrophysiological work showed that active VR reduced early and late pain-related evoked potentials during painful electrical stimulation, supporting the view that immersive engagement can influence both pre-perceptual and later stages of pain processing[73]. These studies suggested that VR does not merely provide a novelty effect, but is a technology capable of modulating pain-related neural responses under controlled experimental conditions.

Clinical trials in acute and procedural pain have generally reinforced this conclusion. One influential randomised comparative effectiveness trial found that VR produced greater immediate pain reduction than a health and wellness television control in hospitalised patients with pain, with benefits also persisting up to 72 hours in some analyses[74]. Recent systematic reviews and meta-analyses conclude that immersive VR is effective for acute and procedural pain management across a range of settings, although heterogeneity in study design, comparator conditions, and

intervention format remains substantial[75, 76]. The acute pain literature therefore provides reasonably consistent proof-of-principle that immersive VR can reduce pain in time-limited painful states, particularly where attentional capture, emotional modulation and perceptual immersion are plausible mechanisms.

Translation into chronic pain has been more complex. Chronic pain is not simply repeated acute pain, and its persistence is shaped by interacting sensory, cognitive, affective, behavioural, and social processes. Attentional modulation may remain a relevant therapeutic mechanism, particularly for short-term symptom relief and engagement, but chronic pain interventions must also address whether benefits can be sustained, repeated, or integrated with broader behavioural and self-management strategies.

Early chronic pain studies suggested that immersive VR could meaningfully reduce pain in some patients. A study in a mixed chronic pain cohort reported substantial short-term pain reduction during an interactive VR session, with some patients also reporting continued benefit after headset removal[77]. Subsequent work evaluating a self-administered, skills-based VR intervention incorporating behavioural pain management content showed improvements in pain related outcomes compared with an audio-only comparator, indicating that VR could act as a platform for delivering structured therapeutic content[69].

The strongest chronic musculoskeletal evidence currently comes from chronic low back pain (cLBP). An 8-week self-administered, at-home behavioural skills-based VR programme delivered through the EaseVRx system demonstrated improvements in pain intensity and interference relative to sham VR in a double-blind randomised placebo-controlled trial[78]. Follow-up durability analyses from this programme and related effectiveness trials have suggested that benefits may be sustained at 12 months and 24 months in some cohorts, with sub-group analyses indicating larger effects in those with high-impact pain[79-81]. These studies move the field beyond single-session paradigms and suggest that home-based VR can act as a scalable delivery platform for behavioural pain management strategies with longer-term clinical relevance.

Other cLBP studies support the view that VR's therapeutic potential is not restricted to one mechanism or intervention model. Improvements in disability, pain

interference and fear of movement have been demonstrated with a 6-8 week embodied psychological and behavioural VR intervention designed for adults with high disability[68]. Broader chronic pain syntheses similarly show that VR interventions have been used to target pain, functioning, mobility, psychological outcomes, quality of life, and body-related perceptions through a wide variety of approaches, including skills-based programmes, embodiment, exergaming, graded exposure, meditation, and pain neuroscience education[82, 83].

Evidence in wider musculoskeletal pain populations is expanding but variable in depth and quality. A narrative review of therapeutic VR in rheumatology identified 56 studies across cLBP, chronic neck pain, complex regional pain syndrome, fibromyalgia syndrome, rheumatoid arthritis, osteoarthritis, and other rheumatology-relevant conditions, with marked heterogeneity in study design, intervention type, comparator selection and outcome assessment[5]. The review showed that cLBP and chronic neck pain account for most of the current literature, whereas evidence in fibromyalgia and other rheumatological conditions remains relatively sparse. Across this broader literature, positive signals are evident across pain, function and mood-related domains, but the intervention content used to produce those signals varies considerably[5, 82]. In fibromyalgia, available studies suggest feasibility and possible benefits for pain and psychological outcomes, while highlighting variability across systems, virtual environments and activities, as well as the need for larger and longer-term trials[5].

The most influential trials also warrant critical appraisal. The pivotal EaseVRx trial compared active VR with a sham condition of two-dimensional nature footage delivered through the same headset[78]. This controlled for hardware and some expectancy but not for the degree of immersion, interactivity or perceived credibility, and participants could plausibly infer their allocation. Outcomes were self-reported and the 12 and 24-month durability analyses were conducted by the same group in samples reduced by attrition[79-81]. Similarly, the inpatient comparative effectiveness trial used an unblinded television comparator[74], and much of the acute pain evidence measures pain during or immediately after a single session. The absence of superiority over sham VR in a recent three-arm trial[84] suggests that headset use, expectancy and non-specific attention account for a meaningful share of the improvement reported in less rigorously controlled studies, and meta-analyses pool

small trials with heterogeneous comparators and frequent unclear or high risk of bias[82, 85]. These issues do not negate the evidence, but they indicate that effect sizes from early trials are likely to overestimate the specific contribution of VR content.

The evidence therefore supports a balanced position. Immersive VR is not merely a technological novelty, and there is credible evidence that it can reduce pain and improve function across a range of contexts. However, in chronic musculoskeletal pain, VR is better understood as a configurable therapeutic medium than as a uniform intervention, with response shaped by content, mechanism, delivery and population. The relevant question is not whether VR 'works', but under what conditions, for whom and through which therapeutic configurations.

2.2.3 Mechanisms of immersive analgesia

No single mechanism fully explains why immersive VR may reduce pain or improve pain-related outcomes. Therapeutic VR is best understood as a multicomponent intervention platform whose effects depend on the interaction between delivered content, user characteristics and clinical context. In acute pain, benefit may arise predominantly through attentional modulation and reduced salience of nociceptive input. In chronic pain, VR may additionally confer benefit by influencing emotional regulation, body representation, fear of movement, behavioural activation and self-management. Recent conceptual reviews therefore frame VR as a medium able to address different processes relevant to chronic pain through immersive biopsychosocial treatments[70, 83, 86].

VR may intensify attentional analgesia by combining presence, multisensory competition and task engagement within a controlled immersive environment. By encapsulating the user within a dynamic audiovisual environment that requires active cognitive engagement, VR competes powerfully with pain-related attention. This does not imply that pain processing is simply switched off. Rather, immersive engagement appears to reduce the subjective salience of pain and alter the way nociceptive information is processed. Experimental work has shown that VR can reduce subjective pain ratings while modulating pain-related neural activity, including activity in regions implicated in sensory and affective pain processing and changes in pain-related evoked potentials[72, 73, 86].

Immersion alone is unlikely to be sufficient to promote analgesia. A growing body of work suggests that analgesic benefit depends on how effectively the virtual experience captures and sustains engagement. Experimental and clinical studies have shown that ‘active VR’ can outperform ‘passive VR’ under certain conditions. This effect has been observed in cold pressor paradigms, painful electrical stimulation and burn care, with greater reductions in pain ratings, pain-related neural responses, or increases in pain tolerance observed when users actively interact with the VR environment[73, 87-90]. However, more interactivity is not uniformly better. Meta-analytic evidence supports the view that the contribution of specific design features, including interaction level, is conditional rather than absolute, so interactivity is better understood as a context-dependent effect modifier than as a uniform driver of efficacy[85].

A further mechanism concerns emotional regulation and the affective context of pain. Chronic pain is shaped not only by nociceptive input, but also by anxiety, distress, catastrophising, low mood, and reduced perceived control. VR can deliver relaxation exercises, guided meditation, breathing or biofeedback-based tasks, nature-based experiences, and cognitive behavioural strategies within an immersive format[5]. The affective content of VR is not merely incidental to analgesia and may be one of its active ingredients. In a within-subject study of chronic temporomandibular disorder, affectively rewarding VR content was associated with greater pain reduction, with improvements in mood and enjoyment mediating analgesic effects for a nature-based ocean condition, whereas autonomic arousal did not explain these effects[91]. Some benefits of VR therefore arise not only from attentional modulation, but from altering the emotional meaning and experiential quality of pain through affectively positive and engaging content.

Embodiment and the modulation of body representation provide another important mechanistic pathway. Foundational work in virtual embodiment has shown that users can experience ownership over a virtual body viewed from a first-person perspective, with physiological responses to threat directed at the virtual body supporting the view that virtual avatars can be incorporated into self-representation[92, 93]. The sense of embodiment has been conceptualised as comprising body ownership, agency, and self-location, providing a framework for understanding how immersive systems may alter the relationship between perceived body state and bodily experience[94]. This is

relevant to chronic pain because some conditions, particularly complex regional pain syndrome, are associated with altered body perception, disturbed body schema and pain linked to how the body is represented[95]. VR allows manipulation of first-person perspective, body ownership, visual appearance, movement congruence and multisensory feedback. Studies in chronic arm pain and complex regional pain syndrome suggest that changing the appearance, size, transparency, or synchrony of a virtual limb or body can reduce pain and alter body perception disturbance in some patients[96, 97]. The relevance of this mechanism varies across conditions, but it broadens the conceptual basis for VR therapeutics in chronic pain.

VR may also confer benefit through graded exposure, movement retraining, and sensorimotor retraining. In chronic musculoskeletal pain, feared or painful movement can become associated with threat, avoidance, deconditioning, and reduced confidence, as described in fear-avoidance models[39, 98]. Immersive environments allow movement to be rehearsed in a controllable and often more engaging context than standard exercise alone, with the ability to manipulate task difficulty, visual feedback, and environmental context[99-101]. VR-based interventions have demonstrated improvements in movement and functional outcomes in conditions such as cLBP and chronic neck pain, supporting the potential for immersive systems to recalibrate threat-related expectations and support more adaptive movement behaviour[100, 102, 103].

A related mechanism is repeated learning and behavioural change. Several promising chronic pain VR interventions incorporate structured, skills-based programmes delivered repeatedly over days or weeks rather than relying on a single-session[5, 69, 70]. These programmes typically include elements of cognitive behavioural therapy, relaxation training and pain education, aiming to build coping skills, reinforce behavioural strategies, and increase self-efficacy. Randomised trials in cLBP demonstrate improvements in pain and pain-related interference following such interventions, consistent with the delivery of behavioural pain management content[68, 79, 104]. Therapeutic benefit may therefore arise from cumulative learning, rehearsal and behaviour change supported by immersive delivery, with adherence, repetition, and perceived relevance likely to determine whether engagement translates into sustained clinical benefit when interventions are home-based.

These candidate mechanisms are not mutually exclusive. Most immersive interventions likely derive their effects from several overlapping processes operating simultaneously, with their relative importance varying according to the virtual environment, task structure, treatment goal and the individual user. For some interventions, attentional modulation may prove the principal mechanism; for others, embodiment, emotional regulation, sensorimotor training, or behavioural skills acquisition may take precedence. The capacity of VR to deliver multiple mechanisms is one of its strengths, but it also creates a challenge for interpretation and design. If different people benefit through different pathways, and if the same person may respond differently depending on fatigue, mood, pain severity, cognitive load, or environmental features, then a fixed immersive intervention is unlikely to optimise benefit reliably.

2.2.4 Limitations, variability and the case for adaptation

Response to VR is not determined by the technology alone. It depends on how the intervention is configured and delivered, and how the user's current clinical and experiential state interacts with the demands of the virtual environment. The mechanisms underpinning VR interventions are unlikely to be recruited consistently across heterogeneous chronic musculoskeletal pain populations without careful attention to intervention design and user state.

Systematic review evidence reflects this complexity. Across chronic pain studies, heterogeneity in study characteristics with inclusion of diverse pain conditions, devices, software, intervention durations, and outcome domains limits interpretation of VR intervention efficacy[82]. Variability is not only a limitation of the evidence base itself but is a defining feature of the therapeutic design problem. Meta-analytic evidence further indicates that response is partly modified by demographic and clinical characteristics. In a systematic review and meta-regression including 122 randomised controlled trials and 9138 participants, immersive VR reduced pain overall, but larger effects were associated with younger age, moderate-to-severe baseline pain and comparator type. Software type and interaction level were not consistently retained as independent predictors in multivariable models[85]. This finding should not be interpreted to mean that design is clinically unimportant. Instead, the impact of design choices is likely to depend on the population, symptom profile, intervention context and comparator used. A virtual task that benefits one

patient group may not produce the same effect under different conditions, and optimisation cannot be reduced to generic increases in immersion, interactivity or exposure time.

Recent cLBP trial evidence reinforces this. A large, randomised trial comparing skills-based VR, 360-degree video attention-based VR and sham VR found modest improvements across all three arms but no superiority of either active VR condition over sham VR for the primary pain interference outcome. The trial reported high variance in dose, and exploratory analyses suggested that higher baseline anxiety predicted greater response to skills-based VR but not to attention-based or sham VR[84]. This result requires cautious interpretation given its exploratory nature. Nevertheless, it illustrates a clinically plausible principle in which effectiveness of a VR programme depends on alignment between therapeutic content and the participant's dominant symptom profile. A skills-based intervention may be more relevant to a person whose pain experience is strongly linked with anxiety, arousal, or coping demands, whereas other users may require different content, pacing, or task structure.

Experimental cognitive literature supports this interpretation. Attentional modulation is not a fixed property of an external task; it depends on the balance between task demand, motivation, and performance. The analgesic value of a demanding task is contingent on engagement, with experimental evidence indicating that increased cognitive demand attenuates pain only when motivation to perform is sufficient[105]. Inter-individual work also indicates that psychological traits shape this trade-off, with pain catastrophising, trait anxiety, and trait mindfulness associated with how pain and cognitive performance interfere with one another during a competing task[106]. These findings are directly relevant to VR because immersive interventions impose dynamic combinations of sensory, cognitive, affective, and motor demand. A task may be absorbing when demand is appropriately calibrated, ineffective when it is too easy, and aversive or fatiguing when it exceeds the user's available capacity.

This issue is salient in chronic musculoskeletal pain, where symptoms fluctuate across and within individuals[9-11]. Pain intensity, fatigue, sleep disturbance, cognitive difficulty, sensory sensitivity, mood, fear of movement, and prior confidence with technology may all influence how a session is experienced. Feasibility work in fibromyalgia comparing different VR systems, activities, and environments found

reductions in pain intensity and pain-related questionnaire scores across multiple conditions, but also substantial variation in the magnitude of response. Differences were observed between virtual environments, with a warm, sunny environment associated with greater pain reduction than the cold, snowy environment, alongside more favourable subjective experience including higher engagement, enjoyment, relaxation, and mood improvements. Responses also varied considerably between participants, with some reporting substantial pain reduction and others reporting little change in symptoms after VR exposure[8]. Despite this study's exploratory design, the findings illustrate that immersive analgesia depends not only on the presence of VR, but on the sensory, affective, and experiential qualities of the virtual context, and on how these interact with individual preference and symptom state.

Intervention type also influences how variability should be interpreted. Therapeutic VR interventions differ substantially in intended mechanism, content, dose, user demands and expected timescale of benefit. A single-session attentional task, an embodied movement intervention, a relaxation environment and a structured behavioural skills programme should not be interpreted as interchangeable forms of the same treatment. Combining such interventions under a single label of "VR" risks obscuring whether non-response reflects limited therapeutic efficacy, insufficient dose, inadequate comparator selection, poor implementation or a mismatch between user and intended therapeutic process. VR trials therefore require detailed reporting of intervention rationale, content, delivery, user interaction, safety, context and implementation. The recent INVIRTUE guidance provides consensus-based reporting items for therapeutic immersive VR interventions, complementing earlier methodological and early-phase evaluation guidance[107]. For chronic pain trials, pain intensity alone is therefore insufficient. Interpretation also requires evidence that the intended therapeutic exposure was delivered, tolerated and engaged with in a manner consistent with the proposed mechanism.

Beyond these design considerations, the existing evidence base is constrained by several limitations. Studies are marked by small samples, short intervention periods, limited long-term follow-up, and substantial heterogeneity in hardware, software, therapeutic content, treatment dose, comparators, and outcome measures[82, 108]. These limitations make it difficult to determine which intervention components account for observed effects. Comparator design is challenging because sham or

low-intensity VR conditions may still provide a degree of immersion, attentional modulation, and positive expectancy. Methodological guidance for VR trials therefore emphasises staged development, human-centred design, feasibility testing, tolerability assessment, appropriate control selection, and adverse event monitoring before definitive efficacy trials are attempted[71, 109].

A further constraint relates to the static nature of VR interventions. Prespecified modules are delivered within fixed environments with predetermined session lengths. This structure suits early clinical evaluation but is poorly aligned with chronic musculoskeletal pain, where symptoms and cognitive capacity may fluctuate within and between sessions. The FDA-authorised EaseVRx system illustrates that structured home-based VR can meet a regulatory standard for adjunctive cLBP treatment, but it also exemplifies a delivery model of preloaded behavioural content administered through a sequence of sessions rather than continuous adaptation to user state[110]. More broadly, current therapeutic VR interventions are under-instrumented from a state-monitoring perspective. They can deliver highly controlled content, but usually provide limited information about whether the user is attending to the task, tolerating the exposure, becoming fatigued or disengaging.

Adverse effects and practical constraints also limit repeated use. Cybersickness is one of the most commonly reported adverse effects of head-mounted VR and can include nausea, oculomotor symptoms such as eye strain and headache, and disorientation[111, 112]. A leading explanation is sensory conflict, in which visual signals of self-motion are incongruent with vestibular and proprioceptive information, while postural instability theory proposes that symptoms arise when users have difficulty maintaining stable postural control in an unfamiliar sensory environment[113]. Symptom burden is influenced by factors including virtual content, locomotion, exposure duration, display latency and individual susceptibility, and transient effects on balance or visuomotor performance may persist after headset removal[114]. Head-mounted displays can also produce visual discomfort through vergence-accommodation conflict, whereby the eyes converge on virtual objects at varying simulated depths while accommodating to a fixed focal plane[115]. Their physical mass may alter head and neck posture[116], which is particularly relevant where musculoskeletal symptoms are already present. Transient dissociative experiences have also been reported following immersive VR exposure[117].

These considerations may be particularly important in chronic musculoskeletal pain, where fatigue, pain and sensory sensitivity can already constrain sustained engagement. In the fibromyalgia feasibility study described in this thesis, fatigue was the adverse effect most likely to persist after headset removal, was reported more frequently with longer exposure, and adverse-effect burden varied between VR systems and environments[8]. In chronic primary pain, motion-related symptoms, headset burden and technical difficulty have similarly been identified as barriers to usability, engagement and retention[118]. Adverse effects are also inconsistently reported across VR trials, supporting recent recommendations that safety and tolerability are explicitly assessed and reported[107, 109]. Therapeutic VR should therefore be evaluated not only according to whether its content is theoretically effective, but also whether users can comfortably complete, engage with and repeatedly return to the intervention. This becomes increasingly important for home-based delivery, where sustained therapeutic exposure depends on the technology fitting reliably into everyday life.

Implementation evidence indicates that adoption of VR depends not simply on device availability, but on whether services can identify a clear therapeutic role, support clinicians and users, and integrate the technology into existing care pathways[119]. In chronic pain, this challenge is amplified because treatment is usually embedded within long-term self-management rather than delivered as a single supervised encounter. The current evidence base provides limited reassurance about equitable deployment. Usability studies have not consistently included older adults, people with lower educational attainment or racially and ethnically diverse populations, limiting confidence that current systems have been designed around the full range of people likely to use them[120]. For home-based VR, equity concerns extend to digital inclusion. NHS England identifies digital exclusion as a driver of health inequalities, highlighting barriers including connectivity, affordability, device access and digital skills[121]. Ofcom data show that home internet access remains socially patterned in the UK, with lack of access more common among older adults and lower socioeconomic households[122]. Home-based immersive therapeutics could therefore widen inequalities if connectivity, device access and technical support are not considered from the outset.

The cumulative implication is that fixed group-level content is unlikely to optimise treatment in a condition characterised by heterogeneity and fluctuation. VR is configurable by nature, but configurability alone does not ensure therapeutic precision. Its value depends on whether the sensory, cognitive, affective and behavioural demands of the virtual environment can be matched to the target mechanism and to the user's baseline characteristics and current state. Addressing this limitation requires a shift from pre-specified immersive content toward systems that can measure relevant aspects of user state during use and adapt therapeutic delivery accordingly.

2.3 Closed-loop personalisation: rationale, biomarkers, sensors and design requirements

2.3.1 Closed-loop therapeutics: framework and rationale

The constraints of current VR interventions point toward a transition from VR as a pre-specified delivery medium to VR as an adaptive therapeutic system. The therapeutic processes outlined earlier, including attentional engagement, affective regulation, movement behaviour and learning, are unlikely to remain stable throughout use. They may shift as the user becomes fatigued, engaged, overloaded, reassured, distressed or more confident. Therapeutic optimisation therefore requires more than selecting the right module at baseline. It requires a system capable of responding to the user's changing state during delivery.

A closed-loop therapeutic system can be understood as a repeated cycle in which user-state information is measured, interpreted through decision rules or an adaptive algorithm, used to modify the intervention, and then monitored again to determine whether the proximal therapeutic state has changed (Figure 2.3)[123]. In the context of VR, the system first captures information about the user during immersion, uses that information to estimate a therapeutically relevant state, modifies some aspect of the virtual environment or task, and then observes whether the modification produces the intended effect. This logic aligns with the broader concept of just-in-time adaptive interventions, which aim to deliver the right type or amount of support at the right moment by adapting to a person’s changing internal state or context[124]. It also aligns with adaptive VR frameworks outside pain, where effective adaptation depends on measurable user information, an adaptive decision logic, and modifiable features of

the virtual environment[124]. For pain therapeutics, this framework shifts VR from a standardised content-delivery tool toward a therapeutic system that can adjust to the user's state.

Adaptation requires information. A system cannot adapt meaningfully on the basis of diagnosis, baseline pain score or prescribed module alone. It requires some estimate of the user's current state, including whether they are engaged or disengaged, challenged or overloaded, relaxed or distressed, tolerating movement or avoiding it, maintaining performance or deteriorating. These are not purely biological states, and they should not be treated as direct proxies for pain mechanisms. They are functional and experiential states relevant to whether a therapeutic exposure is likely to be useful at that moment.

The clinical purpose of adaptation must be specified carefully. A closed-loop VR system should not aim to reduce pain intensity at every moment, because short-term pain reduction may not always align with longer-term therapeutic goals such as movement confidence, behavioural learning, reduced avoidance, or improved self-management. A movement-based intervention may need to preserve tolerable challenge rather than remove all discomfort, while an attention-based intervention may need to maintain attention without producing cognitive overload. The adaptive target is therefore better conceptualised as a proximal therapeutic state. This could include maintaining engagement, preventing cognitive overload, supporting relaxation, encouraging safe movement, optimising learning, or keeping task difficulty within a range that is challenging but tolerable. These states are clinically meaningful because they influence whether the intended therapeutic process is being delivered.

This framing clarifies why closed-loop personalisation requires more than baseline stratification. Baseline measures may help identify broad phenotypes or likely treatment preferences, but they cannot determine whether the intervention is appropriately calibrated at a particular moment. A patient may begin a session motivated and attentive, but become fatigued, anxious, distracted, or overstimulated as the task unfolds. Conversely, a task that is initially too easy may fail to recruit sufficient attention or engagement. This temporal problem is especially relevant for VR, because the system can vary task demand, sensory complexity, movement requirements, feedback, and environmental context with fine control. Static VR uses

this configurability only at the design stage, whereas closed-loop VR would use it therapeutically during delivery. Adaptive VR therefore shifts the design question from "which VR intervention works?" to "which therapeutic state should be supported now, and how should the environment change to support it?"

Evaluation of an adaptive VR system also differs from the evaluation of static interventions. Conventional trials primarily ask whether a pre-specified intervention produces an average treatment effect over a comparator. Adaptive systems raise additional questions about when intervention components should be delivered, for whom they are beneficial, and which time-varying states moderate their effects. Adaptive VR should therefore be developed through staged, theory-informed evaluation rather than treated as an immediately superior form of treatment. VR pain interventions are complex interventions, combining technology, therapeutic content, user interaction, sensory design, and contextual delivery.

The updated Medical Research Council (MRC) framework for complex interventions emphasises programme theory, feasibility, context, implementation and refinement across the development and evaluation pathway[125]. Closed-loop VR extends this complexity because the intervention is not only multicomponent but dynamically contingent on measured user state, and the adaptation logic itself becomes part of the intervention.

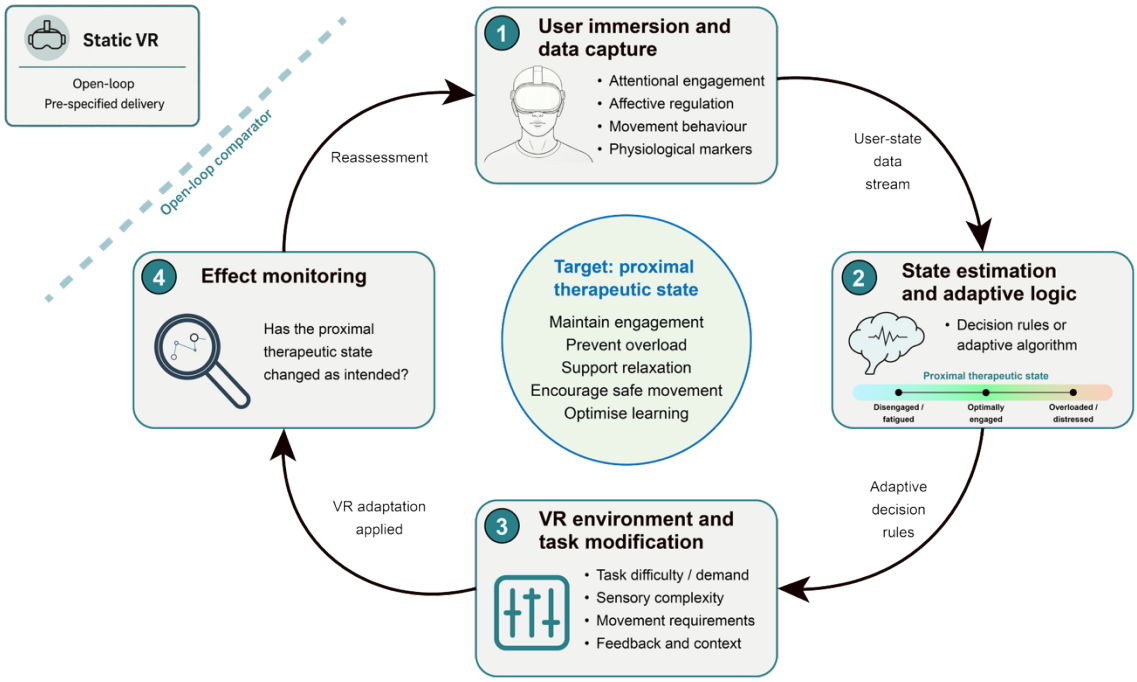


Figure 2.3. Closed-loop framework for adaptive VR therapeutics

The schematic illustrates a repeated cycle of user-state data capture, state estimation, adaptive modification and reassessment during VR delivery. The adaptive target is conceptualised as a proximal therapeutic state, such as maintaining engagement, preventing overload, supporting relaxation or encouraging safe movement, rather than immediate symptom reduction alone.

2.3.2 Mechanistic and state-sensitive biomarker roles

The term biomarker requires careful use in the context of closed-loop VR therapeutics. The MRC framework for stratified medicine defines a biomarker or marker as any measurable trait that could potentially be associated with, and used to define, a patient stratum[126]. This broad definition extends beyond biochemical assays to include imaging, neurophysiological, clinical, psychological, and behavioural measures. The same framework distinguishes mechanistic markers from markers that predict treatment response, including the concept of a “mechagnostic” marker that identifies a distinct mechanistic stratum without necessarily predicting immediate therapeutic benefit[126].

Two translational roles for biomarkers and sensing modalities are relevant to the design of adaptive immersive therapeutics. The first concerns mechanistic biomarkers, used to characterise underlying biology, identify candidate pathophysiological processes, define mechanistically relevant heterogeneity, or test

whether an intervention engages a plausible physiological target. Their value depends on biological interpretability and relevance to the clinical problem, rather than immediate suitability for real-time control. In chronic pain, this role is important because symptom severity and treatment response are not adequately explained by structural pathology alone. Mechanistic biomarkers may help identify altered central processes that could inform the design of future therapeutics. EEG sits primarily within this category. It provides temporally precise information about spontaneous neural activity and has been used to investigate oscillatory and network-level differences in chronic pain[44, 127]. Its relevance is not that it provides an embedded measure of cognitive state during VR, but that it may help characterise neurophysiological differences associated with chronic musculoskeletal pain and identify candidate domains that immersive therapeutics might seek to modulate. Systematic reviews suggest that chronic pain is associated with altered rsEEG and magnetoencephalography (MEG) features, although findings remain heterogeneous and relationships with pain intensity, prognosis, and treatment response are inconsistent[44, 127]. Recent large-scale work supports the potential of EEG-based chronic pain biomarkers while emphasising persistent challenges around replicability, generalisability and interpretation across datasets[128].

The second role concerns embedded state-sensitive sensing. These are measurements used to estimate functional state during the intervention and to provide the information substrate for adaptation. Their value depends less on providing a complete physiological explanation than on being sufficiently reliable, temporally resolved, low burden, and compatible with immersive delivery. Eye tracking sits primarily within this category. Modern VR headsets increasingly incorporate eye tracking, allowing gaze behaviour, fixation patterns, pupil dynamics and blink-related metrics to be captured during the same system that delivers the intervention[67]. Eye tracking is pragmatically attractive for future adaptive VR because it can be embedded within the headset, sampled continuously, and acquired with minimal additional burden. It should not, however, be described as a direct or specific biomarker of pain, cognitive load or therapeutic mechanism, because ocular signals may be influenced by alternate factors such as task demand, visual attention, luminance, arousal, fatigue and emotional state. Its value lies in providing task-

embedded, state-sensitive information that may support inference about functional state during VR use, rather than in offering a standalone physiological explanation.

These two roles are governed by different selection criteria. A signal informative for mechanistic investigation must be physiologically plausible, interpretable in relation to candidate processes, and amenable to controlled acquisition and analysis. A signal useful for in-session adaptation must additionally be temporally resolved on the timescale of intervention decisions, deployable outside laboratory conditions, low burden across repeated sessions, and compatible with the physical and operational constraints of head-mounted VR. A scientifically informative signal is not necessarily a practical control signal, and a practical control signal is not necessarily mechanistically specific. rsEEG is informative for mechanistic characterisation but burdensome to acquire concurrently with VR. Eye tracking is well integrated into VR but less mechanistically specific. The appropriate question is not which sensor is best in general, but which signal is fit for the specific role it is asked to serve.

The utility of a biomarker also depends on the timescale over which it varies and the decision it is intended to inform. For adaptive immersive therapeutics, this can be considered across three nested timescales. At the slowest level, relatively stable between-person characteristics may support stratification, informing selection of an intervention, its content or its initial parameters before treatment begins. Measures used in this way require sufficient test-retest reliability and trait-like stability, and may fulfil susceptibility, diagnostic, prognostic or predictive roles within established biomarker taxonomies[129, 130]. At an intermediate timescale, measures acquired before, after or between sessions may be used to titrate dose, content or progression over days to weeks. This resembles adaptive treatment strategies in which tailoring variables are reassessed at predefined decision points[131, 132]. Measures serving this function are analogous to monitoring or response biomarkers, with sensitivity to clinically meaningful change over the interval of interest being more important than moment-to-moment resolution. At the fastest timescale, continuously sampled signals can potentially inform adaptation within a task, allowing task parameters to be modified over seconds to minutes as a session unfolds. Here, temporal resolution, latency, participant burden and robustness to artefact become particularly important, while mechanistic specificity may be less critical. Such

decision points are conceptually aligned with the logic of just-in-time adaptive interventions[124].

Pain, cognitive load and fatigue map onto these timescales differently. For pain, relatively stable neurophysiological or phenotypic characteristics, such as nociplastic symptom burden, are potential candidates for stratification, while variation in pain intensity and interference over days to weeks may support adaptation between visits[10, 11]. Momentary pain is more difficult to incorporate into within-task control because repeated self-report is intrusive and physiological correlates of ongoing clinical pain currently lack sufficient specificity and validation for reliable moment-to-moment inference[133]. Cognitive load is more amenable to measurement at the fastest timescale. It changes rapidly as task demands vary and has ocular and behavioural correlates that can be sampled continuously, whereas more stable characteristics such as working memory capacity are better suited to stratification. Fatigue may operate across all three levels, from persistent fatigue as a clinical feature of fibromyalgia[41], through day-to-day fluctuation[36], to time-on-task fatigue within a session, which may be accompanied by task disengagement and progressive attenuation of task-evoked pupil responses[134]. These timescales are therefore nested rather than independent. Stratification markers may establish the initial operating parameters of an adaptive system, between-visit measures may recalibrate those parameters as the clinical picture changes, and within-task signals may provide finer control within the range established at slower timescales. A single signal is unlikely to fulfil all three functions, reinforcing the need to match each measure to the decision it is intended to support.

Machine learning is not used to infer mechanism directly, but instead its role is to test whether headset-derived temporal signals can support subject-independent inference of task-defined state[135]. Its translational value lies in determining whether embedded signals are sufficiently informative and generalisable for state estimation during VR use, rather than in establishing those signals as pure measures of latent constructs such as cognitive load. This distinction preserves the separate roles of mechanistic and state-sensitive sensing. rsEEG may indicate which neurophysiological domains are altered in chronic musculoskeletal pain and may be modifiable with immersive exposure. Eye tracking may help estimate whether the user is in a cognitive state compatible with a particular therapeutic task. Machine

learning then provides the analytical framework for testing whether embedded signals can support generalisable state inference across individuals.

2.3.3 Resting-state EEG as a mechanistic measure

rsEEG captures spontaneous neural activity in the absence of an explicit task or externally evoked painful stimulus. This is relevant in chronic musculoskeletal pain because ongoing symptoms are not fully explained by peripheral pathology and alterations in central processing, attentional regulation, affective modulation, and sensorimotor function may all contribute to symptom persistence[41, 44, 127]. In this context, resting-state acquisition therefore allows investigation of the background neurophysiological context in which pain is experienced.

The rationale for rsEEG is strengthened by the wider neuroimaging literature. Functional MRI (fMRI) studies in fibromyalgia have repeatedly implicated altered processing and connectivity within insular, cingulate, prefrontal, default mode, salience, and sensorimotor systems[136-140]. Task-based fMRI work has shown altered responses to both painful pressure and non-painful multisensory stimulation, with augmented activity in insula, operculum, medial prefrontal and posterior cingulate regions, alongside reduced activity in lateral frontal and primary sensory regions[140]. Resting-state fMRI work has linked fibromyalgia pain to altered insula-default mode network connectivity, changes in insular-cingulate coupling, and state-dependent default mode network-insula connectivity[136-139]. UK-based population-level evidence linking Fibromyalgia Index scores to altered functional and structural connectivity within the descending pain modulatory system supports the relevance of central modulatory circuitry in nociplastic symptom burden[42]. Nature-based immersive 360-degree video has been shown to attenuate experimentally induced secondary hyperalgesia, with sustained effects linked to perceived presence and insulo-thalamic effective connectivity[141]. These fMRI findings do not map directly onto EEG alterations, because blood-oxygen-level-dependent connectivity and scalp-recorded oscillations measure different physiological processes. They provide important spatial context, allowing oscillatory and connectivity abnormalities measured at the scalp to be interpreted as candidate temporal-spectral manifestations of altered regulation within distributed sensory, salience, sensorimotor and modulatory systems, rather than as isolated band-power markers.

Existing rsEEG and MEG studies indicate pain-associated alterations in oscillatory activity and network organisation, while also highlighting that no single reproducible biomarker has emerged for clinical application. A systematic review reported recurring cross-sectional differences, particularly involving theta and beta activity, but found that longitudinal associations with symptom change and simple correlations with pain intensity were less stable[44]. Earlier review evidence similarly identified abnormalities across canonical frequency bands while emphasising substantial heterogeneity in study populations, acquisition methods, analytic approaches and interpretation[127]. A recent mega-analysis showed limited replication of single-dataset connectivity associations, but more robust group-level and multivariate patterns involving theta, alpha, and beta frequencies[128]. rsEEG is therefore scientifically valuable for mechanistic characterisation and phenotyping, but current evidence does not support a simple single-feature diagnostic signature of chronic pain.

Chronic pain is unlikely to be represented by a single frequency-band abnormality, and oscillatory and connectivity measures may instead capture distributed processes through which sensory, affective, contextual, and regulatory information are integrated. Broader oscillatory models propose that neuronal rhythms support flexible routing of information across brain systems, with alpha and beta activity contributing to feedback or top-down signalling, and gamma activity more closely associated with feedforward signalling under some conditions[142]. Network-level imbalance models extend this perspective by proposing an altered balance between regions implicated in pain input, including somatosensory cortex and dorsal anterior cingulate cortex, and regions implicated in pain suppression, particularly the pregenual anterior cingulate cortex[143].

Across canonical frequency bands, three findings recur with sufficient frequency to inform candidate mechanistic targets, while none currently provides a sensitive marker for chronic pain. Theta has been linked to cortical slowing, thalamocortical dysrhythmia (TCD) and altered large scale connectivity in fibromyalgia and other chronic pain populations[127, 144-146]. Cortical slowing models, originally developed in neurogenic pain, interpreted increased low-to-mid frequency power, slower dominant rhythms and altered cross-frequency coupling as evidence for disturbed thalamocortical regulation[146]. The framework remains influential but is one mechanistic perspective rather than a complete or universal explanation of chronic

pain. Alpha activity is plausibly relevant because of its associations with intrinsic cortical state, sensory regulation, thalamocortical function and attentional control. Some fibromyalgia studies have reported reduced alpha power, altered alpha connectivity or slower peak alpha frequency, while others have found no significant alpha differences[147]. In chronic musculoskeletal pain more broadly, peak alpha frequency has been linked to pain sensitivity in a model of sustained musculoskeletal pain, suggesting that alpha features may reflect vulnerability, sensory regulation or trait-like differences rather than only established chronic pain state[148]. Alpha also has translational relevance as a candidate modifiable rhythm. Alpha transcranial alternating current stimulation increased somatosensory alpha activity in cLBP, with alpha increases associated with pain relief[149], while pre-sleep 10Hz sensory entrainment increased alpha activity compared with sham stimulation under home-use conditions in CWP and fibromyalgia with sleep disturbance[150].

Beta activity is widely implicated in sensorimotor control, maintenance of the current motor or cognitive set and top-down regulation[151, 152], and broader oscillatory frameworks suggest that alpha and beta activity contribute to feedback signalling and the integration of contextual influences with sensory evidence[142]. Beta abnormalities in chronic pain may therefore relate to altered sensorimotor regulation, sustained pain-related predictions, pain vigilance, movement inhibition or reduced flexibility in updating bodily state. Review-level evidence indicates that beta power is often increased in cross-sectional chronic pain studies[44], and fibromyalgia studies have reported high-beta connectivity abnormalities in prefrontal-cingulate, insular and sensorimotor networks[153, 154]. However, beta should not be interpreted as simply increased in chronic pain. Findings vary according to pain phenotype, symptom burden, disease duration, recording condition and outcome metric[147, 155]. A complementary interpretation links beta activity to inhibitory physiology, with beta and gamma oscillations proposed as indirect indices of GABAergic signalling, and beta typically more practical to measure with scalp EEG than gamma[156]. Reduced low beta power has been reported in chronic peripheral neuropathic pain compared with controls, with preliminary associations between lower beta activity and greater pain intensity before spectral normalisation[156]. Beta reductions cannot be treated as direct GABA measures, nor do they have the same meaning across neuropathic pain, fibromyalgia and CWP. The most defensible conclusion is that beta band

abnormalities are recurrent and mechanistically meaningful, particularly in salience, control and sensorimotor networks, but their direction and interpretation are context-dependent.

A major reason for inconsistency across the literature is methodological heterogeneity. Chronic pain studies differ in phenotype, sample size, pain duration, medication exposure, sleep disturbance, fatigue, affective symptoms, recording condition, reference montage, channel density, artefact handling, source localisation, connectivity metric, and statistical correction[44, 127, 147]. Fibromyalgia studies have used scalp-level power, source-localised current density, coherence, phase-based connectivity, graph or topology measures, microstates, and MEG-derived network analyses[147]. Medication handling is also variable, with some studies withdrawing centrally acting medication before recording, some maintaining usual medication, and others reporting insufficient detail. These factors limit strong biomarker claims and underscore the importance of standardised acquisition and analysis when rsEEG is used as a research measure.

A further methodological consideration is the need to separate oscillatory from aperiodic spectral components. Conventional band power analyses may conflate true periodic oscillations with the 1/f-like aperiodic background of the EEG power spectrum[157]. This aperiodic background is the broadband activity on which oscillatory peaks are superimposed. Its power falls as frequency rises, approximating a $1/f^x$ function, which produces the characteristic pattern of greater power at lower frequencies independent of any true oscillation. Because the slope and offset of this component vary across individuals and recording states and have been linked to aspects of underlying cortical activity, power summed within a canonical band reflects both a genuine oscillatory peak and the aperiodic activity beneath it (Figure 2.4). This is relevant to chronic pain because earlier studies interpreted theta, alpha, or beta power within canonical bands without decomposing periodic peaks from the aperiodic component. Recent work investigating the aperiodic component in chronic pain suggests that aperiodic measures may not show robust group differences in all chronic pain datasets[158]. This does not invalidate oscillatory findings, but it perhaps changes how they should be interpreted. Band power alterations should be treated as candidate periodic effects requiring careful decomposition and sensitivity analysis, rather than as simple raw band power differences.

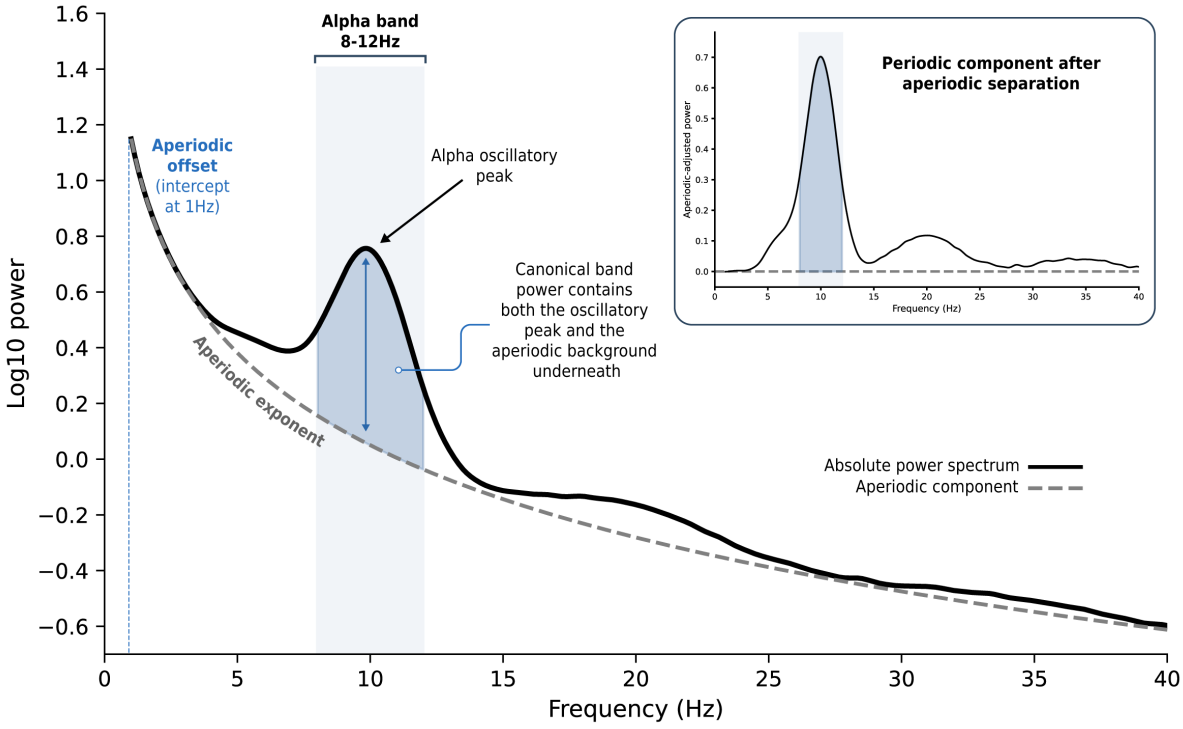


Figure 2.4. Aperiodic separation in EEG spectral analysis

Illustrative EEG power spectrum showing the absolute power spectrum (solid black line) and the fitted aperiodic component (grey dashed line). The shaded region highlights the alpha band (8-12Hz) as a worked example, demonstrating that conventional band power reflects both the oscillatory alpha peak and the underlying aperiodic background. The inset shows the periodic component after subtraction of the aperiodic fit, isolating the residual oscillatory activity. This schematic is intended to illustrate the principle of periodic-aperiodic decomposition, using alpha as an example, rather than to depict all canonical frequency bands.

A complementary interpretive challenge is that rsEEG features are not purely state markers. Quantitative EEG traits, including resting power across frequency bands, alpha peak frequency and spectral connectivity, show substantial within-individual stability across recordings and have well-documented heritability in twin and family studies[159-162]. Heritability estimates for resting alpha and beta band power, alpha frequency and connectivity are typically in the moderate-to-high range, and both adult and developmental data indicate that genetic factors contribute meaningfully to the inter-individual EEG architecture observed at rest[159-162]. This has direct consequences for chronic pain biomarker work. EEG differences identified through conventional case-control designs may therefore partly reflect inherited or familial neurophysiological variation, rather than pain-related effects. This is particularly relevant for chronic widespread and nociplastic pain, which themselves show familial

aggregation[163, 164]. Discordant co-twin designs offer one solution. By comparing twins discordant for chronic pain, with monozygotic pairs sharing 100% of segregating genetic variation and dizygotic pairs sharing approximately 50% on average, conventional case-control comparisons can be supplemented with within-pair analyses that hold genetic background and shared environment partly or fully constant[165]. When EEG differences persist within discordant monozygotic pairs, the case for a pain-related contribution is strengthened relative to a trait or familial explanation.

The translational value of rsEEG therefore lies in mechanistic characterisation rather than direct in-session control. rsEEG can identify neurophysiological differences associated with chronic pain, clarify which features are altered in CWP, and test whether immersive therapeutic exposure is accompanied by measurable changes in resting-state neural activity. These functions are directly relevant to adaptive VR development because they indicate what neurophysiological processes might be altered and potentially modifiable. rsEEG is less attractive as a mainstream embedded control signal. It usually requires controlled resting conditions, dedicated setup, artefact management, offline preprocessing and expert interpretation, and concurrent VR and EEG add further challenges related to headset compatibility, movement artefact, user comfort and synchronisation. rsEEG is therefore positioned as a mechanistic tool that informs the biological rationale for future adaptive therapeutics, while deployable in-session sensing is addressed through eye tracking and behavioural data.

2.3.4 Eye tracking as a pragmatic embedded signal

Eye tracking provides a feasible embedded signal stream for adaptive immersive therapeutics because it can be captured within the same HMD used to deliver the intervention. Modern VR eye tracking captures gaze position, fixation behaviour, pupil size, blink-related features and data-quality indices during immersion, allowing measurement to occur during the therapeutic task rather than just before or after it[67, 166]. The signal is temporally resolved, acquired continuously and does not require additional wearable devices beyond an eye tracking-capable headset. Eye tracking is therefore not the most mechanistically specific measure of pain or cognition, but a sensing approach unusually compatible with deployable VR.

Practical deployability matters because chronic pain populations often experience fatigue, cognitive symptoms, sensory sensitivity, anxiety and variable tolerance of technology. Additional sensors increase setup complexity and the likelihood of non-use. Headset-integrated eye tracking allows signal acquisition without additional electrodes, chest straps, wrist sensors or external cameras[67, 166]. Integration alone is not sufficient. VR eye tracking performance varies across devices, field of view, headset fit, calibration quality, gaze angle and user-specific factors, so data quality must be monitored rather than assumed[166, 167]. The translational case for VR eye tracking is also strengthened by its persistence across multiple hardware ecosystems, with eye tracking incorporated into research, enterprise and consumer systems from manufacturers including Pico, Meta, HTC, Varjo and Apple. Eye tracking also supports gaze-based interaction[168] and foveated rendering, in which rendering quality is prioritised around the user's point of fixation to reduce computational demand[169, 170]. These functions create technical incentives for manufacturers to incorporate eye tracking into VR hardware independently of its therapeutic potential, although device-specific validation remains necessary because sampling rate, calibration, field of view, data access, headset fit and signal quality vary between systems[67, 166, 167].

Pupil size is a candidate signal of central interest for adaptive VR because task-evoked pupillary responses have been associated with mental effort, attention and working memory demand. Classic pupillometry showed that pupil diameter increases with memory load, providing early evidence that the pupil can index cognitive processing demand[171]. Mechanistically, task-evoked pupil dilation is interpreted in relation to locus coeruleus-noradrenergic and autonomic modulation of arousal and attentional control, although this same physiology contributes to the signal's lack of specificity[172, 173]. Subsequent reviews and methodological work support the use of pupil dilation as an index of effort in cognitive control tasks, while emphasising the influence of multiple processes and the need for careful experimental control[174-176]. VR-specific evidence supports the feasibility of pupillometry for assessing cognitive load during virtual training, dependent on task design, visual control and analytic approach[177], and has motivated work on online cognitive-state estimation from pupil dynamics that accounts for disturbances such as illumination change and blink activity[178, 179]. Pupil size may therefore provide temporally resolved information about

whether a task is becoming more cognitively demanding, particularly in tasks requiring working memory updating, sustained attention, response selection or performance monitoring.

Blink-related metrics provide a complementary signal stream. Blink rate and timing have been linked to mental load, attentional state, fatigue and visual task demands[180, 181], and blink behaviour has been used alongside pupil size and other physiological variables to infer cognitive load during work-related and simulated tasks[182]. Some studies suggest that blink rate increases with cognitive load under particular paradigms[183]. Within VR, blink data are useful because blinks are detectable within most eye tracking systems and may provide information complementary to pupil size. Interpretation is complex because blink behaviour is influenced by dry eye, ocular surface factors, visual display use, fatigue, arousal and environmental conditions[184-186]. Within a VR headset, blink patterns may be further influenced by heat, headset fit and display characteristics. Blink metrics should therefore be treated as context-sensitive features rather than direct measures of cognitive load.

The strength of eye tracking for adaptive VR therefore lies not in any single ocular variable, but in the combination of temporally resolved ocular and behavioural features acquired during a controlled task. Pupil size, blink rate, blink duration, missing-data patterns, fixation stability, gaze dispersion, response accuracy, reaction time and task events can be interpreted together to estimate task-defined state. Reviews of gaze and pupil metrics in gamified and simulated sensorimotor tasks support their relevance to mental workload while also emphasising heterogeneity across tasks, populations and analysis methods[187]. Broader reviews of physiological workload measurement show that multimodal and context-aware interpretation is more robust than reliance on a single physiological index[188, 189]. This supports the use of eye tracking as a feature stream for state inference rather than as a standalone biomarker.

Several confounds require explicit handling. The most important is luminance. Pupil diameter is strongly affected by retinal illumination, contrast and visual stimulation, and these effects can be larger than task-evoked cognitive changes. This is acute in VR, where the visual environment can change rapidly and where display luminance depends on virtual scene design, headset hardware, rendering settings and viewing

direction. Methodological guidance in pupillometry emphasises the need to control, model or record visual stimulation when interpreting pupil size[174, 175], and work on the pupillary light reflex and light-adapted pupil size shows that luminance-related effects are fundamental physiological determinants of pupil diameter[190, 191]. In dynamic audiovisual environments, luminance effects may need to be modelled explicitly before smaller attention-related pupil effects can be interpreted[192]. Pupil size cannot therefore be interpreted as cognitive effort unless the visual environment is sufficiently controlled or luminance is explicitly logged and incorporated into the analysis. A second confound is arousal and affective state. Pupil size is influenced by noradrenergic activity, orienting responses, emotional salience, stress and autonomic arousal[172, 173], and these processes are directly relevant to chronic pain because pain, anxiety, fatigue, distress and sensory threat can all alter arousal state. A pupil response may indicate cognitive effort, but it may also reflect emotional salience, anxiety, discomfort or heightened vigilance. This ambiguity does not make the signal unusable, but it requires inference to be task-specific and probabilistic, with ocular signals interpreted alongside task events, behavioural performance, symptom context and environmental parameters. Pain itself adds further complexity. Acute nociceptive stimulation alters pupil responses and autonomic activity, and autonomic responses to noxious stimuli reflect subjective experience as well as nociceptive input[193, 194]. Reviews of autonomic arousal and pain show that physiological responses are not specific to pain and are influenced by stress, attention, emotion and context[133]. If pupil or blink features differ between chronic pain and healthy control groups, the difference cannot be attributed automatically to cognitive load and may reflect pain, fatigue, anxiety, ocular factors, task engagement or some combination of these.

Existing eye tracking work in chronic pain indicates that attentional measures are task and stimulus-dependent, with meta-analytic evidence of chronic pain-related attentional biases to emotional and pain-related information that vary by paradigm and stimulus[195]. This supports the rationale for state-sensitive measurement during VR while highlighting the need for validation within the target population. Task context is therefore central. The same pupil or blink pattern may have different meanings depending on whether the user is completing a working memory task, relaxing in a nature environment, performing a movement exercise or engaging with

pain education. Eye tracking-based inference should be described as estimation of task-defined state, not measurement of a pure latent cognitive construct. Validation must therefore test whether eye tracking features support generalisable inference in the intended population and task context, recognising that a given task level may impose greater subjective effort in someone with chronic pain than in a healthy control, and that reduced performance may reflect disengagement, overload, visual discomfort, medication or misunderstanding rather than cognitive load alone.

Eye tracking should therefore be framed as a pragmatic embedded control signal rather than a mechanistically pure biomarker. It offers a favourable balance of deployability, temporal resolution, low user burden and compatibility with VR, and pupil and blink metrics provide plausible links to cognitive effort, attention, arousal and fatigue, while remaining strongly influenced by luminance, task design, pain, affective state, ocular factors and headset-specific data quality. The empirical question addressed in Chapter 7 is therefore not whether eye tracking measures cognitive load directly, but whether, under controlled VR conditions, it can support reliable inference of task-defined state across individuals with and without chronic pain.

2.3.5 Cognitive load as an adaptive target

Cognitive load is a candidate adaptive target for immersive therapeutics because several proposed mechanisms of VR benefit depend on maintaining an appropriate relationship between task demand, attention and available cognitive resources. Cognitive load is not a direct measure of pain and should not be interpreted as a discrete physiological state. It is instead a task-dependent functional construct that reflects the balance between what the virtual task requires and what the user can currently tolerate, sustain and perform. The same environment or task may be engaging for one person, insufficiently demanding for another and overwhelming for a third, and the same task may be tolerable for one individual at one moment and overwhelming for the same individual at another, depending on pain intensity, fatigue, anxiety, sleep, sensory sensitivity or familiarity.

Cognitive load should be distinguished from, but linked to, cognitive effort and attention. Cognitive load describes the demands placed on working memory by the task and its context, cognitive effort describes the mental exertion invested in

meeting those demands, and attention describes the allocation of processing toward task-relevant information. These constructs overlap in practice and are difficult to separate using self-report alone, which strengthens the rationale for task-linked behavioural and physiological inference rather than reliance on a single subjective rating[196-198].

Cognitive load theory provides a theoretical starting point. It conceptualises performance as constrained by the demands placed on working memory[198]. A task that requires sustained attention, updating, response selection or inhibition imposes demands that must be managed within limited processing capacity. Load is not equivalent to task difficulty. It reflects the interaction between task structure, prior knowledge, sensory context, motivation, affective state and available cognitive resources. In VR, this interaction is salient because cognitive demand is embedded within a multisensory environment, and the user is processing competing sensory inputs while completing a task. Flow theory expresses a related principle, proposing that sustained engagement is most likely when challenge is matched to perceived skill and capability[199]. The therapeutic implication is that absorbed engagement may depend on preserving a tolerable balance between challenge and capability, rather than on maximising challenge.

The relevance of cognitive load to pain modulation follows from the established interaction between pain and attention discussed in Section 2.1.4. Pain demands attention and can interrupt ongoing cognitive activity, while competing cognitive tasks reduce the salience and subjective intensity of pain. Pain and cognition-related networks can remain concurrently active, with the balance between them depending on pain intensity and cognitive task demand. Attentional modulation is therefore not a binary process. The degree of modulation depends on whether the task recruits sufficient attention without exceeding the individual's current capacity, producing a non-linear problem in which both underload and overload may be suboptimal. A task that is too easy may fail to capture attention, generate absorption or compete meaningfully with pain-related processing. A task that is too difficult may impair performance, increase frustration, provoke fatigue, reduce perceived control or lead to disengagement. Experimental evidence supports this view. Cognitively demanding tasks reduce pain more effectively when participants are sufficiently motivated to engage, whereas low motivation weakens the analgesic value of increased task

demands[105]. Conceptual work on flow-induced analgesia proposes that pain modulation may be supported by absorbed engagement, reward, affective regulation and a tolerable balance between challenge and capability, potentially interacting with descending modulatory systems[200].

This balance is consequential for VR. Immersive tasks can be made more or less demanding by altering stimulus complexity, pacing, working memory requirement, visual salience, movement demand, feedback and response constraints. These features are therapeutic opportunities and potential sources of burden. Experimental work on VR attentional modulation suggests that increasing task demand does not always produce greater analgesia. Some studies indicate that more demanding or interactive VR may improve outcomes, while others suggest that the effect depends on motivation, context and user capacity[105, 201, 202]. The therapeutic value of VR therefore depends on how its design features interact with the user's state, rather than on immersion or difficulty alone.

Chronic pain complicates the issue because pain interacts with the cognitive state of the user. Chronic pain is associated with disruption of attention, working memory and executive function[61]. Systematic review and meta-analytic evidence indicate that chronic pain is associated with working memory deficits, and broader reviews support links between chronic pain and cognitive impairment across attention, processing speed, memory and executive control[203-205]. These associations are consequential because many VR interventions depend on the same cognitive functions that chronic pain may disrupt. A task designed to be moderately demanding for a healthy user may impose excessive load on someone with persistent pain. Fibromyalgia is highly relevant because cognitive symptoms are an important part of the clinical phenotype[41]. Fibromyalgia studies report working memory difficulties and associations between cognitive performance, pain, fatigue and clinical variables[206], and experimental evidence suggests that stress can specifically worsen working memory performance in fibromyalgia and peripheral neuropathic pain, indicating that cognitive capacity is not fixed but vulnerable to affective and physiological state[207]. Ecological and longitudinal evidence shows that pain and cognitive function fluctuate within individuals with fibromyalgia, reinforcing the need to consider momentary state rather than relying on baseline assessment alone[36].

If cognitive capacity and symptom burden fluctuate, fixed task difficulty is unlikely to remain appropriate throughout an intervention. A user may begin a session able to tolerate a cognitively demanding task and become overloaded as fatigue develops. Another may require greater challenge to maintain absorption. A third may need reduced sensory complexity or slower pacing because anxiety, pain flare or sensory sensitivity has increased the burden of the virtual environment.

Cognitive load is therefore a useful adaptive target because it is more directly linked to the controllable properties of a VR task than broader clinical states such as pain severity, mood or quality of life. Pain may fluctuate during VR use, but it is influenced by multiple biological, affective, contextual and interpretive processes that cannot be adjusted directly by changing a single task parameter. Cognitive load can be deliberately shaped within a session through design features such as task difficulty, stimulus rate, working memory demand, response timing, visual complexity, feedback and environmental intensity. These features may be adjusted without changing the overall therapeutic aim. An attentional modulation task may be made easier when the user shows signs of overload, or more demanding when engagement appears to decline. A relaxation environment may reduce sensory complexity when arousal or distress is high. A behavioural learning task may preserve challenge while preventing deterioration in performance. In each case, adaptation aims to maintain a therapeutic operating range rather than to maximise difficulty. Some cognitive demand is necessary for engagement, learning and absorption. Excessive demand may undermine tolerability, performance and adherence. Insufficient demand may lead to boredom, mind-wandering or renewed salience of pain. The adaptive aim is therefore to keep the task within a range that supports the intended mechanism, recognising that this range is likely to differ across people and across sessions.

Cognitive load measurement also has clear limits. Estimating cognitive load during VR does not explain the biological mechanisms of pain relief, and it should not be treated as a marker of nociplastic pain, descending modulation or treatment response. It is a pragmatic state estimate relevant to intervention delivery. If a system can infer that a user is likely underloaded, overloaded or appropriately challenged, it could adjust the virtual task to preserve engagement and tolerability.

Working memory tasks are useful in this context because they provide a structured way to manipulate load and generate task-defined labels. The N-back paradigm, in which participants identify whether the current stimulus matches one presented N positions earlier, has been used widely in cognitive neuroscience and applied workload research because the N parameter offers a direct manipulation of working memory load while keeping perceptual and motor demands relatively constant[208, 209]. Although N-back performance reflects multiple processes, including working memory updating, attention control, response strategy and stimulus timing, and should not be treated as a pure measure of any single construct[210], its parametric structure makes it well suited to generating labelled training data for state inference and to evaluating whether sensor signals track graded task demand. This rationale motivates the use of a graded VR N-back paradigm in Chapter 7.

2.3.6 Machine learning for state inference

Closed-loop VR systems require a technique to accurately infer whether complex headset-derived signals contain sufficient information to indicate a specific task-defined state to inform downstream adaptation logic[135, 211]. Sensor data acquired within VR are dynamic, multivariate and high-dimensional. Pupil size, blink behaviour, gaze patterns and behavioural responses unfold over time and interact with task context, and simple aggregate summaries such as mean pupil diameter or overall blink rate may obscure dynamic structure that carries state-relevant information. Sophisticated analytical approaches that preserve temporal structure, such as time-series machine learning, are therefore more likely to recover the relationships needed for inference within a closed-loop system.

Time-series classification has advanced considerably over the last decade, with families of algorithms including dictionary-based methods, convolutional kernel approaches, interval-based ensembles, hybrid meta-classifiers, and deep learning models delivering benchmark performance across diverse domains[211-213]. A common evaluation framework allows these contrasting inductive biases, frequency-based, shape-based, interval-based and deep-learning representations, to be compared on the same headset-derived data and against transparent baselines, without bespoke implementation[211, 213].

Two evaluation principles are central. The first is subject-independent generalisation. Adaptive VR systems must operate on previously unseen users, and within-subject cross-validation can substantially overestimate generalisation by allowing temporal autocorrelation and individual idiosyncrasies to leak between training and test data. Leave-one-subject-out cross-validation provides a more demanding test, training on a population of subjects and evaluating on a held-out individual, and is appropriate when the eventual deployment target is a system that has not yet observed the new user. The second is honest baseline comparison. Benchmark studies in time-series classification have repeatedly shown that simple baselines, including logistic regression on engineered features, can be surprisingly competitive on small or noisy datasets[211, 212]. A classifier should not be considered useful unless it adds value beyond a transparent feature-based baseline.

Machine learning outputs should also be interpreted with care. A model that predicts cognitive task condition, performance state or workload-related class from eye tracking does not prove that the underlying signal is a pure measure of any latent construct, nor that the model identifies the mechanism of analgesia. Predictive success demonstrates that signals available during VR use contain information relevant to state estimation under defined task conditions, which is the kind of evidence required before eye tracking can be considered a candidate input for future adaptive control. This framing motivates the analytical approach taken in the VR-FOCUS study (Chapter 7), which compares conventional and time-series classifiers under subject-independent evaluation.

2.3.7 Alternative peripheral signals, briefly considered

Several peripheral and behavioural signals could contribute to adaptive immersive therapeutics, including functional near-infrared spectroscopy (fNIRS), electromyography (EMG), electrodermal activity (EDA), cardiac measures, movement tracking, respiratory signals, skin temperature, photoplethysmography (PPG) and multimodal wearable streams. Their relevance lies not in providing direct measures of pain, but in capturing physiological or behavioural states that influence whether a VR intervention is appropriately calibrated. Systematic reviews of physiological mental workload measurement indicate that cardiac, electrodermal, respiratory, ocular, muscular and neural signals have all been used to estimate task

demand or mental workload, with validity varying substantially by task, sensor, analytic method and population[188, 189, 214, 215].

fNIRS provides a portable central neurophysiological alternative to EEG and fMRI, measuring cortical haemodynamic responses with sensitivity to cognitive state and workload, including in immersive VR working memory paradigms[216, 217], and has been used to monitor pain-related cortical processing in prefrontal and somatosensory regions[218]. As a pragmatic sensor to inform closed-loop systems it is limited because it still requires additional head-mounted instrumentation, has lower temporal resolution than EEG, is vulnerable to extracerebral physiological confounds and is not routinely integrated into consumer VR systems. EMG can capture muscle activation, guarding, co-contraction, relaxation and movement-related motor patterns, and surface EMG has been used to examine flexion-relaxation behaviour and lumbar muscle function in low back pain, with systematic review evidence supporting its biomechanical relevance while also highlighting variability in reliability and diagnostic utility[219]. EMG is more closely aligned with peripheral motor behaviour than with headset-embedded cognitive state estimation, and usually requires additional instrumentation, correct sensor placement and management of motion artefact. Recent developments in wrist-worn surface EMG, including consumer neural-input bands[220] and large-scale work showing that generic decoding models can support computer interaction across users without person-specific calibration[221], suggest EMG could become a feasible sensor in future closed-loop immersive systems, although it lies outside the scope of the present empirical work.

EDA, often described as galvanic skin response, indexes sympathetic sudomotor activity and is widely used as a measure of physiological arousal[222, 223]. Skin conductance may reflect arousal, stress, threat, engagement or task demand, and could support relaxation, emotional regulation or distress monitoring. EDA is not specific to attention or cognitive load, and its utility for rapid adaptation is constrained by response latency, habituation, thermoregulation, movement and skin contact[222, 223]. Cardiac measures, including heart rate, heart rate variability (HRV), electrocardiography and PPG-derived pulse metrics, may provide information about autonomic regulation, arousal, cognitive effort, relaxation, recovery and intervention tolerability[188, 189, 214, 215]. HRV is relevant to relaxation and self-regulation paradigms, where changes in parasympathetic or vagal regulation may correlate with

downregulation rather than directly index pain relief[224]. Both EDA and cardiac signals are influenced by respiration, posture, medication, fitness, sleep and other contextual factors, and frequency-domain HRV requires sufficient recording duration and stable conditions, which limits utility for rapid within-session adaptation[224]. Sensors for cardiac, EDA and EMG are not routinely embedded in current VR systems, leaving them better positioned as complementary modalities than as primary control signals.

Movement tracking is relevant to chronic musculoskeletal pain because movement avoidance, kinesiophobia, reduced activity and altered biomechanics are central therapeutic targets[39]. VR systems already capture head and often hand and body movement, while external or wearable inertial sensors extend tracking to trunk, limb, gait and posture[225]. Systematic review evidence in cLBP suggests that motion sensor-based remote rehabilitation and VR may support real-time movement tracking, biofeedback, personalised exercise and adherence, while also indicating the need for adequately powered longitudinal trials and economic evaluation[226]. Movement data show how a user behaves but not why, since reduced movement may reflect fatigue, fear, pain, low confidence, biomechanical limitation or disengagement. Movement is therefore best viewed as a readily available behavioural signal for future multimodal systems rather than as a primary measure of cognitive state.

Other wearable signals, including respiration, skin temperature, sleep, physical activity and multisensor smartwatch-derived features, may be useful for between-session adaptation and longitudinal phenotyping. Fitbit data have been used within a 10-week mind-body and quota-based pacing programme for adults with chronic pain, where step-goal achievement was associated with improvements in physical function and anxiety[227]. Large-scale cLBP work combining ecological momentary assessment with wearable motion sensors has shown that real-world sensor data can characterise activity and sedentary behaviour while revealing discrepancies between perceived and objectively measured activity[228]. Consumer and research-grade wearables could therefore support future VR therapeutic pathways by tracking whether repeated use is accompanied by changes in activity, sleep, pacing or recovery outside the headset, but they lie outside the scope of the empirical work that follows.

2.3.8 Design requirements for future adaptive systems

Future adaptive immersive therapeutics require a design architecture that links mechanistic rationale, deployable sensing, analytical inference and therapeutic adaptation. The central challenge is not simply to add sensors to VR, but to determine which measurements are suited to which translational role. Mechanistic measures, including rsEEG, can identify neurophysiological processes that may be altered in chronic musculoskeletal pain and modifiable through immersive intervention. Embedded measures, including eye tracking and task performance, are better suited to estimating user state during VR use because they can be captured continuously within the therapeutic environment[67, 166].

Adaptive systems also require analytical approaches capable of handling temporally structured, multivariate data. State during VR is unlikely to be captured by a single feature or static threshold, because cognitive load, engagement, arousal, fatigue and performance unfold over time and interact with task context. Machine learning is relevant because it can test whether complex headset-derived signals contain sufficient information to support subject-independent prediction of task-defined state[135, 211]. Predictive accuracy, generalisability and robustness across individuals are therefore important, while outputs must retain enough interpretability and clinical meaning to guide adaptation.

The adaptation logic itself must be explicit. Future systems require defined decision points, tailoring variables and modifiable intervention components, so that changes in task difficulty, pacing, sensory intensity, movement demand or feedback are linked to a plausible therapeutic purpose[124, 229]. Evaluation will need to move beyond average effects of fixed VR packages to examine when, for whom and under what state conditions adaptive components are beneficial[230], in keeping with established methodological guidance for complex interventions[125].

Adaptive VR must also remain tolerable, acceptable and deployable. Systems designed for chronic pain populations must account for fatigue, sensory sensitivity, disability, technology confidence and cybersickness risk[111, 112]. They must be evaluated across diverse users and settings and designed for digital inclusion, so that home-based delivery does not widen the inequalities to which the populations most affected by chronic pain are already exposed.

2.3.9 Rationale for the empirical studies

Static immersive therapeutics are poorly matched to the dynamic nature of chronic musculoskeletal pain, and optimisation of these therapies relies upon a transition toward closed-loop systems that adapt to measured user state. This requires three separable competencies, each answering a different question. Mechanistic characterisation identifies what is biologically altered in chronic pain and is therefore worth targeting. State-sensitive measurement captures the user's functional state during intervention delivery in a form that can inform adaptation. Predictive inference converts temporally structured sensor data into decision-relevant estimates of that state. These competencies are governed by different selection criteria and are unlikely to be satisfied by any single signal.

The empirical work is organised around this decomposition. Rather than holding population and signal constant, each study matches its cohort, measure and analytical approach to the competency it addresses. rsEEG is used to investigate mechanistic neurophysiology in the widespread, centrally mediated phenotypes in which oscillatory abnormality is most plausible. Headset-derived eye tracking, interpreted through time-series machine learning, is used to test whether a deployable embedded signal can support inference of task-defined state across a clinically heterogeneous pain cohort. No study attempts to deliver a working closed-loop system; the aim is to establish the components on which one could credibly be built.

PainWaves (Chapter 5) addresses mechanistic characterisation in chronic widespread pain. Case-control rsEEG findings in chronic pain remain inconsistent in direction and topography, with studies failing to account for the genetic and environmental factors that may influence rsEEG signals, or to adjust for the aperiodic background which can conflate results. PainWaves uses a discordant co-twin design, which partly holds genetic and shared-environmental background constant, and applies periodic-aperiodic decomposition to separate oscillatory effects from the broadband background, testing which spectral differences are more plausibly pain-related.

VIPA-EEG (Chapter 6) extends mechanistic characterisation from description toward modulation. Identifying candidate neurophysiology is of limited value unless that

physiology is responsive to an intervention. VIPA-EEG examines whether candidate rsEEG features change around immersive VR exposure in fibromyalgia, a preliminary test of whether immersive exposure is accompanied by measurable change in candidate neurophysiology and therefore whether any such feature could be treated as a modifiable target.

VR-FOCUS (Chapter 7) addresses state-sensitive measurement and predictive inference. The reliability of ocular features within VR, their interpretation in chronic pain, and the incremental value of time-series methods over aggregate summaries all remain underexplored. This information is important when considering deployable systems that generalise to users for whom no training data exist. VR-FOCUS evaluates whether headset-integrated eye tracking, analysed under subject-independent cross-validation, can support inference of task-defined cognitive load in adults with and without chronic musculoskeletal pain.

These studies address the three competencies with empirical evidence, while leaving their integration to future work. The aims that follow set out what each study was designed to achieve.

3. Aims

The overarching aim is to investigate whether sensor-derived measures, interpreted using conventional statistical and machine learning frameworks, can contribute to the development of future closed-loop VR therapeutics for chronic musculoskeletal pain. The specific aims are:

1. To characterise rsEEG differences associated with CWP using a discordant co-twin control design that partly accounts for shared genetic and environmental background.
2. To examine whether rsEEG features change around immersive VR exposure in people with fibromyalgia.
3. To investigate whether headset-derived eye tracking acquired during a VR cognitive task can support subject-independent inference of task-defined cognitive load in healthy controls and in a chronic musculoskeletal pain cohort.

4. Core methods: EEG, eye tracking, and machine learning

This chapter summarises common methodologies used within the empirical work, which comprises three studies. PainWaves (Chapter 5) examines resting-state EEG (rsEEG) in twins discordant for chronic widespread pain (CWP). VIPA-EEG (Chapter 6) measures rsEEG around immersive VR exposure in fibromyalgia. VR-FOCUS (Chapter 7) evaluates headset-derived eye tracking during a VR cognitive task in adults with and without chronic musculoskeletal pain. The chapter describes the technologies, data preparation steps and analytical pipelines used, and clarifies how these shared approaches were applied across the EEG, VR eye tracking and machine learning components, while leaving study-specific implementation detail to the individual chapters that follow.

4.1 EEG framework

4.1.1 Acquisition

EEG data were acquired using Brain Products LiveAmp amplifiers (Brain Products GmbH, Germany) with BrainVision Recorder Professional software (v1.23.0003)[231] at a 500Hz sampling rate. Two electrode configurations were used across the EEG studies, reflecting the distinct acquisition contexts of PainWaves (Chapter 5) and VIPA-EEG (Chapter 6).

PainWaves used the 32-channel R-Net system (Brain Products GmbH, Germany)[232], a sponge-based passive electrode cap conductive through saline solution rather than gel. R-Net preparation typically requires under 15 minutes, which made it feasible within the existing TwinsUK multimodal research workflow that included downstream MRI scanning, and within the operating constraints of acquisition by research nursing staff without prior specialist EEG experience. The reduction in spatial sampling relative to higher-density configurations was an accepted trade-off given the planned sensor-level spectral focus of the study.

VIPA-EEG used the 64-channel actiCAP active electrode gel-based system (Brain Products GmbH, Germany)[233], selected for the higher signal stability and spatial sampling appropriate to a controlled repeated-measures intervention design. Setup typically required around 45 minutes per participant, which was compatible with the parent feasibility study[8] workflow but would not have scaled to the PainWaves acquisition context.

In both studies, electrode placement followed the international 10-20 system, impedances were checked prior to recording and adjusted as required, and event markers were inserted to denote the onset of resting-state conditions. Analyses focused on eyes-closed recordings throughout to minimise concurrent sensory input and visual processing while improving sensitivity to intrinsic oscillatory activity.



Figure 4.1. EEG acquisition hardware used across the rsEEG studies

The left panel shows the Brain Products LiveAmp amplifier used with both EEG systems. The second panel shows the R-Net sponge-based EEG cap used in PainWaves. The third panel shows a 64-channel actiCAP gel-based EEG cap used in VIPA-EEG. The right panel shows the actiCAP gel application process. Images provided by Brain Products GmbH and reproduced with permission.

4.1.2 EEG preprocessing pipeline

EEG data were preprocessed using an automated pipeline adapted from the DISCOVER-EEG framework[234], implemented in MATLAB R2025b[235] using EEGLAB[236] and FieldTrip[237]. Automated preprocessing was selected to reduce operator-dependent variability, improve consistency across datasets and support reproducibility. DISCOVER-EEG was chosen because it provides an open, fully automated pipeline designed for biomarker discovery in clinical EEG datasets, with documented use in chronic pain EEG research[128, 234].

Prior to preprocessing, raw recordings were converted to BIDS format[238] using MNE[239] and MNE-BIDS[240], with standardised subject identifiers, recording metadata, channel information and coordinate files generated where available. Only eyes-closed segments were retained using event markers to determine recording onset. Where event markers were absent, a deterministic fallback offset from recording onset was applied to preserve reproducibility. Electrode coordinates derived from manufacturer layout files (.bvef) were remapped to a head-centred Cartesian convention within EEGLAB[236] and normalised within FieldTrip[237] to ensure consistent

scalp geometry for spatially informed analyses before formal preprocessing was completed.

The preprocessing sequence comprised eight steps applied uniformly across datasets. Continuous data were downsampled to 250Hz to reduce computational load while preserving sufficient temporal resolution for spectral analyses up to 40Hz. Line noise was attenuated using the CleanLine algorithm[241] with an additional 49-51Hz notch filter to remove residual mains interference. Noisy or non-physiological channels were identified using DISCOVER-EEG criteria incorporating signal variance, flatlining and channel-wise statistical deviation, and removed before further processing. Data were then re-referenced to the common average reference. Independent component analysis was performed using the extended infomax algorithm[242] with two repetitions, and components were automatically classified by ICLabel[243] with those identified as blink or muscle artefact removed. Removed channels were reconstructed using spherical spline interpolation[244]. Residual high-amplitude artefacts were attenuated using Artifact Subspace Reconstruction[245] (ASR) with a conservative burst criterion of 35 standard deviations, meaning that only transient signal subspaces showing large deviations from the reference EEG covariance structure were corrected. After ASR, residual contaminated periods were removed when more than half of the channels in a time window exceeded an upper abnormality threshold of 1.5 robust standard deviations from the reference windowed amplitude/power distribution. Finally, continuous data were segmented into non-overlapping 2s epochs. Segmentation into 2s epochs is consistent with prior rsEEG work in chronic pain, where 2s epochs have been used to provide sufficient frequency resolution for spectral estimation, and support robust averaging across segments[246, 247]. Overall, conservative parameter selection after iterative testing was used to maximise usable EEG retention while providing sufficient artefact attenuation for spectral analysis. This was consistent with recent evidence cautioning that unnecessarily aggressive EEG preprocessing can attenuate meaningful neural signal as well as artefact[248].

Preprocessed datasets were saved as EEGLAB-compatible files preserving epoched data, channel metadata and provenance. Audit logs were generated for each participant, recording the number of channels removed, retained post-clean duration, proportion of rejected data and number of final epochs.

4.1.3 Quality-control framework

Quality-control was treated as part of the analytical framework. A tiered sensitivity approach was applied to the EEG studies, with progressively stricter inclusion thresholds applied to the same inferential framework so that any change in findings reflected data inclusion rather than analytical inconsistency.

Tier 0 represented the full preprocessed dataset with no participants excluded. Subsequent tiers applied increasingly stringent thresholds based on number of channels removed and retained post-clean duration, excluding participants (or twin pairs in PainWaves) that did not fulfil these thresholds. Tier thresholds were defined a priori based on data distributions and the requirements of downstream analyses, with study-specific details provided in the corresponding study chapters.

Effects that remained stable or strengthened across stricter tiers were interpreted as more likely to reflect a true physiological signal, whereas effects that attenuated under stricter thresholds were treated more cautiously. This approach balanced statistical sensitivity with signal stability and supported transparent reporting of inclusion decisions.

4.1.4 Spectral analysis

Spectral analysis was used to quantify resting-state oscillatory activity within predefined frequency bands and to provide a consistent basis for subsequent topographical, region-of-interest (ROI), alpha peak, and aperiodic-periodic analyses across the EEG studies. All spectral analyses were conducted using an analogous framework to maximise comparability between datasets and to support cross-study synthesis.

Power spectral density (PSD) was estimated from preprocessed 2s epochs using a Fourier-based approach with Hanning tapering. Spectral estimates were calculated at the epoch level and averaged within participants in linear power space before conversion to decibel units, avoiding distortion that can arise when averaging log-transformed spectra[249].

Analyses were restricted to 1-40Hz to reduce sensitivity to ultra-slow drift and non-neural low-frequency artefact at the lower bound and to limit muscle-related contamination at the upper bound. This frequency range therefore reflected a balance between physiological coverage and signal reliability[250]. Frequency bands

were defined consistently across studies as delta (1-3Hz), theta (4-7Hz), alpha (8-12 Hz), broad beta (13-30Hz) with sub-bands low beta (13-20Hz) and high beta (21-30Hz), and gamma (31-40Hz). Non-overlapping boundaries were used throughout to avoid ambiguity in band assignment and to ensure that adjacent bands did not share spectral bins.

ROI analyses were used to further quantify and interpret effects identified through cluster-based testing rather than as primary inferential endpoints. A frontocentral ROI (Fz, FC1, FC2, Cz) was selected because cluster-based EEG effects across the thesis were often centred on, or extended into, midline/frontocentral sensors. This region also overlaps with scalp locations commonly used to index central sensorimotor and frontocentral oscillatory activity. The ROI choice was further supported by chronic pain EEG studies reporting frontocentral beta and central theta alterations in musculoskeletal pain, and beta/theta abnormalities in fibromyalgia involving middle frontal, midcingulate and central-parietal networks[251, 252]. Using a common frontocentral ROI therefore provided a biologically plausible and consistent summary region for effect-size estimation, avoiding selection of individual peak sensors from significant clusters, and supporting comparability across studies.

Alpha peak characteristics were examined as a complementary measure of oscillatory activity, providing information not captured by band power alone[148, 253]. Two ROIs were used: the posterior ROI (O1, O2, Oz, POz, Pz, PO3, PO4) to capture canonical occipital alpha, and a secondary central ROI (C3, Cz, C4) targeting sensorimotor alpha given prior pain research suggesting that central alpha abnormalities may be relevant in chronic pain[254]. Within each ROI, the dominant peak in the 8-12Hz range was identified from the ROI-averaged spectrum, with peak frequency and peak power refined by parabolic interpolation. Alpha peak measures were treated as secondary and interpretive, supporting differentiation between changes in peak prominence and shifts in dominant frequency.

4.1.5 Aperiodic-periodic decomposition

EEG power spectra were also decomposed into aperiodic and periodic components using the Fitting Oscillations and One-Over-F (FOOOF; now specparam) framework, which addresses evidence that conventional band power can conflate genuine oscillatory activity with shifts in the broadband 1/f-like aperiodic component[157, 255].

Spectral parameterisation modelled the power spectrum as the combination of an aperiodic background characterised by an offset and exponent, and superimposed periodic peaks fitted as Gaussians with estimates of centre frequency, power and bandwidth. FOOOF was applied to ROI-averaged spectra over 2-40Hz to reduce dimensionality and improve fitting stability. The aperiodic component was modelled using a fixed (non-knee) formulation, which assumes an approximately linear 1/f-like relationship in log-log space and was selected because the analysis range made low-frequency knee-like curvature unlikely to dominate the fit[256]. Peak-fitting parameters were chosen to balance sensitivity to physiologically plausible peaks with robustness to noise.

This decomposition allowed spectral effects to be interpreted with greater specificity. Offset and exponent were used to determine whether group or condition differences were better characterised as broadband aperiodic changes. Residual, aperiodic-adjusted band power was used as a secondary confirmatory measure to assess whether conventional band-power effects persisted after removal of the fitted aperiodic component. Effects that persisted after aperiodic adjustment were interpreted as more consistent with band-limited periodic activity, whereas effects that were attenuated after adjustment were interpreted more cautiously as potentially reflecting broader aperiodic spectral differences rather than oscillatory activity alone.

4.2 Eye tracking and VR design framework

4.2.1 Hardware and software choices

VR-FOCUS (Chapter 7) used the Pico Neo 3 Pro Eye headset (Pico Interactive)[257], a standalone consumer VR system with integrated Tobii binocular eye tracking[258] and onboard processing. The headset has dual 1832 x 1920 LCD displays, a 90Hz refresh rate and a nominal horizontal field of view of 98°. Eye tracking data were acquired at 60Hz for most participants and 90Hz for a later subset following a system update, with subsequent harmonisation as described in 4.2.4.

The VR application was developed in Unity 2021.3.45f1[259] using the Universal Render Pipeline, with the Tobii Ocumen SDK (v2024.1.0)[260] used to access binocular pupil size, gaze direction and blink-related outputs. Deployment to the headset used the PICO XR Unity Integration SDK with Android build settings configured for IL2CPP, ARM64 and Pico-only XR plugin management. The application ran offline on

the headset rather than via a tethered PCVR workflow, with output mirrored to a research laptop via USB-C for real-time researcher observation.

Temporal resolution was an important consideration in VR system selection because the intended analyses extended beyond static gaze estimation to task-evoked pupil and blink dynamics as candidate markers of cognitive load. Cognitive pupillometry examines phasic, stimulus-linked changes in pupil size, making adequate temporal granularity important despite pupil dynamics being slower than many eye-movement measures[174, 261]. This is particularly relevant in immersive tasks, where cognitively driven pupil changes must be interpreted alongside luminance-related responses[187]. A sampling rate of 60Hz was therefore treated as the minimum acceptable acquisition rate, given that visually evoked pupil responses emerge on the order of hundreds of milliseconds[261] and that 60Hz sampling has direct precedent in working-memory pupillometry using N-back paradigms and pupil-based cognitive load classification[262, 263]. An alternative VR system (HTC Vive XR Elite) was piloted during early development but was not adopted because available pupil sampling was insufficient for the intended analyses.

Calibration used the Tobii Ocumen workflow embedded within the Unity application. An initial position-guide step aimed to optimise headset placement and lens separation, followed by a 5-point gaze-based calibration. Calibration was considered successful when all points showed bias of no more than 3° and precision of no more than 1° for both eyes, with all samples valid. Where criteria were not met, calibration was repeated. Calibration could also be relaunched between tasks if headset position or visual clarity appeared to have shifted. Calibration metrics were exported to CSV for independent quality-control review.

4.2.2 Cognitive task framework

VR-FOCUS used a graded N-back working-memory paradigm[171, 264, 265] to elicit parametrically varying levels of cognitive load within the same immersive system used for eye tracking acquisition. The N-back task requires participants to monitor a continuous sequence of stimuli and indicate when the current item matches one presented N steps earlier, with demands on working memory updating, maintenance and attentional control increasing with N-back level. The paradigm was selected because it provides a well-established and scalable manipulation of working-memory

demand suitable for repeated within-session sampling under controlled visual conditions. The specific load range, stimulus parameters, response logic, lure structure and counterbalancing scheme used in the present study are reported in Chapter 7.

4.2.3 Signal considerations

Pupil size, blink-related signal interruption and gaze direction are all available from headset-integrated eye tracking. Their analytical roles within the thesis differ, and the rationale for emphasising pupil dynamics and blink behaviour over gaze in the present application is developed in Chapter 2 (Section 2.3.4). Briefly, pupil size is treated as the principal signal of interest, blink-related metrics are treated as informative both as a state feature and as a determinant of pupil signal completeness, and gaze is recorded but not prioritised analytically given the centrally constrained visual design of the VR-FOCUS task.

Luminance is logged as a contextual variable rather than as a primary predictor. The VR-FOCUS task panel was deliberately designed to stabilise scene brightness across blocks (Chapter 7), and frame-level luminance was computed from RGB values rendered to a low-resolution 512 x 512 render texture using a standard linear transform (Lum = 0.2126R + 0.7152G + 0.0722B). Whole-scene, Gaussian-weighted and gaze-weighted luminance approaches were explored during development; whole-scene luminance was used in the final workflow because the task design rendered the more computationally intensive approaches unnecessary.

4.2.4 Preprocessing

Raw .ocumen files exported from the headset were converted in Python to participant-level CSV representations containing pupil diameter and gaze-vector data for both eyes, then combined into participant-level workbooks. The staged conversion preserved raw recordings while generating progressively analysis-ready derivatives.

Recordings acquired at 90Hz were harmonised to 60Hz using nearest-neighbour downsampling rather than interpolative resampling. Nearest-neighbour selection preserved empirically observed measurements within the original signal, whereas interpolation would have introduced synthetic values. Downsampling was applied within task blocks using the recorded time axis to construct the target 60Hz grid.

Blink preprocessing was treated as both an artefact-handling step and a feature-relevant operation. Candidate blink segments were identified from contiguous missing periods in pupil or gaze channels and filtered using sampling-rate-specific frame thresholds and physiologically plausible duration criteria (50-500ms)[174, 266]. Segments shorter than 50ms were excluded as artefactual; segments longer than 500ms were classified as data-loss periods and removed rather than interpolated.

For valid in-range blink segments, interpolation boundaries were extended beyond the strict period of complete signal loss to account for blink onset and offset slopes, which are commonly treated as part of the blink artefact in pupillometry[267-269]. A local SD-based approach was used: for each candidate blink, reference windows were sampled at 100ms before blink onset and 150ms after blink offset, with up to five valid samples from each side used to estimate a local median and standard deviation. Adjacent samples were added to the interpolation mask when they deviated by more than 3 standard deviations from the local median, allowing the interpolation region to extend where the signal showed blink-related distortion while retaining a local rather than global definition of onset and offset.

Blink reconstruction used cubic spline interpolation as the primary method, requiring up to two valid anchor samples on each side of the masked interval. Where these four anchors were unavailable, for example near task-block boundaries or where neighbouring data-loss periods left insufficient valid samples, linear interpolation between the nearest available bounding samples was used. The final dataset retained channel-specific interpolation labels (0 for observed data, 1 for samples reconstructed within the blink mask), preserving traceability of blink-affected samples.

Two parallel feature representations were generated from the preprocessed data. Summary features comprised mean pupil diameter and blink rate at task-block level, with performance-derived behavioural features (hits, misses, false alarms, discrimination index d′, response bias c and reaction times) computed using standard signal-detection conventions and linked to the corresponding ocular summaries. Time-series representations preserved continuous multivariate sequences for use in machine learning analyses, with equal-length segments enforced through padding or trimming where required, and missing values imputed using per-case, per-channel means to avoid leakage across participants. A binary

blink-mask channel was developed to preserve explicit blink timing within the feature representation and was evaluated as an additional input alongside ocular signals.

4.3 Predictive modelling and machine learning framework

4.3.1 Rationale

Machine learning was used because the primary question addressed in VR-FOCUS concerned individual-level prediction rather than population-average inference. Mixed-effects models can establish whether measures such as mean pupil diameter or blink rate differ across tasks at a group level, but such differences do not demonstrate that ocular signals from a previously unseen individual contain sufficient information to identify the task condition associated with them. This distinction is important for adaptive VR systems intended to infer user state during use. Ocular signals may also contain information in their temporal structure, including the timing, magnitude and consistency of responses across a task block, which is lost when a sequence is reduced to a single summary statistic. Time-series classification therefore provided a means of testing whether discriminative information was present within these temporal patterns and whether it generalised out of sample.

The machine learning framework in VR-FOCUS was consequently designed to evaluate whether time-resolved eye tracking signals contained information relevant to task-defined cognitive load beyond that captured by simple summary measures. The broader conceptual rationale for state inference and subject-independent evaluation is developed in Chapter 2 (Section 2.3.6). Several practical considerations shaped the study-specific framework. The cohort included both healthy controls and people with chronic pain, allowing assessment of whether task-related ocular information remained discriminable in the presence of clinical heterogeneity. Importantly, although continuous eye tracking generated a large number of observations, these observations were nested within a relatively modest number of independent participants. The effective number of independent training exemplars was therefore determined primarily by participant number rather than by the number of timepoints or derived sequences. This informed the choice of algorithms, feature handling and participant-level validation procedures described in the following sections. The intended translational endpoint was generalisation to

previously unseen users rather than optimisation of within-sample fit, which therefore shaped the evaluation strategy described in Section 4.3.3.

4.3.2 Model classes

The modelling framework combined a conventional summary-feature baseline with multiple time-series machine learning approaches drawn from the AEON toolkit[211]. The aim was not to identify a single best-performing algorithm but to compare model families that differ in representational assumptions, computational burden and translational practicality.

A logistic regression baseline using task-level summary features (principally mean pupil diameter and blink rate) provided an interpretable benchmark against which the value of time-resolved modelling could be judged. Summary features were winsorised and z-scored prior to cross-validation.

Time-series classification used the AEON toolkit, with model inputs configured in AEON .ts format to preserve temporal ordering. Channel configurations included single-pupil, dual-pupil and pupil-with-blink-mask combinations. The final framework emphasised pupil-centred inputs because gaze contributed little additional discriminative value under the constrained visual design of the task.

The Random Convolutional Kernel Transform (ROCKET)[212] was treated as the primary time-series classifier. ROCKET convolves each series with 10,000 kernels whose length, weights, bias, dilation and padding are sampled at random, and extracts two features from each resulting feature map, the maximum value and the proportion of positive values, giving 20,000 features per case. In multivariate inputs each kernel is applied to a randomly selected subset of channels. Because the kernels are generated at random and are not fitted to the data, the transform has no learned parameters. The features are scaled and passed to a ridge classifier, in which an L2 penalty shrinks coefficients towards zero, with the penalty strength selected from ten logarithmically spaced values between 10^{-3} and 10^{3} by efficient leave-one-out cross-validation within the training data. ROCKET captures local temporal shape without manual feature engineering, is computationally efficient relative to larger ensembles and is well suited to settings in which discriminative motifs are not known a priori.

The Diverse Representation Canonical Interval Forest (DrCIF)[270] and QUANT[271] were included as interval-based classifiers. DrCIF builds an ensemble of multiple decision trees. For each tree, intervals of random position and length are sampled from three representations of the series, the raw series, its first-order differences and its periodogram, and ten attributes are drawn at random from a pool comprising the catch22 features and seven summary statistics. Each tree is therefore trained on a small, randomly constructed feature space, and predictions are averaged across trees. QUANT computes quantiles over a fixed hierarchy of dyadic intervals from the raw series, its first and second differences and its Fourier transform, and classifies the resulting features with an extremely randomised ensemble of decision trees in which each split considers a random 10% of features. QUANT offered a computationally efficient option with rapid training and prediction times.

Rotation Forest[272] was included as a strong classifier for high-dimensional vector inputs that, unlike the time-series classifiers above, does not explicitly model temporal ordering. It constructs an ensemble of decision trees and, for each tree, features are divided into random subsets and principal component analysis is applied separately to bootstrap samples of these subsets to generate a rotated feature space. All principal components are retained, so this transformation does not reduce dimensionality; rather, it decorrelates and rotates the input features while introducing diversity between trees. The rotated representations are then used to train the individual decision trees whose predictions form the ensemble.

HIVE-COTE 2.0[273] was evaluated as a high-capacity hierarchical ensemble with state-of-the-art performance across a wide range of time-series classification problems. It combines four components, the Shapelet Transform Classifier (STC), the Temporal Dictionary Ensemble (TDE), DrCIF and ARSENAL, an ensemble of smaller ROCKET classifiers that provides probability estimates. Each component's probabilities are weighted by its training accuracy raised to the fourth power, estimated internally on the training data, so that more reliable components dominate the combined prediction. It was included as a methodological reference point for the trade-off between predictive performance and deployability.

For exploratory regression of continuous subjective workload measures (NASA-TLX), ROCKETRegressor was used[274]. This algorithm applies the ROCKET random convolutional-kernel transform to time-series inputs and fits a RidgeCV regression

model to the transformed features. Workload regression was treated as exploratory rather than as a primary inferential endpoint, given that subjective ratings reflect appraisal and response style as well as task demand and are therefore conceptually distinct from task-state classification.

The model-class framework was hierarchical in its inferential weight. Logistic regression provided the conventional interpretable baseline, ROCKET was the primary time-series classifier, larger ensembles and alternative classifier families functioned as comparators, and continuous workload regression was exploratory. This hierarchy reflected the question of whether eye tracking data captured from a consumer VR system contains sufficient task-relevant temporal structure to support cognitive load inference, while also considering which classes of model are most plausible for future adaptive VR implementation.

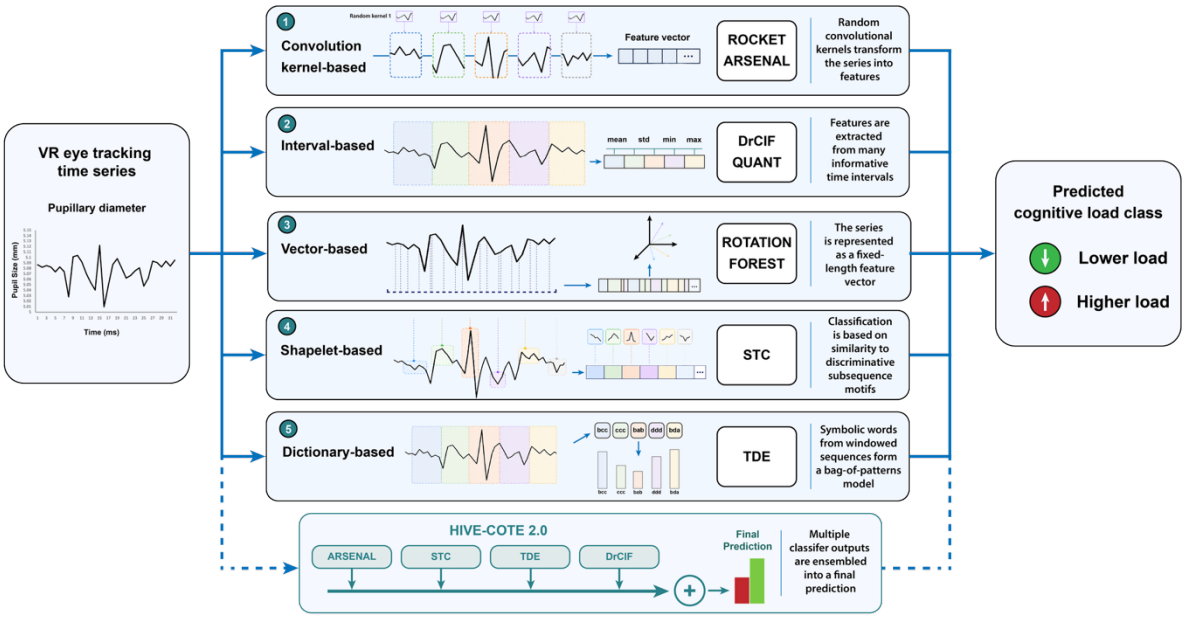


Figure 4.2. AEON time-series classifier families evaluated for VR eye tracking cognitive load classification

Conceptual schematic showing how different time-series classifier families transform pupil-diameter signals into representations for lower-versus-higher cognitive-load classification. ROCKET, the RandOm Convolutional KErnel Transform, extracts random convolutional-kernel features that are typically classified using a linear ridge classifier. DrCIF and QUANT extract interval-based features before classification using tree-based ensemble methods. Rotation Forest is an ensemble classifier applied to fixed-length transformed feature representations. STC identifies discriminative shapelet-based features before classification, while TDE represents time series as symbolic pattern histograms within a dictionary-based ensemble. HIVE-COTE 2.0 is a heterogeneous meta-ensemble that combines predictions from multiple time-series classifier families into a single final classification. The schematic is illustrative rather than a direct representation of every implementation detail in the analytical pipeline.

4.3.3 Evaluation

The primary evaluation strategy was grouped leave-one-subject-out (LOSO) cross-validation, in which all data from a given participant were held out together in the test fold and did not appear in the corresponding training set. This approach approximates the intended deployment problem of inferring task state in a previously unseen user, and prevents the optimistic bias that would arise if repeated observations from the same individual were split across training and testing folds.

Subject-level separation was particularly important because each participant contributed multiple correlated observations, including repeated task blocks derived from the same recording session. Random resampling at trial or segment level would have allowed participant-specific structure, including baseline pupil characteristics, blink tendencies and other recording-specific features, to leak across folds.

Where computational constraints made full LOSO evaluation impractical, particularly for HIVE-COTE 2.0, contracted fold times were used to ensure feasibility. Outputs from contracted models were interpreted as comparative evidence regarding model behaviour and feasibility rather than as directly comparable substitutes for full LOSO results. Model computational tractability was also evaluated, on the basis that models requiring substantially greater runtime or specialist hardware are less attractive for future adaptive VR implementation.

Performance metrics were derived from pooled out-of-fold predictions and included accuracy with Wilson 95% confidence intervals, and ROC AUC for classifiers producing probabilistic outputs. Sensitivity, specificity and confusion matrices were reported alongside accuracy. Standard deviations across folds were not reported because fold-level accuracies are highly discretised given small held-out test sets per fold, which inflates variance estimates and reduces their value for assessing model stability.

Cross-validation was applied after task and participant filtering and, where relevant, after selection of the N-back contrast of interest. Principal classification analyses focused on the binary contrast between 1-Back and 4-Back because this provided a clear discrimination target between lower and higher working-memory demand while preserving a stable class structure. Broader multiclass formulations were not treated as the primary evaluation endpoint.

4.3.4 Dimensionality, over-fitting and validation

The choice of input representations, model classes and evaluation strategy reflected the need to preserve task-relevant temporal information while accounting for a limited number of independent training examples. Continuously sampled recordings can contain thousands of measurements within each labelled sequence, but these measurements describe the sequence rather than constitute separate training examples. Repeated sequences from the same participant may also share physiological and recording-specific characteristics. The information available for generalisation was therefore constrained by participant number, despite the volume of recorded data. This was particularly relevant when using high-dimensional representations, where flexible models may fit noise or sample-specific patterns in the training data. Such over-fitting can produce strong training performance while generalising poorly to previously unseen participants, with predictive performance estimates remaining uncertain when the independent sample is small[275, 276].

Input complexity was limited by emphasising signals with a physiological and practical rationale, as described in Section 4.2.3. Continuous sequences were retained because the timing, magnitude and consistency of responses could contain information not represented by task-level summaries. No separate principal component analysis or automated feature-selection stage was applied before classification. Retaining the sequences allowed the classifiers to derive their own representations of temporal structure without first imposing a common reduced feature space. Dimensionality reduction and regularisation were distinguished within this approach: dimensionality reduction decreases the number of input features, whereas regularisation constrains fitted coefficients while retaining the feature representation.

The model classes described in Section 4.3.2 provided different ways of extracting information from these sequences and managing the resulting feature space. ROCKET combined a randomly generated convolutional representation with a regularised linear classifier, allowing temporal patterns to be represented without learning the convolutional kernels from the limited labelled examples. Interval-based methods used features calculated over sub-intervals, with randomised feature use and averaging across trees introducing diversity between learners. These approaches supported evaluation of continuous signals while providing safeguards

against unstable fitting. Higher-capacity ensembles provided comparators for assessing whether additional model complexity offered sufficient predictive benefit to justify its computational burden. The established AEON implementations supported consistent and reproducible comparisons. This balance between predictive performance and computational tractability was relevant to the future closed-loop infrastructure, in which state estimates would need to inform adaptation using the processing resources available on consumer hardware.

The grouped LOSO strategy described in Section 4.3.3 addressed the corresponding requirement for generalisation to previously unseen users. Grouped k-fold cross-validation also maintains participant separation, but LOSO retains all other participants for training at each iteration and evaluates every participant in turn. It was therefore preferred when independent training examples were limited. A single train-test split would reserve a larger proportion of participants from training and produce an estimate dependent on one allocation, while record-wise resampling would not preserve the separation needed for the intended prediction requirements[276, 277].

Internal selection of regularisation parameters or ensemble weights formed part of model fitting and was distinct from the pooled out-of-fold predictions used to estimate performance. Any additional tuning would require a nested evaluation procedure or an untouched independent test dataset, and preprocessing learned across participants would similarly need to remain within the training folds[278-280]. Regularisation and participant separation do not remove uncertainty arising from a small development sample. Independent validation and assessment of inference latency and adaptation logic would therefore be required before the models could support closed-loop therapeutic use[281]. Study-specific implementation and limitations are reported in Sections 7.3.13 and 7.5.

4.4 Reproducibility and data management

Reproducibility and data management were treated as part of the analytical framework rather than as administrative considerations. The thesis combined multiple modalities, heterogeneous acquisition contexts, automated preprocessing pipelines and several stages of downstream analysis. Reliable inference depended on the ability to trace data from raw acquisition to final derived outputs.

For the EEG studies, data standardisation used the BIDS framework[238], providing consistent naming, metadata, channel information and recording descriptors. BIDS conversion preceded all preprocessing and supported re-execution of the analysis pipeline. Automated preprocessing through the adapted DISCOVER-EEG framework reduced operator-dependent variability and allowed the same workflow to be rerun systematically across datasets and quality-control tiers. Audit logs recording channel rejection, retained data duration, proportion of rejected data and final epoch counts formed part of the inferential framework, supporting the tiered quality-control sensitivity analyses.

For the eye tracking and VR work, the Unity application generated task-linked outputs in a deterministic structured form, including raw .ocumen files (containing ocular data), behavioural task logs, questionnaire responses, calibration metrics and luminance records. These were written to the headset's persistent data path and exported for staged downstream processing. Raw outputs were preserved in their original form, with conversion to task-level CSV, participant-level workbooks and AEON-compatible time-series representations performed through versioned Python scripts rather than through manual handling.

Temporal alignment across data streams was managed through millisecond-resolution timing referenced to the behavioural tracker start (using Unity's System.Diagnostics.Stopwatch), synchronised to the onset of eye tracking recording. System-clock-based recording was avoided because of insufficient precision and frame-drop susceptibility.

The machine learning workflow imposed additional reproducibility requirements because time-series classification is particularly vulnerable to inconsistent input formatting, leakage and preprocessing variation. All time-series data were transformed into standardised AEON .ts representations with explicit metadata regarding channel structure, task labels, equal-length status and missing-data handling. Preprocessing operations including winsorisation, z-scoring, and fold-wise transformations were implemented through versioned scripts, preventing undocumented differences in data preparation from impacting the results.

The overarching principles were preservation of raw source data, generation of derivative datasets in parallel rather than through destructive overwriting, script-

based deterministic processing, and structured auditability of preprocessing and analytic steps. These principles were applied differently across EEG and eye tracking workflows because the modalities imposed different technical demands, but the underlying logic was consistent across the thesis.

5. Study 1: The PainWaves Study: Resting state EEG alterations in twin pairs discordant for chronic widespread pain

This chapter presents the PainWaves study, a co-twin control investigation of resting-state electroencephalography (rsEEG) in monozygotic (MZ) and dizygotic (DZ) twin pairs discordant for chronic widespread pain (CWP). Within-pair spectral contrasts are examined alongside periodic-aperiodic decomposition to clarify the oscillatory or broadband nature of observed effects, with exploratory extension to alpha peak properties, functional connectivity, microstate dynamics and symptom-EEG associations. The study contributes the trait-level neurophysiological component of the empirical work, framed within a design that accounts for shared genetic and environmental factors that might otherwise contribute to underlying EEG differences between individuals.

The principal finding was reduced frontocentral beta power in twins with CWP, the most reproducible spectral effect. It persisted under stricter quality-control criteria, was not attributable to a shift in the aperiodic background, and a beta-range cluster remained present even in monozygotic (MZ) pairs, in whom genetic background is fully shared. Alpha differences were weaker and less consistent and are interpreted as a secondary signal, while no convincing delta, theta or gamma signals emerged. Exploratory connectivity, microstate and symptom-association analyses showed no effects surviving correction.

5.1 Introduction

Chronic widespread pain (CWP) is a prevalent phenotype with substantial clinical impact, closely related to fibromyalgia and characterised by widespread musculoskeletal pain, fatigue, sleep disturbance, cognitive symptoms[282-284]. While precise aetiology remains unclear, many studies support its association with centrally mediated nociplastic pain mechanisms[1, 2, 17, 40]. Such mechanisms offer a tractable target for identifying neurophysiological features that would inform future immersive therapeutic development.

Resting-state electroencephalography (rsEEG) provides a non-invasive, scalable methodology to examine spontaneous neural activity in CWP. Previous studies in chronic pain and fibromyalgia have reported alterations across multiple spectral bands and in functional connectivity, but findings remain inconsistent in candidate features, effect direction and reproducibility[44, 127, 128, 147]. This likely reflects heterogeneity in clinical populations, EEG acquisition, preprocessing, regions of interest and analytical methods. It may also reflect limitations of conventional band-power analysis, which can conflate periodic oscillatory activity with shifts in the aperiodic background spectrum, an issue addressed here using periodic-aperiodic decomposition[157].

A less frequently considered source of inconsistency is the heritability of EEG itself. Core rsEEG traits, including spectral power and aspects of connectivity, show moderate to high heritability in twin and family studies[159-162]. CWP itself is influenced by familial and genetic factors[163]. Conventional case-control comparisons between unrelated participants may therefore capture shared familial liability rather than pain-associated neurophysiological variation. EEG studies that control for genetics in chronic pain remain rare, limiting confidence in candidate biomarkers identified using standard between-group designs.

The PainWaves study addresses this gap using a discordant co-twin control design within an established adult twin registry. Monozygotic (MZ) and dizygotic (DZ) twin pairs in which one twin met criteria for CWP and the other did not completed standardised rsEEG acquisition, enabling within-pair comparison while partially controlling for shared genetic and environmental background. The primary analysis examined within-pair differences in spectral power, with secondary periodic-aperiodic decomposition and alpha peak analyses, and exploratory analyses of functional connectivity, microstate dynamics and EEG-symptom associations. The objective was to identify rsEEG features of CWP that remain detectable under a design that partially addresses EEG heritability and shared environmental exposure.

5.2 Study aims

The primary aim of this study was to determine whether CWP is associated with reproducible alterations in rsEEG that are independent of shared genetic and environmental influences. To address this, the study used a co-twin control design in

twin pairs (MZ and DZ) discordant for CWP, enabling within-pair comparisons that reduce confounding from shared genetic background and environmental exposures.

A secondary aim was to characterise the nature of any observed spectral differences, specifically whether they were better explained by alterations in oscillatory (periodic) activity within canonical frequency bands, or by changes in the aperiodic background component of the EEG spectrum.

Additional exploratory aims were to examine whether alpha peak properties, functional connectivity, and microstate dynamics were associated with CWP within this genetically informed framework, and whether observed differences varied with zygosity.

5.3 Methods

5.3.1 Study design and participants

This study employed a co-twin control design using MZ and DZ twin pairs discordant for CWP recruited from the TwinsUK cohort. Participants were enrolled between February and August 2025 while already participating in an existing TwinsUK neuroimaging study. TwinsUK is an adult twin registry recruited from the general population through national media campaigns and characterised by extensive longitudinal phenotyping, including pain-related traits[285]. Zygosity has been assigned within TwinsUK using questionnaire methods[286] and confirmed by genotyping.

CWP status was determined using a standardised questionnaire-derived phenotype based on the London Fibromyalgia Epidemiology Study Screening Questionnaire (LFESSQ)[287], in line with American College of Rheumatology definitions[284] used in prior TwinsUK work[288, 289]. The CWP screening process included an initial question regarding the presence of musculoskeletal pain in the preceding 3 months, followed by a more detailed questionnaire assessing pain distribution across upper limb, lower limb, and axial regions, together with pain duration. CWP was defined as pain affecting both sides of the body, above and below the waist, involving the axial skeleton, and persisting for at least 3 months. Each twin completed this screening process independently.

For this study, twin pairs were selected using TwinsUK LFESSQ questionnaire data. Given questionnaires were completed pre-2020, affected twins were therefore treated as having a longstanding questionnaire-derived CWP phenotype, present for

at least 5 years before the present EEG analysis. Discordant pairs were defined as pairs in which one twin met questionnaire criteria for CWP and the co-twin did not. Screening, selection and consent procedures were conducted by the TwinsUK research team.

5.3.2 Ethics and governance

Existing ethical approval for the TwinsUK Programme of Research was applicable for EEG data collection, participant management and data analysis conducted at King's College London (REC 24/NW/0107; IRAS project ID 342169, Appendix 5). Secondary ethical approval from the University of East Anglia (UEA) Faculty of Medicine and Health Research Ethics Sub-committee was obtained for analysis of de-identified EEG data at UEA (REC reference ETH2425-2568, Appendix 5). A formal data-sharing agreement was established for this collaborative study between the Department of Twin Research and Genetic Epidemiology (King's College London), and the UEA prior to data transfer and analysis. All procedures followed relevant guidelines and regulations, including the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrolment.

5.3.3 EEG acquisition

EEG was acquired using the 32-channel R-Net configuration described in Chapter 4 (Section 4.1.1). For the R-Net sponge-based system, impedance values of approximately 60-100kΩ are recommended with tolerance up to 150kΩ[290], reflecting the different electrode-to-scalp contact characteristics of saline-based acquisition relative to gel-based systems. Across 72 participants, mean electrode impedance, excluding sentinel out-of-range values, was 60.7 ±20.3kΩ (median 57.5kΩ). On average, 84.9% of electrodes per participant were ≤100kΩ, 7.4% were within the 100-150kΩ range, and 7.5% exceeded 150kΩ or were labelled out-of-range. Only three participants exhibited more than 25% out-of-range channels. Poorer impedance values, defined as >150kΩ or out-of-range, were distributed across the montage rather than concentrated at a small number of systematically problematic sites. The principal frontocentral ROI electrodes used in later effect-size analyses were not among the most affected sites. Overall, these distributions were consistent with stable contact quality for saline-based R-Net acquisition.

Participants were seated comfortably in a quiet research clinic room at TwinsUK. The acquisition protocol comprised 7 minutes of eyes-closed (EC) recording followed by 5 minutes of eyes-open recording with visual fixation. At the onset of each recording condition, the researcher manually inserted an event marker into the acquisition software to denote the start of the EC and EO periods. The present analysis focused exclusively on the EC segment, consistent with the resting-state framework described in Chapter 4.

5.3.4 Self-reported measures and participant characteristics

Existing TwinsUK study records were used to extract participant demographics, CWP status, medication data and co-morbidity data. Questionnaire-derived medication and co-morbidity records were linked to the PainWaves cohort using participant identifiers and summarised descriptively for clinical context.

For co-morbidity data, the most recent relevant questionnaire record was reviewed for each participant, with earlier questionnaire entries retained where they provided additional historical diagnostic information. The 2024 longitudinal core questionnaire delivered by TwinsUK as part of their twin follow-up primarily captured self-reported doctor-diagnosed musculoskeletal and rheumatological conditions, including osteoarthritis, rheumatoid arthritis, gout, lupus, other inflammatory arthritis, osteoporosis and polymyalgia rheumatica, with information on age at first diagnosis where available. Earlier questionnaires varied in scope and included some additional non-MSK conditions.

Medication fields captured oral medications and selected topical medications judged relevant to clinical or neurophysiological interpretation. Over-the-counter supplements were not included, except for vitamin D or calcium/vitamin D preparations where recorded. Medication entries were summarised using the most recent available questionnaire record; questionnaire years were interpreted as the medication record source rather than as treatment start dates or confirmation of use at EEG acquisition.

Medication variables were grouped descriptively into clinically relevant, non-mutually exclusive categories. Analgesic medication included paracetamol, NSAIDs, opioid or opioid-combination analgesics, nefopam, lidocaine patches and methocarbamol. Neuropathic or adjuvant pain medication included gabapentinoids and tricyclic

antidepressants. A separate EEG-relevant centrally acting medication category was used to capture medication with plausible effects on arousal, cortical excitability or resting EEG rhythms, including centrally acting analgesics, gabapentinoids, antidepressants, sedatives, hypnotics, anti-seizure or mood-stabilising medications, dopaminergic medications and other centrally acting agents.

Co-morbidity and medication variables were used for descriptive clinical characterisation only and were not used for formal covariate adjustment because questionnaire timing and scope varied; medication dosage, indications and adherence data were inconsistent; diagnoses and ongoing medication use were not verified at EEG acquisition.

Shortly before EEG set-up on the day of testing, participants completed standardised self-report measures to characterise current symptom burden and relevant factors related to pain and mental health. Pain intensity and fatigue were assessed using 0-10 VAS, where 0 indicated no pain or fatigue and 10 indicated worst possible pain or fatigue. Depressive and anxiety symptoms were measured using the Patient Health Questionnaire-2 (PHQ-2)[291] and Generalised Anxiety Disorder-7 (GAD-7)[292], respectively. Widespread pain and associated symptom burden were quantified using the Widespread Pain Index (WPI) and Symptom Severity Scale (SSS), together forming the Fibromyalgia Severity (FS) score as described in the revised 2016 fibromyalgia criteria[282]. Participants also reported total sleep duration on the previous night. Appendix 2 provides further detail on the self-reported measures used.

Measures were selected to characterise the pain phenotype and participants' state at the time of recording, within the constraints of an existing multimodal TwinsUK visit in which assessment time and participant burden were limited. CWP status was defined using the LFESSQ to maintain continuity with the established TwinsUK phenotype and its previous use in genetic and epidemiological studies[288, 289]. As the questionnaires used to define CWP predated the EEG visit, the WPI, SSS and composite FS score were also collected on the day of testing to provide a contemporary, continuous measure of widespread pain and associated symptom burden, features relevant to the nociplastic pain phenotype[282, 293]. Pain and fatigue were rated using a VAS immediately before EEG recording because both fluctuate over short timescales and may influence resting measures; single-item visual

analogue ratings also provide a brief and practical means of quantifying current symptom severity[294, 295]. Depressive and anxiety symptoms were assessed using the PHQ-2 and GAD-7 because both are commonly associated with CWP and may influence symptom reporting and rsEEG measures. The PHQ-2 was selected in preference to the PHQ-9 to reduce participant burden while retaining acceptable performance as a brief depression screening measure[291]. Previous-night sleep duration was recorded because acute sleep restriction can alter waking resting-state spectral activity[296]. Collectively, these measures captured several clinically relevant domains recommended for chronic pain research, including pain intensity, emotional functioning and broader symptom burden[297]. Pain interference and physical function were not assessed, however, limiting interpretation of the functional impact of pain within this cohort.

Missing values in core study-visit self-report variables were imputed using status-stratified median imputation, whereby missing values were replaced with the median value calculated from participants with the same CWP status. Imputation was restricted to core study-visit self-report variables and was not applied to EEG-derived measures, medication or co-morbidity variables.

5.3.5 Data preparation and preprocessing

Raw EEG recordings were converted to BIDS format and preprocessed using the automated pipeline described in Chapter 4 (Section 4.1.2). To preserve the co-twin pairing structure throughout the workflow, participant identifiers were parsed deterministically from source filenames and linked to family-level BIDS subject labels. Across the dataset, 72 EC recordings from 36 twin pairs were successfully converted. EC onset markers were available for 66/72 recordings (91.7%); where no marker was present, a deterministic fallback start time of 10s after recording onset was applied.

To support within-pair comparability while maximising retention, each recording was trimmed to a fixed maximum processing window anchored to EC marker onset where possible. Up to 390s of data were retained per participant. Where insufficient data remained because of shorter recordings, the maximum available duration was retained without excluding the participant (n=8, range 304.4-375.6s). Electrode coordinate metadata were incorporated during BIDS conversion using the Brain

Products RNP-LA-32.bvef template montage, allowing standardised electrode positions to be written into BIDS sidecars.

5.3.6 Processed data quality

As expected for a pragmatically acquired passive-electrode dataset collected within a broader non-EEG specific research pathway, data quality was more variable than would typically be anticipated in a tightly controlled high-density gel-based laboratory acquisition. Nevertheless, preprocessing audit outputs indicated that the retained recordings remained suitable for downstream spectral analyses, with visual review of power spectral density plots confirming preservation of plausible brain-related EEG spectral characteristics. Across 72 recordings, a mean of 3.9 ±2.7 channels per participant were removed before interpolation (median 3, range 0-13). The number of retained clean epochs was 93.1 ±27.7 on average (median 96, range 30-145), corresponding to a mean retained post-clean duration of 186.23 ±55.5s (median 192s, range 60 to 290s). These audit outputs were used to characterise retained data quality across participants and to support later sensitivity analyses.

5.3.7 Sample size and power

The co-twin control design provides efficiency gains relative to unmatched designs when the outcome exhibits substantial within-pair similarity. rsEEG traits, particularly alpha and beta power, showed high MZ twin correlation ($r \approx 0.8$-0.9) and moderate DZ correlation ($r \approx 0.4$-0.5). Given the present sample composition (17 MZ and 19 DZ discordant pairs), the expected overall within-pair correlation for beta power was approximately 0.60-0.65, corresponding to a relative efficiency of approximately 2.5-2.9 compared with an unmatched design.

With 36 discordant twin pairs, the study had 80% power to detect a moderate within-pair effect size (Cohen's $dz=0.48$) using a two-tailed paired t-test ($\alpha=0.05$). Under stricter quality-control tiers, detectable effect sizes increased as expected due to reduced sample size. For Tier 1 ($n=30$ pairs), 80% power corresponds to $dz = 0.53$, for Tier 2 ($n=22$ pairs), $dz = 0.63$, and for Tier 3 ($n=12$ pairs), $dz = 0.89$.

5.3.8 Inferential framework

All inferential analyses preserved the matched co-twin structure of the study design, with primary comparisons based on within-pair differences between the CWP-

affected twin and the unaffected co-twin rather than on unmatched group comparisons. This is an approach widely used in co-twin control studies[165].

5.3.9 Sensitivity analyses

To assess robustness to residual data-quality variation, relevant analyses were repeated across progressively stricter quality-control tiers, excluding a pair if either twin exceeded predefined thresholds for channel interpolation burden or retained post-clean duration.

Tier 0 included all discordant twin pairs (n=36). Tier 1 was defined as a minimal sensitivity tier, excluding recordings with ≥10 channels removed or <75s retained data (n=30 pairs), in order to remove outliers characterised by substantial channel loss or limited retained data while preserving a broad analysis sample for spectral outcomes. Tier 2 applied stricter criteria, excluding recordings with ≥8 channels removed or <120s retained data (n=22 pairs), and was used for more conservative spectral analyses together with secondary aperiodic and microstate sensitivity analyses. These parameters were chosen to reflect evidence that aperiodic estimation and microstate metrics benefit from longer resting segments than basic spectral power analysis, while remaining pragmatic in relation to the available sample size[298-300]. Tier 3 was reserved for exploratory connectivity analyses only and applied a higher retained-duration threshold (≥8 channels removed or <180s retained data; n=12 pairs), reflecting the greater data-length demands of resting-state connectivity estimation relative to spectral power analyses[301].

To examine potential differences by zygosity, exploratory stratified analyses were conducted separately within MZ and DZ discordant pairs. Within-pair contrasts were preserved in all strata, and analytical code was identical across tiers and zygosity groups, varying only the eligible pair manifest. These stratified analyses were considered exploratory and were interpreted cautiously.

5.3.10 Primary spectral analysis

PSD and ROI-based effect size estimation were completed in line with the methodology described in Chapter 4 (Section 4.1.4). Statistical inference for the primary spectral analysis was implemented in MATLAB R2025b[235] using FieldTrip[237] and custom analysis scripts. The analysis tested within-pair differences in band-limited sensor-level power between the twin with chronic widespread pain and the

co-twin without chronic pain. Differences were coded as CWP1-CWP0, where positive values indicated higher power in the twin with CWP.

Sensor-level inference used cluster-based permutation testing. The design matrix specified the twin pair as the unit variable and pain status as the within-pair independent variable, thereby preserving the matched-pair structure throughout inference. Paired comparisons used a dependent-samples t-statistic. Spatially adjacent sensors were defined using triangulation of the recorded 32-channel R-Net electrode geometry, and clusters were formed using a cluster-forming threshold of $p<0.05$. Cluster-level statistics used the sum of t-values within each spatial cluster, and statistical significance was evaluated against a null distribution generated using 5,000 two-sided random permutations of the paired observations. Cluster-level p-values were not subjected to additional correction, because the cluster-based permutation procedure controls the family-wise error rate at the cluster level.

Paired frontocentral (Fz, FC1, FC2, Cz) ROI-based effect sizes were expressed as Cohen's dz, calculated as the mean within-pair difference divided by the standard deviation of the within-pair difference. Bootstrap resampling of complete twin pairs with 5,000 iterations was used to derive 95% confidence intervals. ROI-based estimates were used to summarise the direction and magnitude of effects identified in the sensor-level analyses and were interpreted descriptively alongside the primary cluster-based inference.

5.3.11 Secondary aperiodic-periodic spectral parameterisation analysis

Aperiodic-periodic decomposition followed the FOOOF/specparam framework described in Chapter 4 (Section 4.1.5) and was implemented using a custom MATLAB pipeline integrating EEGLAB[236], FieldTrip[237] and Python's FOOOF package called from within MATLAB.

The analysis was conducted on the Tier 2 dataset using the predefined frontocentral ROI (Fz, FC1, FC2, Cz). An ROI-based approach was used because this analysis was intended as a secondary spectral decomposition of predefined regional activity, rather than as a separate whole-scalp spatial discovery analysis. The frontocentral ROI provided a biologically plausible and consistent region for parameterisation, reduced the multiplicity and model-fit complexity associated with channel-wise FOOOF fitting, and improved the stability of spectral estimates by averaging spectra

across neighbouring electrodes. For each participant, ROI-averaged power spectra were derived from the preprocessed 2s eyes-closed epochs and parameterised over 2-40 Hz using a fixed aperiodic model. Periodic peaks were modelled as Gaussian functions with peak width limits of 1-8Hz, a maximum of six peaks, a minimum peak height of 0.1 and a peak threshold of 2.0. Offset and exponent parameters were extracted from the fitted aperiodic component, and residual aperiodic-adjusted band power was calculated within predefined frequency bands after subtraction of the fitted background component in log10 power space.

Within-pair differences were evaluated through paired t-tests for each FOOOF/specparam-derived outcome, with effect sizes expressed as Cohen's dz and 95% bootstrap confidence intervals based on 5,000 iterations. Holm-Bonferroni correction was applied separately across the predefined aperiodic outcome family (offset, exponent) and the residual band-power family (residual delta, theta, alpha, broad beta, low beta, high beta and gamma).

5.3.12 Secondary alpha peak analysis

Individual alpha peak frequency (IAF) and alpha peak power were estimated using the posterior (O1, O2, Oz, POz, Pz, PO3, PO4) and central (C3, Cz, C4) ROIs. Analyses were implemented using custom MATLAB and Python scripts applied to the preprocessed eyes-closed spectra. Within-pair differences were evaluated using paired t-tests.

Effect sizes were expressed as Cohen’s dz with 95% bootstrap confidence intervals based on 5,000 resamples of complete twin pairs. Holm-Bonferroni correction was applied across the alpha peak outcome family to account for testing of alpha peak frequency and alpha peak power across the predefined ROIs.

5.3.13 Exploratory connectivity analysis

Exploratory resting-state functional connectivity was implemented in MATLAB R2025b[235] using EEGLAB[236], FieldTrip[237] and custom analysis scripts. Connectivity was assessed using amplitude envelope correlation (AEC), an amplitude-based measure that quantifies the temporal co-fluctuation of band-limited oscillatory envelopes and has prior use in chronic pain rsEEG biomarker work[128]. The analysis was conducted on the Tier 3 dataset and was restricted to frequency bands implicated by the primary spectral findings.

To reduce volume conduction, data were transformed to scalp current source density (CSD) before connectivity estimation[302]. Band-limited signals were obtained by finite impulse response (FIR) filtering, analytic envelopes were extracted using the Hilbert transform, and channel-level Pearson correlations were calculated on concatenated envelopes across epochs. Correlation coefficients were Fisher z-transformed prior to ROI aggregation.

Connectivity estimates were summarised within and between four predefined sensor groupings: frontal (Fp1, Fp2, F3, F4, F7, F8, Fz), frontocentral-central (FC1, FC2, C3, C4, Cz), parietal (CP1, CP2, P3, P4, Pz) and occipital (O1, Oz, O2). For within-region estimates, diagonal elements and duplicate edges were excluded so that summary values reflected unique inter-sensor connections only.

Within-pair differences were analysed separately for each region-to-region connection using two-sided paired sign-flip permutation tests with 5,000 permutations. Effect sizes were expressed as Cohen's dz with 95% bootstrap confidence intervals from 5,000 resamples of complete twin pairs. Within each frequency band, p-values were adjusted across tested region-to-region comparisons using the Holm-Bonferroni procedure. These analyses were interpreted as exploratory and were not weighted equivalently to the primary spectral or aperiodic-periodic analyses.

5.3.14 Exploratory microstate analysis

Microstate analysis was implemented in MATLAB R2025b[235] using EEGLAB[236] and custom scripts. The analysis was performed to characterise the temporal organisation of quasi-stable scalp topographies over short time scales, with key parameters aligned with prior microstate research[154, 298]. It was conducted on the Tier 2 dataset using a microstate-specific 1-40Hz filtered copy of the preprocessed eyes-closed data.

Global field power (GFP) was calculated at each time point as the spatial standard deviation of voltage across electrodes, providing a reference-independent measure of instantaneous scalp field strength. Local GFP maxima were used for map sampling because these time points have higher signal-to-noise and are commonly treated as the most stable moments for estimating scalp topographies. Candidate maps were sampled at local GFP maxima using a minimum inter-peak distance of

10ms and a maximum of 4,000 peaks per participant. The scalp maps sampled at these GFP peaks were then submitted to a two-stage k-means clustering procedure using correlation distance. In this context, k-means clusters scalp topographies according to spatial similarity, with each cluster representing a recurring microstate map. Candidate solutions from K=4 to K=8 were examined. K=5 was selected as the primary descriptive solution because it offered an extension of the traditional four-class resting-state model[303] while better fitting the present data. Relative to K=4, the five-class solution increased the mean total global explained variance and reduced the proportion of unassigned time points, and the additional fifth class itself accounted for a meaningful share of coverage and explained variance rather than a negligible one. This indicated that the fifth class captured genuine structure in the data, supporting K=5 as an interpretable intermediate solution that improved on K=4 without the greater fragmentation risk of higher-class models. The neighbouring K solutions were examined descriptively to confirm that the interpretation did not depend on the chosen class number.

The resulting group-level microstate templates were then backfitted to each participant's continuous data. Backfitting refers to assigning each time point in the EEG recording to the template map with the highest spatial similarity. Assignment was performed using polarity-invariant absolute spatial similarity, meaning that maps with the same spatial configuration but opposite polarity were treated as equivalent. A minimum similarity threshold of 0.50 was applied. Segments shorter than 30ms were removed using minimum-duration temporal smoothing to reduce implausibly brief class switching. For each microstate class, mean duration, temporal coverage, occurrence rate and GFP-weighted global explained variance (GEV) were computed. Mean duration quantified the average length of continuous segments assigned to a class, temporal coverage quantified the proportion of recording time assigned to that class, occurrence rate quantified how often the class appeared, and GEV quantified how much of the GFP-weighted scalp variance was accounted for by the fitted templates.

Within-pair differences were evaluated using paired t-tests, with Holm-Bonferroni correction applied across tested microstate outcomes within each K solution. As with the connectivity analysis, this work was treated as exploratory rather than as a confirmatory test of the primary spectral findings.

5.3.15 Clinical descriptive statistics and missing-data handling

Data normality was assessed using Shapiro-Wilk tests. All clinical measures demonstrated significant deviation from normality ($p<0.05$). Accordingly, medians and IQRs with non-parametric summary statistics are reported. Missing data were imputed using group-specific median substitution, whereby missing values were replaced with the median of the observed values within the corresponding group (CWP case or co-twin control), preserving between-group distributional differences and minimising distortion of within-pair contrasts.

5.3.16 Exploratory symptom-association analyses

Univariate ordinary least squares regression analyses were conducted in Python (v3.12) to examine association between within-pair difference in EEG power and within-pair difference in clinical characteristics. EEG outcomes were defined as CWP1 minus CWP0 band-limited power averaged within prespecified sensor-level regions of interest. Analyses were restricted to bands implicated by the primary spectral results: alpha (8-12Hz), broad beta (13-30Hz), low beta, and high beta (21-30Hz). A primary frontocentral ROI (Fz, FC1, FC2, Cz) was specified a priori to reflect the principal spatial locus of the spectral effects. A secondary parietal ROI (CP1, CP2, P3, P4, Pz) was examined as a limited secondary region to assess whether clinical associations extended beyond the frontocentral topography evidence in spectral analyses.

Five clinical predictors were examined: pre-EEG pain intensity (VAS), pre-EEG fatigue intensity (VAS), Fibromyalgia Severity (FS) score, depressive symptoms (PHQ-2), and anxiety symptoms (GAD-7). Each predictor was entered separately because these symptom measures were moderately intercorrelated and the sample size did not support stable multivariable modelling. Continuous predictors were standardised (z-scored) to facilitate comparison of regression coefficients across models.

The exploratory zygosity analysis was performed in the full cohort using interaction models (EEG outcome ~ predictor + zygosity + predictor × zygosity), to test whether association differed between MZ and DZ pairs. To aid interpretation of any interaction effects, parallel subgroup models were also estimated separately within MZ and DZ twins. All models were estimated with heteroskedasticity-robust standard errors

(HC3), which provide more reliable variance estimates in small samples and under heteroskedasticity. Regression coefficients, 95% confidence intervals, R^2 values, and associated p-values were reported, and model assumptions were assessed using residual and influence diagnostics. Holm-Bonferroni correction was applied across the five clinical predictors within each band-by-ROI analysis.

5.4 Results

A total of 36 discordant twin pairs (total n=72) were included with 17 MZ and 19 DZ pairs. Table 5.1 summarises participant characteristics for twins with CWP and their co-twin controls.

For the core study-visit self-reported measures included in Table 5.1, missing data frequency was low, affecting only three participants (4%), with 46 missing values across 3024 cohort-wide observations (2%) requiring imputation. Across participants, those with CWP had mild symptom burden, but reported higher pain, fatigue and anxiety ratings than participants without CWP at group level.

Questionnaire-derived co-morbidity and medication data were available for most participants and were summarised descriptively. A dated relevant co-morbidity questionnaire was available for 70/72 participants, comprising equal numbers of twins with and without CWP. Most participants had a 2024 questionnaire record (50/72), while 20 participants had only earlier questionnaire data. Co-morbidities were predominantly musculoskeletal or rheumatological. Osteoarthritis, inflammatory arthritis or connective tissue disease, and gout were more frequent in twins with CWP, whereas osteoporosis was balanced between groups. Analgesic medication and opioid or opioid-combination analgesic use were also more frequent in twins with CWP. In contrast, neuropathic or adjuvant pain medication use was similar between groups, while EEG-relevant centrally acting medication exposure was modestly more frequent in twins with CWP.

Table 5.1. Participant characteristics for the PainWaves study

Characteristic	Twins with CWP (n=36)	Twins without CWP (n=36)
Age in years, mean (SD)	69.9 (8.1)	69.9 (8.1)

Female sex, n (%)	34 (94%)	33 (92%)
Ethnicity, n (%)		
White British	34 (94%)	34 (94%)
Asian British	1 (3%)	1 (3%)
Other (unspecified)	1 (3%)	1 (3%)
Widespread Pain Index (WPI) score, median (IQR)	3.0 (2.0-6.3)	2.0 (1.0-3.0)
Symptom Severity Scale (SSS) score, median (IQR)	3.0 (2.0-5.0)	2.0 (1.0-4.0)
Fibromyalgia Severity (FS) score, median (IQR)	7.0 (3.8-10.0)	3.5 (2.0-7.8)
Pain VAS score, median (IQR)	1.0 (0.0-3.0)	0.5 (0.0-2.3)
Fatigue VAS score, median (IQR)	2.0 (0.0-4.0)	1.0 (0.0-3.5)
PHQ-2 score, median (IQR)	0.0 (0.0-2.0)	0.0 (0.0-1.5)
GAD-7 score, median (IQR)	2.0 (0.3-4.0)	1.0 (0.0-5.0)
Quantity of sleep on the previous night in hours, median (IQR)	5.0 (4.0-6.0)	5.0 (4.0-6.0)
Questionnaire-derived co-morbidities, n/N (%)		
Osteoarthritis	23/35 (66%)	16/35 (46%)
Osteoporosis	7/35 (20%)	7/35 (20%)
Inflammatory arthritis or connective tissue disease	4/35 (11%)	0/35 (0%)
Gout	2/35 (6%)	0/35 (0%)
Questionnaire-derived medication, n/N (%)		
Analgesic medication	19/34 (56%)	7/35 (20%)
Opioid or opioid-combination analgesic	9/34 (26%)	3/35 (9%)

Neuropathic or adjuvant pain medication	4/34 (12%)	5/35 (14%)
EEG-relevant centrally acting medication	14/34 (41%)	10/35 (29%)

Values are presented as mean (SD), median (IQR), n (%) or n/N (%), as appropriate. CWP, chronic widespread pain; CTD, connective tissue disease; EEG, electroencephalography; FS, Fibromyalgia Severity; GAD-7, Generalised Anxiety Disorder-7; PHQ-2, Patient Health Questionnaire-2; SSS, Symptom Severity Scale; VAS, visual analogue scale; WPI, Widespread Pain Index. FS score was calculated as WPI + SSS. Questionnaire-derived co-morbidity and medication variables were summarised descriptively; denominators vary by data availability. Medication categories are non-mutually exclusive and defined in Section 5.3.4.

Across 36 complete discordant pairs within the primary analysis, paired cluster-based permutation testing demonstrated a robust reduction in beta band power (13-30Hz) in pain-affected twins relative to their unaffected co-twins, expressed as a broad frontocentral negative cluster (cluster-level p=0.005) with a moderate ROI-based effect size (dz -0.57, 95% CI -0.986 to -0.260). Evaluation of beta sub-bands (low beta 13-20Hz; high beta 21-30Hz) showed comparable cluster significance with overlapping frontocentral distributions and similar ROI-based effect sizes. Significant beta and sub-band clusters with comparable topography and larger ROI-based effect sizes were seen under stricter quality-control tiers on sensitivity analyses. Alpha (8-12Hz) and gamma (31-40Hz) also showed significant negative clusters on primary analysis (cluster-level p-values: alpha p=0.044; gamma p=0.026), with small-moderate ROI-based effect sizes (alpha: dz -0.38, 95% CI -0.737 to -0.050, gamma: dz -0.46, 95% CI -0.805 to -0.152). On sensitivity analyses, clusters remained statistically significant across Tiers 1-2. Delta and theta bands showed no significant clusters with primary or sensitivity analyses. See Table 5.2 for full band-specific dz and p-values across sensitivity tiers.

When analyses were stratified by zygosity, DZ twin pairs (n=19) showed significant negative clusters in the alpha, broad beta, low beta and high beta bands (cluster-level p-values 0.011-0.038). The largest ROI-based effect size was observed for low beta (dz -0.683, 95% CI -1.360 to -0.253). In MZ twin pairs (n=17), significant cluster-level effects were more selective, with a negative frontocentral high beta cluster persisting as the only significant band-level finding (cluster-level p=0.009). ROI-based effect size was large (dz -0.819, 95% CI -1.635 to -0.339).

Table 5.2. Within-pair spectral effect sizes and cluster-level p-values across primary and sensitivity quality-control tiers

Tier	n	Delta dz (p-value)	Theta dz (p-value)	Alpha dz (p-value)	Beta dz (p-value)	Low Beta dz (p-value)	High Beta dz (p-value)	Gamma dz (p-value)
Tier 0	36	-0.28 (0.114)	-0.37 (0.101)	**-0.38 (0.044)**	**-0.57 (0.005)**	**-0.55 (0.006)**	**-0.54 (0.003)**	**-0.46 (0.026)**
Tier 1	30	-0.31 (0.111)	-0.39 (0.092)	**-0.52 (0.015)**	**-0.70 (<0.001)**	**-0.63 (0.007)**	**-0.73 (0.001)**	**-0.48 (0.050)**
Tier 2	22	-0.35 (0.118)	-0.45 (0.088)	**-0.58 (0.021)**	**-0.87 (0.004)**	**-0.77 (0.006)**	**-0.92 (0.002)**	**-0.69 (0.009)**

Values show within-pair ROI-based Cohen's dz effect sizes and cluster-level p-values for rsEEG band-power differences between twins with CWP and their co-twins without CWP. n denotes the number of discordant twin pairs retained within each quality-control tier. Negative dz values indicate lower band power in the twin with CWP relative to their co-twin. Frequency bands were defined as delta 1-3Hz, theta 4-7Hz, alpha 8-12Hz, broad beta 13-30Hz, low beta 13-20Hz, high beta 21-30Hz and gamma 31-40Hz.

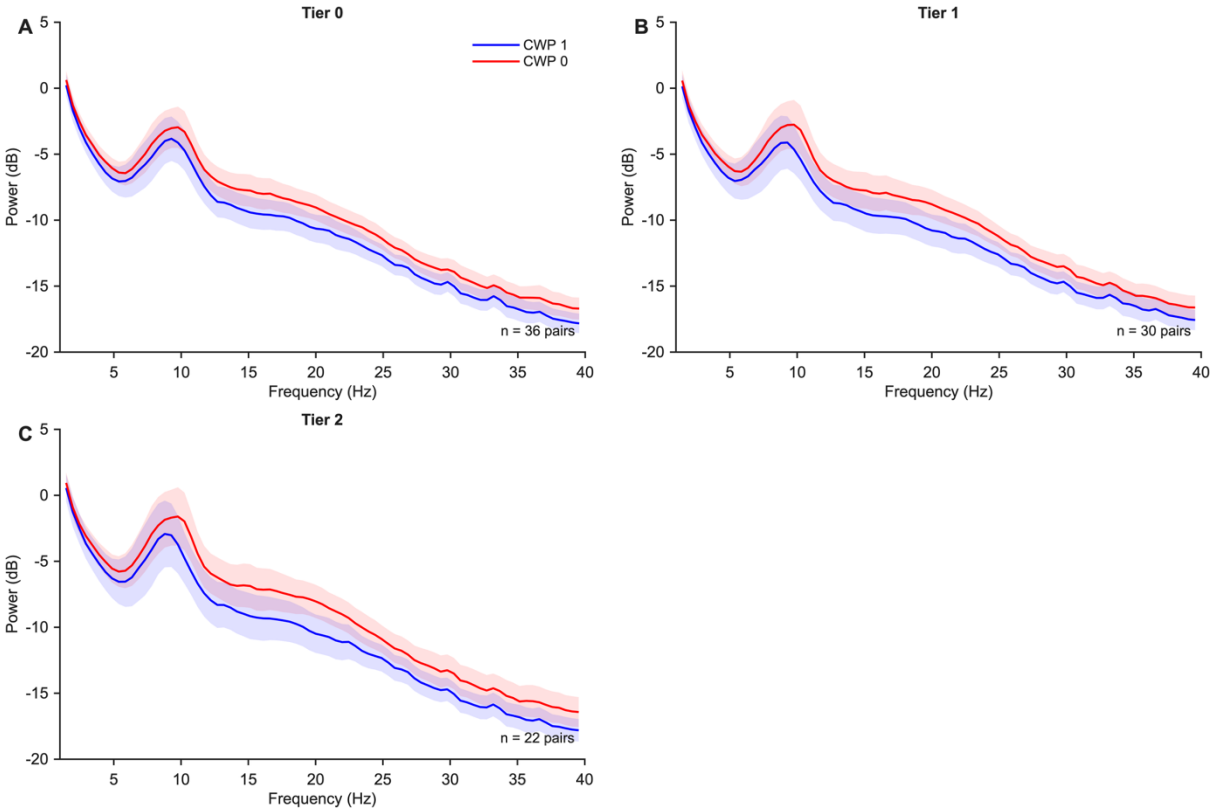


Figure 5.1. Group-level resting-state power spectral density across quality-control tiers

Mean power spectral density (PSD), averaged across all channels, for twins with CWP (CWP1, blue) and co-twins without CWP (CWP0, red). Shaded regions indicate 95% confidence intervals around the group mean. Panels show progressively stricter quality-control tiers: (A) Tier 0, all complete pairs (n=36 pairs); (B) Tier 1 (n=30 pairs); (C) Tier 2 (n=22 pairs).

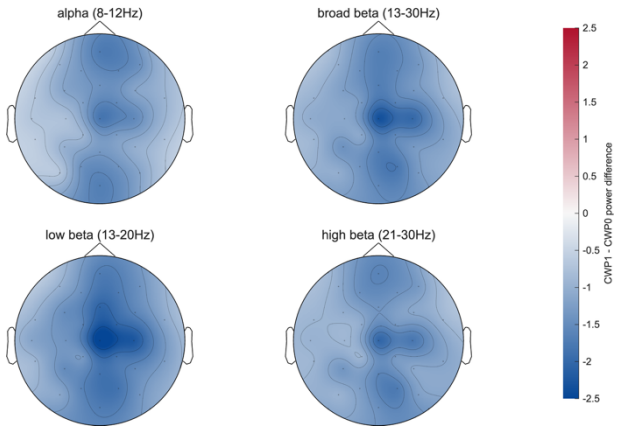


Figure 5.2. Band power topographies for alpha and beta bands

Topographies show mean within-pair differences in EEG band power across the scalp for alpha (8-12Hz), broad beta (13-30Hz), low beta (13-20Hz) and high beta (21-30Hz) for Tier 0 (n=36 pairs). Blue indicates lower band power in the co-twins with CWP (CWP1) relative to co-twins without CWP (CWP0). The maps are presented as descriptive spatial visualisations of the conventional spectral contrasts.

5.4.2 Secondary aperiodic-periodic spectral parameterisation

In the paired Tier 2 analysis (n=22 pairs) neither the aperiodic exponent nor offset differed significantly between twins with and without CWP (exponent: Holm-corrected p=0.525, dz 0.25, 95% CI -0.173 to 0.652; offset: Holm-corrected p=0.525, dz -0.23, 95% CI -0.708 to 0.193). Thus, there was no evidence for a robust between-twin difference in the broadband aperiodic spectral background.

Following aperiodic adjustment, residual broad beta showed an uncorrected within-pair reduction (p=0.024, Holm-corrected p=0.156, dz -0.52, 95% CI -1.080 to -0.114), with a similar pattern in low beta (p=0.022, Holm-corrected p=0.156, dz -0.53, 95% CI -1.097 to -0.109). Residual high beta effects were smaller and not statistically significant (uncorrected p=0.204, dz -0.28, 95% CI -0.803 to 0.123), and residual alpha likewise showed no reliable difference (uncorrected p=0.158, dz -0.31, 95% CI -0.934 to 0.144). No residual gamma difference was identified (uncorrected p=0.701, dz -0.08, 95% CI -0.410 to 0.485).

Alongside the power spectral density findings, these results are more consistent with altered oscillatory activity than with a broad shift in the aperiodic spectral background, given the absence of between-twin differences in either exponent or offset. Within the aperiodic-adjusted analyses, the strongest residual signal was observed in broad beta, with a closely similar nominal pattern in low beta. Nevertheless, these effects did not survive Holm correction and therefore provide only cautious, non-confirmatory support for a beta-range oscillatory reduction in CWP. Residual high beta, alpha, and gamma showed no robust corrected effects, indicating that inferences outside the broad beta range are particularly uncertain at this tier. Overall, the Tier 2 aperiodic decomposition suggests a directionally preserved but statistically attenuated beta-dominant oscillatory pattern in the absence of evidence for a global aperiodic shift.

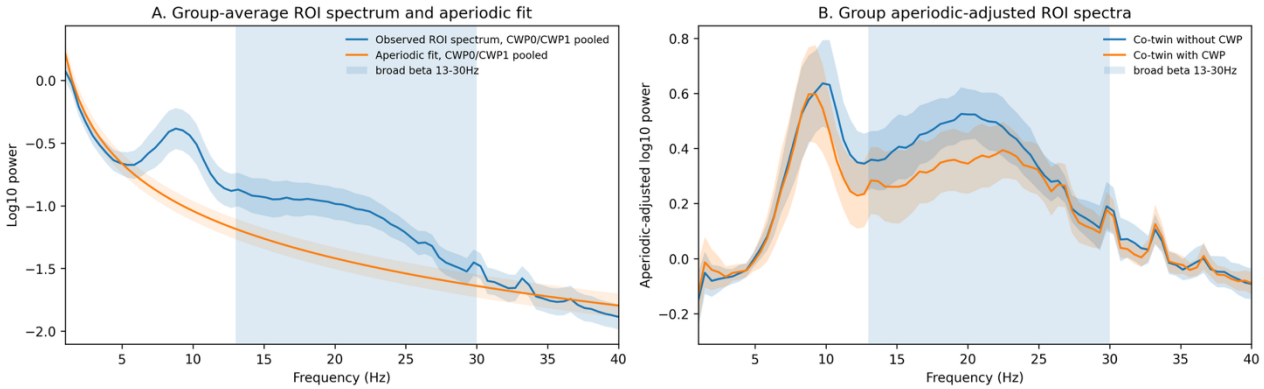


Figure 5.3. Aperiodic-adjusted frontocentral broad beta power in co-twins discordant for chronic widespread pain

Mean frontocentral ROI spectra are shown before and after aperiodic adjustment. Panel A shows the mean ROI power spectrum and fitted aperiodic component, plotted in log10 power units. Panel B compares co-twins without CWP (CWP0) and co-twins with CWP (CWP1) after subtraction of the fitted aperiodic component, shown as aperiodic-adjusted log10 power. The shaded vertical region indicates the predefined broad beta band, 13-30Hz, used for residual broad beta extraction. Shaded ribbons indicate 95% confidence intervals around the mean. Although spectra are displayed as group averages for visualisation, all formal analyses used paired CWP1-CWP0 contrasts within discordant twin pairs.

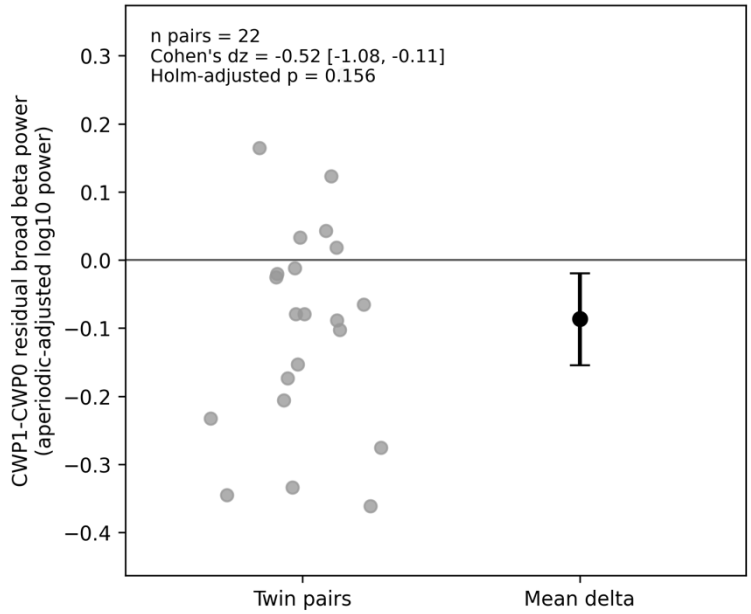


Figure 5.4. Within-pair difference in aperiodic-adjusted broad beta power in co-twins discordant for chronic widespread pain

Grey points show individual twin-pair differences in residual broad beta power within the predefined frontocentral ROI for co-twins with CWP (CWP1) relative to co-twins without CWP (CWP0). The black point shows the mean within-pair difference, with error bars representing the 95% bootstrap confidence interval. Negative values indicate lower aperiodic-adjusted broad beta power in the co-twin with CWP relative to the co-twin without CWP.

5.4.3 Secondary alpha peak analysis

For the occipital ROI, there was no evidence of a consistent within-pair difference in either IAF or alpha peak power at any tier. In the primary analysis of all twin pairs, IAF did not differ between twins with and without CWP (mean difference -0.132Hz, p=0.469), and alpha peak power was similarly unchanged (mean difference 0.100dB, p=0.926). This pattern remained directionally and statistically similar across stricter tiers.

In contrast, the central ROI (C3, Cz, C4) showed a more consistent directional pattern, with lower alpha peak measures in CWP across sensitivity tiers. In the primary analysis, central IAF was numerically lower in CWP but did not reach statistical significance (mean difference -0.193Hz, p=0.274), and central alpha peak power was also not statistically significant (mean difference -0.983dB, p=0.229). At Tier 1, both measures again showed reductions in the same direction, with a larger reduction in central alpha peak power (mean difference -1.621dB, p=0.086), although neither reached statistical significance. At Tier 2, central IAF was significantly lower in CWP (mean difference -0.551Hz, p=0.028), representing the clearest evidence in the alpha peak analyses for a within-pair central alpha frequency reduction. Central alpha peak power at Tier 2 remained directionally lower in CWP but did not reach statistical significance (mean difference -1.736dB, p=0.157).

These findings do not support a systematic shift in occipital alpha peak frequency or power, arguing against a generalised posterior alpha slowing effect in CWP. Instead, the stricter-tier analyses provide tentative evidence for a more localised reduction in central alpha peak frequency, with the clearest support emerging at Tier 2. Central alpha peak power showed a directionally similar reduction across tiers, but the statistical evidence was weaker and less consistent. Overall, the alpha peak analyses are therefore more consistent with alteration in central or sensorimotor alpha dynamics than with a robust global change in intrinsic occipital alpha rhythm, although this interpretation is made cautiously given the limited and tier-dependent nature of the central effects.

5.4.4 Exploratory functional connectivity

Given that the primary spectral analyses identified the most consistent effects in the alpha and beta ranges, exploratory resting-state functional connectivity was assessed using amplitude envelope correlation (AEC) in alpha, broad beta, low beta and high beta bands within the connectivity-only Tier 3 subset (n=12 pairs).

In the alpha band, no within-pair difference survived Holm correction across ROI-to-ROI comparisons. The strongest uncorrected effects were observed for central-central connectivity (mean difference z -0.082, dz -0.606, 95% CI -1.634 to -0.045, permutation p=0.062, Holm-corrected p=0.624), frontal-central connectivity (mean difference z -0.045, dz -0.581, 95% CI -1.282 to -0.061, permutation p=0.071, Holm-corrected p=0.637), and central-parietal connectivity (mean difference z -0.052, dz -0.574, 95% CI -1.136 to -0.082, permutation p=0.075, Holm-corrected p=0.637). All indicated lower alpha envelope coupling in the twin with CWP than in the co-twin control, but none reached corrected statistical significance. Remaining alpha-band contrasts were smaller and not statistically significant.

In the broad beta band, again no comparison survived correction for multiple testing, but the clearest exploratory effect was observed for central-central connectivity (mean difference z -0.063, dz -0.791, 95% CI -2.166 to -0.217, permutation p=0.029, Holm-corrected p=0.292). Smaller uncorrected effects in the same direction were observed for frontal-central connectivity (mean difference z -0.027, dz -0.509, 95% CI -1.578 to 0.010, permutation p=0.104, Holm-corrected p=0.939) and central-parietal connectivity (mean difference z -0.028, dz -0.516, 95% CI -2.265 to 0.057, permutation p=0.107, Holm-corrected p=0.939). Other broad beta contrasts were small and non-significant.

In the low beta band, no within-pair difference survived Holm correction. The strongest uncorrected effects were again centred on central connectivity, with lower central-central coupling in the twin with CWP (mean difference z -0.070, dz -0.659, 95% CI -1.671 to -0.102, permutation p=0.054, Holm-corrected p=0.540) and a similar pattern for central-parietal connectivity (mean difference z -0.041, dz -0.572, 95% CI -1.458 to -0.038, permutation p=0.075, Holm-corrected p=0.678). Frontal-parietal and frontal-central contrasts showed smaller effects in the same direction, while remaining comparisons were negligible.

In the high beta band, no comparison survived multiplicity correction, but the strongest nominal effect once more involved central-central connectivity (mean difference z -0.062, dz -0.766, 95% CI -2.014 to -0.173, permutation p=0.033, Holm-corrected p=0.330). Central-occipital (mean difference z -0.014, dz -0.540, 95% CI -1.382 to -0.0003, permutation p=0.083, Holm-corrected p=0.749) and central-parietal connectivity (mean difference z -0.022, dz -0.537, 95% CI -1.419 to 0.011, permutation p=0.085, Holm-corrected p=0.749) showed weaker effects in the same direction. All other high beta contrasts were small and clearly not statistically significant.

Collectively, these exploratory Tier 3 AEC analyses did not identify any ROI-level connectivity differences that survived correction for multiple comparisons. Nevertheless, a recurrent directional pattern was evident across alpha, broad beta, low beta, and high beta, with lower connectivity in contrasts involving central electrodes, particularly central-central and central-parietal coupling in CWP. This pattern was most prominent in broad beta and high beta, where the largest nominal effect sizes were observed, with similar but slightly weaker trends in alpha and low beta. Within the limits of this exploratory sensor-space analysis, these findings suggest that any connectivity abnormality in CWP is likely to be modest, centred primarily on central network coupling, and insufficiently robust at this sample size to support strong inferential claims.

5.4.5 Exploratory microstate analysis

Exploratory microstate analyses were conducted in the Tier 2 dataset (n=22 pairs). For the primary K=5 solution, no microstate metric demonstrated a significant within-pair difference between twins with and without CWP. Across all microstate classes, mean within-pair differences were small and effect sizes were uniformly low. For example, MS1 duration was essentially unchanged (mean difference -0.01ms, dz -0.001, p=0.669), while MS2 duration showed only a small positive difference (mean difference 1.59ms, dz 0.175, p=0.421). Coverage, occurrence, and GEV were similarly null across the K=5 solution, with all Holm-adjusted p-values equal to 1.0.

Findings were comparable across alternative clustering solutions (K=4, K=6, K=7, and K=8), with no reproducible state-specific pattern emerging across the wider K sweep. At K=4 and K=6, all effects remained small and not statistically significant,

with paired effect sizes generally close to zero with no suggestion of a coherent directional profile across duration, coverage, occurrence, or GEV. At K=8, the largest effects were still modest, for example MS1 GEV (dz 0.259, p=0.238) and MS1 coverage (dz 0.255, p=0.245), but these did not approach statistical significance and were not supported by consistent findings across related metrics.

The only signals that approached nominal statistical significance were observed at K=7 for microstate 6, where GEV showed the strongest trend across the entire microstate analysis (mean difference -0.0062, dz -0.413, p=0.067). However, these effects remained not statistically significant, all Holm-adjusted p-values were 1.0, and no corresponding robust difference was observed across the primary K=5 solution. As such, these isolated signals are most appropriately interpreted as weak exploratory trends rather than evidence of a stable disease-related microstate alteration.

Overall these results provide no strong evidence that resting-state microstate temporal organisation or explained variance differs systematically between twins with CWP and their co-twins. In contrast to the robust oscillatory differences identified in the spectral analyses, the present exploratory microstate analysis did not identify a reproducible within-pair microstate signature of CWP.

5.4.6 Exploratory symptom-association analyses

Across the full cohort, there was no evidence that within-pair difference in clinical characteristics were associated with within-pair differences in ROI-averaged alpha, beta, low beta or high beta power, with no predictor surviving Holm correction across both ROIs. The most prominent uncorrected pooled association was observed for GAD-7 in the frontocentral high beta model (β -0.97, 95% CI -1.92 to -0.03, p=0.043, Holm p=0.217) and PHQ-2 in the parietal high beta model (β -0.98, 95% CI -2.04 to 0.08, p=0.071, Holm p=0.355). Fatigue intensity, pain intensity and FS showed no corrected-significant pooled association in any band or ROI, suggesting that any effects were not uniform across the full cohort.

Interaction analyses then tested whether these associations differed by zygosity. The clearest effect was observed for fatigue in the frontocentral low beta outcome, where the association between fatigue discordance and EEG power discordance differed between MZ and DZ pairs (interaction coefficient -2.68, 95% CI -4.58 to -0.77,

p=0.006, Holm p=0.030). Directionally similar but non-significant interaction effects were seen for fatigue in frontocentral broad beta (interaction coefficient -2.18, p=0.013, Holm p=0.066), parietal alpha (interaction coefficient -2.51, p=0.014, Holm p=0.069), and parietal low beta (interaction coefficient -2.08, p=0.028, Holm p=0.139). No robust interaction effects were identified for pain intensity, FS, depressive symptoms, or anxiety symptoms.

Further subgroup analyses indicated that this pattern was driven primarily by the MZ cohort**.** In MZ pairs, greater fatigue discordance was positively associated with beta-family power differences across both ROIs, with multiple associations surviving Holm correction in frontocentral broad and high beta and across parietal broad, low, and high beta (β range 0.97 to 1.25; Holm-corrected p range 0.003 to 0.039). A similar association was also observed for frontocentral low beta (β 1.34, 95% CI 0.24 to 2.44, p=0.017), although this did not survive correction (Holm p=0.067). In contrast, DZ pairs showed no Holm-significant associations in any band or ROI. Several uncorrected effects were observed for PHQ-2 in the parietal beta-family models (β range -1.70 to -1.89; p-value range 0.028 to 0.038), but none remained significant after correction.

5.5 Discussion

In this co-twin control study of MZ and DZ twin pairs discordant for CWP, the primary and most reproducible rsEEG finding was reduced frontocentral beta power in the CWP-affected twin relative to the co-twin without CWP. This effect was evident across broad beta and both beta sub-bands, persisted across the stricter Tier 1 and Tier 2 spectral quality-control analyses, and showed the strongest nominal residual effects after aperiodic adjustment in broad beta and low beta, although these residual effects did not survive correction for multiple comparisons. Alpha also showed significant negative clusters in the primary spectral analysis and across stricter spectral quality-control tiers. Occipital alpha peak frequency and peak power were not altered, arguing against a generalised posterior alpha slowing phenotype. The clearest alpha peak finding was a Tier 2 reduction in central IAF, while central alpha peak power showed only a non-significant directional reduction, and residual alpha did not survive correction after aperiodic separation. Theta-band power was not altered, arguing against a universal low-frequency spectral signature of CWP. Gamma showed a significant negative cluster in the primary spectral analysis, but no

residual gamma difference was identified after removal of the aperiodic component. Aperiodic exponent and offset were unchanged, and exploratory connectivity and microstate analyses were not statistically significant after correction. These findings indicate that reduced frontocentral beta-range activity is the most robust electrophysiological correlate of CWP in this sample of MZ and DZ discordant twins.

The zygosity-stratified analyses further qualified this interpretation. DZ pairs showed broader negative clusters across alpha, broad beta, low beta and high beta bands, whereas MZ pairs showed a more selective pattern, with high beta as the only band retaining a significant cluster-level difference. This pattern supports beta-range physiology as the most credible signal across analyses, but suggests that the breadth of spectral differences is reduced when genetic background is most tightly controlled. These subgroup findings should be interpreted cautiously because the MZ and DZ analyses were not powered for formal interaction testing; nevertheless, the persistence of a high beta effect in MZ pairs strengthens the interpretation that at least part of the beta-range phenotype is associated with CWP rather than reflecting shared inherited neurophysiological variation alone.

These findings are concordant with and refine the existing rsEEG literature in chronic pain. Prior work has been heterogeneous, with the strongest cross-sectional support for differences in theta and beta activity, but with marked inconsistency in direction, topography, and reproducibility across studies[44, 128]. More specifically, recent synthesis of quantitative EEG findings in fibromyalgia, chronic fatigue syndrome and long covid concluded that the literature remains mixed and methodologically variable[147]. Against that background, the present study does not simply replicate one side of an inconsistent literature. Rather, it suggests that when shared familial and genetic factors are tightly controlled, the most stable spectral correlate is lower frontocentral beta power. This work raises the possibility that at least some previously reported case-control EEG findings may partly reflect genetic liability, sampling differences, or broader phenotypic heterogeneity rather than pain-relevant neurophysiology alone.

This interpretation is supported to some extent by the aperiodic-periodic analysis. A recurring concern in EEG biomarker studies is that conventional band-power measures may conflate oscillatory changes with shifts in the aperiodic background spectrum. In the present study, neither aperiodic exponent nor offset differed

significantly between discordant twins, arguing against a robust between-twin difference in the broadband aperiodic spectral background. Following subtraction of the aperiodic component, broad beta and low beta showed the strongest residual reductions in CWP, but these effects remained significant only at the uncorrected level. The findings therefore provide cautious support for the view that the beta-range signal of interest is more consistent with oscillatory alteration than with a nonspecific shift in broadband spectral slope. This interpretation also aligns with recent larger-scale work showing that the EEG aperiodic component did not differentiate people with chronic pain from healthy participants, nor correlate with pain intensity[158]. Accordingly, the current data argue against a simple model in which CWP is characterised by a broad resting-state change in spectral background activity, and instead support a more selective disturbance in beta-range oscillatory dynamics.

The centrality of beta is biologically plausible. Beta oscillations are strongly linked to sensorimotor processing, maintenance of the current sensorimotor or cognitive set, and top-down regulation of ongoing processing[151, 152]. Within sensorimotor systems, beta activity is typically most prominent during relative stability and is reduced when the system shifts toward change, updating, or altered state estimation[151, 152]. In that context, lower resting frontocentral beta in CWP may be consistent with a less stable or less efficiently regulated sensorimotor state, potentially reflecting altered integration of bodily signals, impaired inhibitory control, or weaker top-down constraints on ongoing somatic input. This interpretation is also broadly compatible with prior pain literature implicating beta band activity in chronic pain pathophysiology, including work proposing beta oscillations as candidate inhibitory or GABA-related markers[156]. However, the present results do not justify a strong mechanistic claim. Resting-state beta reduction is not specific to chronic pain, and these data do not determine whether the observed effect indexes pain persistence, nociplastic processing, fatigue-related burden, compensation, or another correlated process. The most defensible conclusion is that reduced frontocentral beta identifies a reproducible neurophysiological state associated with CWP that remains detectable even after controlling for major shared familial confounders.

The current findings also help contextualise prior fibromyalgia-specific EEG research. Some case-control studies have described increased beta or high-beta

activity and increased beta-range connectivity in fibromyalgia[153, 154], whereas others have reported lower beta activity or a negative relationship between beta oscillations and pain severity[155, 156]. The present co-twin design is important because it reduces the influence of between-person confounding that may contribute to these discrepancies. Rather than implying that prior studies were incorrect, the present data suggest that directionality may depend in part on study design, sample composition, and the extent to which familial or genetic background is controlled. In a community-based twin sample with mild symptom burden, the signal that survives the strictest comparison is lower frontocentral beta. That is arguably a more conservative and therefore more credible biomarker candidate than effects that depend on unmatched case-control contrasts alone.

Alpha appears to be a secondary finding rather than a co-primary signal. The broader pain literature has often focused on peak alpha frequency, especially posterior alpha slowing, as a putative marker of pain sensitivity or chronic pain vulnerability[148]. In the present dataset, however, occipital alpha peak frequency and alpha peak power were not consistently altered between discordant twins, arguing against a strong global posterior alpha slowing phenotype in this cohort. Instead, the alpha peak analyses suggested a more localised and tier-dependent central effect. Patterns observed in central IAF and alpha peak power are consistent with the possibility that alpha alterations in CWP are spatially selective, potentially involving central or sensorimotor rhythms, rather than representing a dominant global posterior rhythm change. The residual alpha findings reinforce the need for caution, as alpha differences weakened after aperiodic separation and did not survive correction. Alpha therefore represents a credible candidate alteration in CWP, but should be considered a secondary signal relative to the more robust beta-family findings in the current study.

Gamma should be interpreted with appropriate restraint, rather than as a robust mechanistic finding. Although cluster-corrected analyses showed significant negative gamma-band clusters, no residual gamma difference was identified after aperiodic adjustment. This weakens the interpretation that the gamma effect reflects a specific oscillatory alteration. Similarly, the vulnerability of higher-frequency scalp EEG signals to non-specific variance reduces the confidence that significant findings

reflect true changes in CWP. Therefore, the present study does not provide convincing evidence that gamma is a stable oscillatory feature of CWP.

Perhaps surprisingly given the previous chronic pain EEG literature, this study did not find a significant alteration in theta-band power. Systematic review evidence suggests that increased theta power is among the most commonly reported cross-sectional rsEEG findings across chronic pain cohorts[44], and fibromyalgia-specific work has reported augmented prefrontal and anterior cingulate theta activity associated with tenderness and fatigue[145]. Low-frequency theta prominence has also been interpreted within TCD frameworks derived mainly from neurogenic or neuropathic pain, where excess low-frequency power and slowing of the dominant rhythm have been linked to deafferentation-related thalamocortical disturbance[146]. However, that framework has not generalised uniformly across pain phenotypes, and a controlled study of moderate chronic back pain did not identify a statistically significant TCD-like group effect[304]. In that context, the present null theta result is important. It suggests that low-frequency spectral abnormalities are not a necessary or universal rsEEG feature of CWP, particularly in a mildly symptomatic, genetically informed sample. It is also notable that the recent mega-analysis identified the most robust theta-related associations at the level of large-scale connectivity with pain intensity, rather than a replicated theta power difference per se[128]. The null theta power finding therefore does not so much contradict the broader literature as refine it, suggesting that theta may be more relevant to network-level pain processing, specific clinical phenotypes, or more severe pain states than to within-pair spectral power differences in CWP.

The exploratory network-level analyses are also informative because they did not identify corrected group differences. No connectivity or microstate effect survived correction for multiple comparisons. Prior fibromyalgia studies have reported increased beta band connectivity across distributed networks and altered microstate parameters[153, 154], as well as disrupted theta-band resting-state networks[144]. The current findings do not reproduce those observations at a corrected level in a co-twin framework. This should not be interpreted as evidence that network-level alterations are absent in CWP, particularly given recent mega-analytic results suggesting that theta-related abnormalities may be more apparent at the level of large-scale connectivity with pain intensity than as replicated theta power differences[128]. Rather,

the present data indicate that any network-level signal was not sufficiently robust in this cohort and analytical framework to support strong inferential claims. Although no connectivity comparison survived correction, the strongest nominal effects repeatedly involved lower central network coupling in CWP, particularly central-central connectivity in broad beta and high beta, with weaker directional patterns in alpha and low beta. These exploratory findings are directionally consistent with the central beta band spectral emphasis. Within the limits of a 32-channel sensor-space dataset, the present results suggest that sensor-level beta power may be a more robust phenotype than higher-order connectivity or microstate features in this cohort. An equally plausible explanation is that the study was underpowered for these more complex measures, particularly given the reduced sample sizes in the Tier 2 and Tier 3 exploratory analyses.

The univariate analyses investigating relationships between rsEEG band power differences and clinical variables add further nuance. Within-pair differences in pain intensity, fibromyalgia severity, depression, and anxiety were not robustly associated with within-pair EEG differences after correction, indicating unsurprisingly that the beta phenotype is not simply a linear reflection of contemporaneous symptom ratings. rsEEG may index a more trait-like physiological disposition, whereas single-time-point self-report measures capture only part of the clinically relevant variance. The exploratory fatigue-by-zygosity pattern is potentially interesting, particularly because the MZ subgroup showed stronger associations between fatigue discordance and beta band power. However, this finding should be interpreted cautiously. It arose in exploratory analyses, subgroup sizes were modest, and the pattern was not mirrored consistently across symptom domains. It is therefore better framed as a signal-generating observation than as evidence of a specific fatigue-beta mechanism.

A major strength of this study is the co-twin control design itself. rsEEG traits are substantially heritable, particularly spectral power measures[159, 161, 162], and conventional case-control studies therefore risk attributing to pain what may actually reflect underlying familial or genetic variation. By comparing twins discordant for CWP within pair, and by including both MZ and DZ pairs, the present study directly addresses that limitation[165]. The design does not provide formal heritability estimates, nor do the exploratory zygosity findings support a simple causal model.

Nonetheless, it allows a materially stronger claim than is usually possible in chronic pain EEG research, namely that reduced frontocentral beta is associated with CWP over and above age, sex, early shared environment, and, in MZ pairs, virtually all shared genetic variation. That substantially increases the credibility of beta as a candidate biomarker phenotype. An additional strength is the use of an adapted automated rsEEG preprocessing framework. In a field where methodological heterogeneity has been a major barrier to replication, automated preprocessing reduces analyst-dependent subjectivity and decision variability, improves transparency and reproducibility, and supports more standardised feature extraction across studies[44, 128, 234]. Although implementation was adapted to the requirements of this study and is therefore not directly identical to other applications of the pipeline, the use of a standardised automated preprocessing framework strengthens methodological rigour and enhances the interpretability of the findings within the broader rsEEG biomarker literature. From a practical perspective, the acquisition paradigm was also deliberately pragmatic. The brief eyes-closed recording protocol and rapid sponge-based 32-channel setup enabled EEG collection to be embedded within the wider TwinsUK workflow with low participant burden and limited operator burden, making the approach feasible within a real-world registry setting. This pragmatic design is important for future biomarker development, because scalable longitudinal and nested cohort studies will require methods that are not only analytically robust but also operationally deliverable across large research ecosystems.

There are several limitations to this study. The sample, although strong for a co-twin EEG study, remains modest for stratified and multivariate inference, especially in Tier 2 and Tier 3 analyses. The cohort was predominantly female with a mild symptom phenotype overall, which may limit generalisability to more severe pain populations. Relatedly, questionnaire-derived co-morbidity and medication data were available for most participants and provided useful clinical context, but these variables were not confirmed at the time of EEG acquisition due to feasibility challenges. Questionnaire timing varied, some diagnoses were historical, and medication dose, indication, adherence and ongoing use were not confirmed. Twins with CWP had greater recorded osteoarthritis and analgesic/opioid exposure, which is clinically consistent with the pain phenotype but also indicates individual-level clinical heterogeneity not

controlled for. Medication exposure and acquired co-morbidities therefore remain possible contributors to observed EEG differences. The same pragmatic 32-channel sponge-based acquisition that supported feasibility within the TwinsUK ecosystem also limits spatial precision and may reduce sensitivity for higher-order measures such as connectivity and microstates. The trade-off, therefore, was between scalability and neurophysiological granularity. Reassuringly, the principal beta findings strengthened rather than weakened under stricter quality thresholds, which argues against poor data quality as the explanation, although more spatially resolved acquisitions may be needed to evaluate network-level and source-level abnormalities more sensitively. Finally, the symptom, connectivity, microstate, and zygosity-stratified analyses should all be regarded as secondary or exploratory relative to the primary spectral result.

Future work should prioritise replication of the beta finding in larger genetically informative samples and in clinically richer nociplastic pain cohorts. Furthermore, future studies should also determine whether theta-related abnormalities in CWP are better captured at the level of large-scale connectivity, dynamic coupling, or more severe clinical phenotypes than by simple spectral power alone. Longitudinal designs would be especially valuable in determining whether reduced frontocentral beta tracks persistence of widespread pain, fatigue fluctuation, treatment response, or transition between pain states. Denser montages, source-space modelling, and dynamic analyses such as beta-burst quantification may also provide more mechanistic precision than static band averages alone. More broadly, the present results support a biomarker strategy focused on interpretable oscillatory phenotypes, especially beta, rather than assuming that increasingly complex derived measures will necessarily prove more robust. In this respect, beta appears to be the most promising signal emerging from the current dataset, while alpha may reflect a weaker secondary sensorimotor feature and gamma does not presently warrant mechanistic interpretation.

In conclusion, this co-twin control study provides evidence that CWP is associated primarily with reduced frontocentral beta oscillatory power at rest, with weaker and more localised alpha abnormalities, and without evidence for a theta-band power abnormality, a persistent gamma-band effect, or a global shift in the aperiodic spectral background. The convergence of sensor-level and aperiodic-adjusted

analyses indicates that beta reduction is the most reproducible electrophysiological feature in this dataset. While this should not yet be considered a diagnostic biomarker, it represents a biologically plausible and methodologically robust candidate biomarker phenotype that remains detectable after substantial control for shared familial and genetic confounding.

6. Study 2: VIPA-EEG: Investigating resting-state EEG changes following virtual reality exposure in fibromyalgia

This chapter presents the VIPA-EEG study, a secondary analysis of repeated-measures resting-state EEG (rsEEG) acquired around an attention-based immersive VR intervention in fibromyalgia. Six within-session recordings support both sequential post-activity contrasts and an overall pre-post comparison, with periodic-aperiodic decomposition used to distinguish oscillatory modulation from broadband spectral change, and secondary analyses examining alpha peak properties and EEG-symptom associations. The study contributes the state-modulation component of the empirical work, evaluating whether candidate rsEEG features are responsive to immersive therapeutic exposure.

The principal finding was increased alpha and beta band power following VR exposure, with alpha the clearest and most progressive within-session effect. The aperiodic background itself was unchanged. The alpha increase survived adjustment for the aperiodic component and was supported by higher occipital and central alpha peak power, whereas the broad and low beta increases weakened after adjustment and did not survive correction, appearing more activity-sensitive than the progressive alpha response. Delta, theta, high beta and gamma were not significant. EEG-symptom association analyses showed no effects surviving correction.

6.1 Introduction

Identifying neurophysiological features associated with chronic pain is only one part of biomarker development. For such features to support immersive therapeutic design, they need to be at least partially modifiable around intervention. Modifiability is not a prerequisite for biomarker relevance, but it is essential if a candidate domain is to act as a process indicator, an engagement marker or a target for therapeutic adaptation.

Fibromyalgia is an appropriate phenotype for this question as it represents an archetypal nociplastic pain condition in which altered central nervous system processing contributes to symptom persistence and amplification[40, 41]. Although cross-sectional rsEEG differences in fibromyalgia have been characterised in some detail[127, 145, 147], the within-subject question of whether candidate features are dynamically responsive to intervention has comparatively little empirical evidence. It therefore remains unclear whether reported features represent stable trait-level abnormalities, dynamic state-sensitive markers, or both.

Immersive VR provides a plausible context in which to examine this question. VR can influence attention, sensory context, affective state and behavioural engagement, all of which are relevant to chronic pain processing[50, 51, 55]. Attention-based VR has shown feasibility and preliminary signals of symptom benefit in fibromyalgia, although responses vary across systems, environments and activities[8]. Direct evidence linking VR exposure with EEG modulation in chronic pain remains limited, with the small existing literature concentrated on populations and acquisition designs distinct from those examined here[305].

The VIPA-EEG study addresses this gap through secondary analysis of repeated-measures rsEEG data within a previously completed VR feasibility study in fibromyalgia[8] (Appendix 1). Six within-session eyes-closed rsEEG recordings were obtained per participant, comprising a pre-intervention baseline, four post-activity recordings under headset-on conditions, and a final post-intervention recording obtained after headset removal. This structure allowed separation of activity-specific change from cumulative within-session effects, while acknowledging the asymmetry between matched intermediate recordings and the pre-post contrast. Primary spectral analyses are complemented by periodic-aperiodic decomposition, alpha peak analyses and examination of associations between EEG change and symptom measures. The objective was to test whether candidate rsEEG features in fibromyalgia are dynamically modifiable around immersive VR exposure.

6.2 Study aims

The primary aim of this study was to determine whether exposure to an attention-based immersive VR intervention, designed for chronic pain, was associated with

measurable changes in rsEEG activity in individuals with an archetypal nociplastic pain condition, fibromyalgia syndrome.

Within the context of the thesis, this study addressed whether candidate rsEEG features identified in CWP within PainWaves and the wider chronic pain literature, were dynamically responsive to immersive intervention. It therefore represents a translational step from trait-level biomarker identification towards evaluation of potential physiological modifiability.

A secondary aim was to characterise the nature of any observed spectral changes, specifically to determine whether they were better explained by alterations in oscillatory (periodic) within canonical frequency bands or by shifts in the aperiodic background component of the EEG spectrum.

Additional aims were to examine whether observed EEG changes varied across individual VR activities and/or following the overall intervention session, and to explore whether changes in candidate EEG domains corresponded with contemporaneous symptom measures, including pain intensity.

6.3 Methods

6.3.1 Study design and participants

This study employed a repeated-measures design in a cohort of participants with fibromyalgia undergoing an attention-based virtual reality (VR) intervention. The present study reports secondary analysis of rsEEG data acquired within a previously completed feasibility study (the VIPA Study, Appendix 1)[8]. VIPA investigated attention-based VR interventions in fibromyalgia with comparison of different VR systems, activities, and environments.

Participants fulfilled the 2016 fibromyalgia diagnostic criteria and were recruited through Rheumatology services into the parent feasibility study[8]. The present EEG analysis was restricted to the activity-comparison session of that study (Data Collection Session 2), in which participants completed four interactive VR activities within a common virtual environment. A total of 23 participants contributed complete repeated-measures EEG data and were included in the analysis. The repeated-measures design enabled both sequential within-session comparisons and an overall pre-post intervention contrast within participant.

6.3.2 Ethics and governance

The parent feasibility study received ethical approval from the Cornwall and Plymouth Research Ethics Committee (REC reference 20/SW/0050)[8], with all procedures conducted in accordance with the Declaration of Helsinki. Written informed consent (Appendix 6) was obtained from all participants prior to data collection.

Further ethical approval was obtained from the University of East Anglia (UEA) Faculty of Medicine and Health Research Ethics Sub-committee (REC reference ETH2425-2568, Appendix 5; the same approval as Study 1) for secondary analysis of data conducted in the present study, with appropriate data governance in place.

6.3.3 VR intervention context

The present analysis focused specifically on the activity-comparison session (Data Collection Session 2) within the parent VIPA study (Appendix 1). Participants used a head-mounted VR system (Oculus Rift S with hand-held controllers) to complete four interactive tasks within a common virtual environment depicting a nature-based scene with a cold and snowy weather system. Activities were designed to engage attention through distinct cognitive demands, including visuospatial processing, working memory, multitasking, and sensorimotor coordination. The order of activities was randomised between participants to minimise systematic order effects.

Participants remained seated throughout VR use, and interaction was enabled via controller-based input (right-handed controller) with minimal physical movement requirements. Following each activity, participants completed brief in-headset questionnaires, after which an rsEEG recording was obtained.

The activity-comparison session was selected for EEG analysis because it held the VR hardware and virtual environment constant while varying the activity performed, provided four post-activity rsEEG recordings acquired using the same protocol within a single continuous VR exposure, and had the largest sample size in comparison with the environment-comparison session. The activities were developed within the parent study as active, attention-based tasks, reflecting a well-established rationale for VR analgesia in which engagement with competing cognitive demands may reduce the attentional resources available for pain processing[55-58]. Active interaction was prioritised over passive viewing because several experimental and clinical

studies have reported greater analgesic effects during interactive than passive VR exposure[73, 87-90]. The four activities delivered different forms of engagement, including visuomotor coordination, working memory, planning and divided attention, and visuospatial pattern recognition (Table 6.1). This provided a broader range of cognitive demands than a single repeated task and allowed investigation of whether post-activity rsEEG changes were common across attention-based VR activities or differed according to the form of engagement.

The virtual environment remained the same across activities so that activity-related comparisons were not additionally confounded by changes in scene or environmental context. A cold, snowy, nature-based setting was chosen within the parent feasibility protocol, informed in part by the winter environments used in early VR analgesia research[72]. The activities were not primarily designed to manipulate cognitive load and differed in pacing and visual content as well as in their predominant cognitive demands. Consequently, the analysis can examine whether rsEEG changes occur consistently around attention-based VR and whether post-activity rsEEG differs between activities with broadly different demands, but it cannot attribute any observed difference to a specific cognitive process.

Accordingly, the repeated-measures EEG design was used to examine two related but distinct effects: (1) short-term changes in resting-state neural activity following individual VR activities under consistent acquisition conditions, and (2) overall pre-post intervention changes across the full VR session.

Table 6.1. Comparison of the VR activities investigated in the parent feasibility study[8]

Activity Type	Main Skills Required	Short Activity Description	Image
Rail-shooter	Hand-eye co-ordination and agility.	To aim at and shoot targets appearing in the sky whilst travelling steadily along a calm river.	

Memory	Short-term and working memory.	To recall sequences of objects emerging from plots within a 3x3 grid. Upon correct recall, sequences increased in length.	
Multitasking	Strategic planning and attention allocation.	To follow a series of steps to plant, maintain and harvest trees across five plots simultaneously.	
Match-3	Pattern recognition, visuospatial awareness and strategic planning.	To match three of the same objects in a row across three vines by moving vines up or down.	

6.3.4 EEG acquisition

rsEEG data were acquired using the 64-channel actiCAP active electrode configuration described in Chapter 4 (Section 4.1.1). Lower impedance values are expected with active gel-based acquisition than with the saline-based sponge system used in Study 1, reflecting the different electrode-scalp interface[306]. Impedance information recoverable from the BrainVision header files was summarised across the repeated-measures archive. Across 165 resting-state files, representing 10,560 channel-level observations, 60.4% of channel observations were classified as good (≤25kΩ), 12.1% as acceptable (>25 to 50kΩ), 1.1% as poor (>50 to 100kΩ), and 1.0% as very poor (>100kΩ). A further 2.1% were labelled out-of-range. Across valid numeric entries, mean electrode impedance was 15.5 ±24.6kΩ, with a median of 11.0kΩ. Median impedance remained stable across session stages, ranging from 10.0-12.0kΩ, with interquartile ranges consistently between 3.0-22.0kΩ. Poorer or unavailable impedance values were not concentrated within the principal frontocentral ROI channels used in later effect-size analyses, instead being more evident at peripheral and lateral sites, particularly TP10, FT10, TP9 and FT9, with additional burden at some posterior channels including O1, O2 and POz. Overall,

these distributions were consistent with acceptable and stable acquisition quality across the repeated-measures session.

Participants were seated comfortably in a quiet research room at the University of East Anglia. The repeated-measures rsEEG protocol comprised six recording stages across the intervention session. A pre-intervention baseline recording was acquired before VR exposure and before placement of the head-mounted display (HMD), comprising 2 minutes of eyes-closed recording followed by 2 minutes of eyes-open recording with visual fixation. Four intermediate post-activity resting-state recordings were then obtained after completion of each VR activity. These post-activity recordings consisted of 2 minutes of eyes-closed recording only and were typically acquired within 1-2 minutes of activity completion, after participants had completed the brief in-headset virtual questionnaire. During these intermediate recordings, the VR headset remained in place over the EEG cap and the headset display was set to black. A final post-intervention resting-state recording was then acquired after the full VR sequence had been completed and the headset had been removed, again comprising 2 minutes of eyes-closed recording followed by 2 minutes of eyes-open recording with visual fixation.

This acquisition structure was important for later interpretation. The four post-activity recordings provided the most internally standardised repeated-measures observations because they were all obtained under the same headset-on, black-screen, eyes-closed condition. By contrast, the overall pre- versus post-intervention comparison was informative as a session-level endpoint but differed in acquisition context because both recordings were obtained without the headset in place. Accordingly, the repeated post-activity sequence and the final pre-post comparison were treated as related but not identical inferential components of the study.

At the onset of each recording condition, a manual event marker was inserted to denote the start of the relevant resting-state segment. Analyses focused exclusively on eyes-closed recordings, consistent with the resting-state framework described in Chapter 4, both to maximise comparability across the six repeated-measures observations and to preserve methodological continuity with Study 1.

6.3.5 Self-reported measures

Participants completed a range of self-reported measures within the parent feasibility study to characterise symptom burden, mood, and pain response. Baseline measures included fibromyalgia severity indices, depressive symptoms (PHQ-9), and anxiety symptoms (GAD-7)[8]. Pain intensity was also assessed repeatedly using VAS ratings obtained at baseline, after each VR activity and following the intervention. The McGill Pain Questionnaire short-form (MPQ-SF)[307] was administered before and after the intervention to characterise changes in pain intensity and quality.

For the present EEG analyses, pain VAS was the only clinical measure acquired repeatedly at within-session timepoints corresponding to the repeated EEG recordings. This enabled modelling of changes in pain response across the session alongside neurophysiological change. A single-item rating was also practical within the VR protocol because it could be completed between activities without removing the headset, minimising disruption, while visual analogue ratings are sensitive to short-term changes in pain intensity[294]. The MPQ-SF provided a complementary pre- to post-intervention measure incorporating sensory and affective dimensions of pain[307]. This was relevant to attention-based VR, where changes in the affective experience of pain may not necessarily parallel changes in pain intensity. The FIQR, PHQ-9 and GAD-7 were used for baseline characterisation only, providing established measures of fibromyalgia impact, depressive symptoms and anxiety and allowing comparison with other fibromyalgia cohorts[292, 308, 309]. Collectively, these measures captured pain intensity, emotional functioning and broader symptom burden relevant to chronic pain research[297]. Appendix 1 provides full details of all measures used.

6.3.6 Data preparation and preprocessing

Raw EEG recordings were converted to BIDS format and preprocessed using the automated pipeline described in Chapter 4 (Section 4.1.2). Conversion focused exclusively on eyes-closed rsEEG recordings from the repeated-measures intervention session. Across the analysed dataset, 138 eyes-closed recordings from 23 participants were included for downstream processing, corresponding to six repeated-measures observations per participant. All recordings had eyes-closed

event markers present prior to conversion, with at least 120s of eyes-closed data available after marker onset.

To ensure comparability across repeated measures, all recordings were trimmed to a fixed 120s eyes-closed segment anchored to the onset of the eyes-closed period using event markers. This standardisation minimised variability arising from differences in recording duration and ensured that downstream analyses were performed on consistent temporal windows across all timepoints. Electrode coordinate metadata were incorporated during BIDS conversion using the manufacturer 64-channel actiCAP montage, allowing standardised channel locations to be written into the BIDS sidecars.

6.3.7 Processed data quality

Across the repeated-measures dataset, retained post-clean duration was broadly consistent across timepoints, supporting comparability of usable signal across the intervention session. Variability in data quality was expected given the repeated-measures design, the use of a gel-based system across multiple recordings per participant, and the practical demands of acquisition before and after VR exposure. However, visual inspection of power spectral density outputs confirmed preservation of physiologically plausible EEG spectral structure. Across 138 recordings, a mean of 2.7 ±1.4 channels per participant were removed before interpolation (median 2, range 0-13). The number of retained clean epochs was 28.4 ±4.5 on average (median 29, range 6-47), corresponding to a mean retained post-clean duration of 56.8 ±9.0s (median 57s, range 12-94s). These audit outputs were used to characterise retained data quality across recordings and to support the later tiered sensitivity analyses.

6.3.8 Sample size and power

A formal a priori sample size calculation was not undertaken for EEG efficacy outcomes given that this study reports secondary analyses of data captured during a parent feasibility study. In the parent study, recruitment pragmatically aimed for 20-30 participants, a range considered sufficient to evaluate feasibility rather than to provide definitive efficacy testing[8]. The present EEG analyses therefore included all available participants from Data Collection Session 2.

Accordingly, power considerations for the present manuscript are descriptive rather than confirmatory. With 23 participants contributing complete repeated-measures EEG data, a two-tailed paired comparison at α=0.05 would provide 80% power to detect a within-subject effect of approximately Cohen's dz = 0.61. In sensitivity analyses using the stricter data-quality tier, the corresponding detectable effect increased to approximately Cohen's d*z* = 0.64 for Tier 1 (n=21). These thresholds indicate that the study was positioned to detect moderate-to-large within-subject EEG effects in the primary and Tier 1 analyses. These detectable-effect estimates are presented as a guide to interpretation rather than as a formal power calculation for every analytical component.

6.3.9 Inferential framework

All primary analyses were conducted using a within-subject repeated-measures framework, with each participant serving as their own control. Two complementary inferential approaches were employed.

First, a sequential contrast framework compared pre-intervention baseline recordings with each post-activity recording and with the final post-intervention recording. This approach assessed whether spectral changes emerged progressively across the session and whether they persisted at the endpoint.

Second, a post-activity modelling framework examined differences among the four post-activity recordings, enabling evaluation of activity-specific effects while accounting for cumulative exposure across the session.

6.3.10 Sensitivity analyses

To assess robustness to residual data-quality variation, relevant analyses were repeated across prespecified quality-control tiers. The same inferential framework was retained across tiers so that any change in results reflected data inclusion rather than analytical inconsistency.

Tier 0 was the inclusive dataset and applied no additional exclusions beyond the main preprocessing and feature-extraction pipeline (n=23). A single sensitivity tier, Tier 1, excluded recordings with retained post-clean duration <30s or ≥8 rejected channels (n=21). Tier 1 was therefore intended to define a more stable analysis subset through removal of clear low-quality outliers, rather than to function as part of a multi-level quality-control ladder.

The tiering variables were selected to capture the two features most likely to influence the stability of the planned downstream analyses: usable post-clean data length and channel-rejection burden. Retained post-clean duration was prioritised because all recordings had first been standardised to a fixed 120s eyes-closed segment, making post-clean duration the most direct index of how much analysable signal remained after artefact attenuation. The choice of a <30s exclusion threshold was informed by evidence that stable aperiodic spectral estimates can stabilise over relatively short resting-state recordings, including durations on the order of 30 to 120s, while also recognising the substantially shorter repeated-measures segments available in VIPA relative to conventional resting-state datasets[300]. Channel rejection burden was included because extensive channel removal and interpolation can reduce spatial fidelity and increase noise heterogeneity, which is particularly relevant for ROI-based spectral summaries and repeated topographical comparisons. A threshold of ≥8 rejected channels was retained as a marker of substantial spatial attrition within the 64-channel montage, while not further reducing the analysable sample.

Unlike PainWaves, VIPA did not use a multi-tier sensitivity framework. This reflected both the smaller sample size and the repeated-measures structure of the study, in which applying progressively stricter thresholds across all sessions rapidly produced a highly depleted subset that was no longer informative as a robustness check. In addition, the principal aims of VIPA centred on repeated spectral and aperiodic analyses, rather than on the broader connectivity and microstate analyses undertaken in PainWaves. A single pragmatic sensitivity tier was therefore considered more appropriate, allowing exclusion of clear low-stability recordings while preserving an interpretable within-subject sample.

6.3.11 Primary spectral analysis

Statistical inference for the primary spectral analysis was implemented in MATLAB R2025b[235] using FieldTrip[237] and custom analysis scripts. PSD and ROI-based effect size estimation, and cluster-based permutation testing were completed in line with the methodology described in Chapter 4 (Section 4.1.4), analogous to methods used in PainWaves (Chapter 5, Section 5.3.10). This included a two-sided framework with 5,000 randomisations, a cluster-forming threshold of $p<0.05$, paired effect size expressed as Cohen's dz with 95% bootstrap confidence intervals from 5,000

iterations, and frontocentral (Fz, FC1, FC2, Cz) ROI-based effect sizes used descriptively to complement the primary cluster-based inference. Two features distinguished the present implementation from PainWaves: the contrast structure described below, and the sensor layout in which electrode geometry was derived from the recorded 64-channel actiCAP positions.

The principal repeated-measures contrast structure compared the pre-intervention baseline recording with each of the four post-activity recordings in turn and, separately, with the final post-intervention recording. This framework was used to determine whether spectral differences from baseline were evident after early or later post-activity recordings, and whether any differences were present at the overall session endpoint.

To complement these paired contrast analyses, supplementary linear mixed-effects models were fitted to ROI-averaged post-activity band power. These models focused on the four headset-on post-activity observations and were used to examine whether post-activity spectral power varied as a function of session order and/or activity identity. This provided a separate ROI-level assessment of cumulative within-session effects and activity-specific differences, but was treated as supplementary to the primary sensor-level cluster-based inference. This is described further in 6.3.14.

6.3.12 Secondary aperiodic-periodic analysis

Aperiodic-periodic decomposition followed the FOOOF/specparam framework introduced in Chapter 4 (Section 4.1.5) and was implemented as in PainWaves (Chapter 5, Section 5.3.11), including the use of a fixed aperiodic model, parameterisation over 2-40Hz, the same Gaussian peak-fitting parameters, and the same Holm-Bonferroni correction structure applied separately across the aperiodic outcomes (offset, exponent) and the residual power related to canonical frequency bands.

Because stable estimation of aperiodic parameters depends on adequate retained broadband spectral structure, and may be more sensitive than conventional band-power analysis to shorter retained segments and heavier channel rejection, this analysis was conducted on the Tier 1 dataset. For each participant and recording state, ROI-averaged power spectra were derived from cleaned 2s eyes-closed epochs and parameterised using the FOOOF/specparam model.

Within-subject differences were calculated using the same repeated-measures contrast structure as the conventional spectral analyses, comparing baseline rsEEG with each subsequent post-activity recording and, separately, the final post-intervention recording. Paired t-tests were used as the inferential test for each FOOOF/specparam-derived outcome, with effect sizes expressed as Cohen's dz and 95% bootstrap confidence intervals (5,000 iterations).

6.3.13 Secondary alpha peak analysis

Individual alpha peak frequency (IAF) and alpha peak power were estimated as in PainWaves (Chapter 5, Section 5.3.12), using the same posterior (O1, O2, Oz, POz, Pz, PO3, PO4) and central (C3, Cz, C4) ROIs and methodology introduced in Chapter 4 (Section 4.1.4). Within-subject differences were assessed using paired t-tests across the same repeated-measures contrast structure as the conventional spectral analyses.

6.3.14 Repeated-measures modelling of post-activity EEG effects

To examine whether post-activity rsEEG differed according to the specific VR activity most recently experienced, linear mixed-effects models were fitted to ROI-averaged post-activity band power. These analyses were restricted to the four intermediate post-activity recordings, because these observations were all acquired under the same headset-on, black-screen, eyes-closed condition and therefore provided the most internally standardised basis for repeated-measures modelling.

For each ROI-band outcome, models took the general form: outcome ~ activity + order_in_session + (1|participant). Activity was treated as a four-level categorical fixed effect and was therefore represented in the model matrix by three contrast-coded terms, with one activity serving as the reference level. Order_in_session was entered as a continuous fixed effect to estimate linear change across the four post-activity recordings. This model structure allowed activity-related differences to be estimated whilst adjusting for cumulative within-session order effects. Participant was included as a random intercept to account for repeated observations within individuals. Random slope was not used because each participant contributed only four post-activity observations, making a random-intercept structure the more stable and interpretable choice.

The principal ROI-band outcomes examined were those most relevant to the spectral findings, namely frontocentral alpha, low beta, high beta and broad beta, and posterior alpha, low beta, high beta and broad beta. Where relevant, exploratory interaction models of the form outcome ~ activity * order_in_session + (1|participant) were also examined to assess whether any apparent activity effects were modified by where in the session that activity occurred.

Model outputs included fixed-effect coefficients, an omnibus test of the activity factor evaluating whether the set of activity contrasts improved model fit after adjustment for order in session, an omnibus test of the linear order effect, and pairwise activity contrasts derived from estimated marginal means where appropriate. Multiplicity correction was applied to omnibus order effects, omnibus activity effects, pairwise activity contrasts within outcome, and interaction omnibus effects across the different dependent variables using Holm-Bonferroni adjustment.

6.3.15 Clinical descriptives and missing-data handling

Participant characteristics and self-reported clinical measures were summarised descriptively to characterise the fibromyalgia sample and to contextualise the EEG analyses. Continuous variables were assessed for normality using the Shapiro-Wilk test and are reported using summary statistics appropriate to their distribution. Categorical variables are reported as counts and percentages.

Baseline descriptive variables were drawn from the parent feasibility study and included demographic and symptom measures relevant to fibromyalgia and mood[8]. For the present EEG work package, the principal repeated self-report outcomes were baseline pain VAS, post-activity pain VAS, post-intervention pain VAS, and MPQ-SF outcomes.

All relevant self-reported data were complete for the analysed dataset, and no imputation was required.

6.3.16 Models to evaluate EEG-clinical associations

Secondary EEG-clinical association analyses were performed to examine whether the principal EEG changes related to concurrent or overall changes in pain across the intervention session. These analyses focused on the main spectral findings identified from the conventional PSD analyses, namely frontocentral broad beta, frontocentral low beta, and posterior alpha. Clinical variables were selected to reflect

either repeated within-session pain response or participant-level overall change across the intervention session.

Repeated-measures mixed-effects analyses examined associations between EEG measures and pain ratings collected after VR activities. These models were restricted to the four post-activity rsEEG recordings and used current post-activity pain VAS as the primary outcome. The principal model took the form: current activity pain VAS ~ EEG + baseline_pain_VAS + order_in_session + activity + (1|participant), thereby accounting for repeated observations within participant, baseline pain severity, cumulative exposure across the session, and the identity of the immediately preceding activity. A lag-adjusted sensitivity model additionally included previous activity pain VAS to address short-lag pain autocorrelation across the session. Random-intercept models were used because each participant contributed only four post-activity observations, making more complex random-effects structures less stable and less interpretable. Mixed-effects models were fitted in MATLAB using restricted maximum likelihood. Holm-Bonferroni adjustment was used to account for multiplicity.

Separate participant-level pre-post modelling analyses examined whether overall session-level EEG change related to overall session-level change in pain outcomes. These analyses used one row per participant and modelled the change in post-intervention pain VAS or MPQ-SF outcomes as a function of the change in EEG metrics. Robust linear regression was used as the principal approach because of the modest sample size and the possibility of influential observations. Age was included as a covariate in all participant-level models. Numeric variables were standardised within the analysed dataset to facilitate coefficient comparability. As above, multiplicity correction across the primary EEG predictor p-values used Holm-Bonferroni.

6.4 Results

A total of 23 participants from Data Collection Session 2 contributed repeated-measures EEG data to the present analysis. Baseline characteristics are summarised in Table 6.2. Overall, the cohort comprised predominantly middle-aged, white British females with established fibromyalgia, substantial symptom burden, and moderate mood symptoms.

Descriptive self-report data indicated lower pain scores during and after the VR session. Mean pain VAS was 5.46 ±2.35 at baseline, 4.00 ±2.13, 4.17 ±2.46, 4.09 ±2.52, and 4.13 ±2.20 after Activities 1 to 4, respectively, and 4.48 ±2.55 post-intervention. Mean MPQ-SF total score decreased from 21.74 ±8.79 pre-intervention to 16.91 ±10.49 post-intervention.

Table 6.2. Baseline participant characteristics for the VIPA-EEG cohort

Characteristic	Total n=23
Age in years, mean (SD)	46.4 (13.8)
Female sex, n	20
Ethnicity, n	
White British	21
Black/African/Caribbean	1
Other (unspecified)	1
Duration of fibromyalgia diagnosis in years, mean (SD)	4.3 (5.9)
Fibromyalgia Severity (FS) score, mean (SD)	21.7 (4.5)
FIQR total score, mean (SD)	65.4 (16.6)
PHQ-9 score, mean (SD)	14.1 (5.5)
GAD-7 score, mean (SD)	9.9 (5.6)

Values are presented as mean (SD) or n. FIQR, Revised Fibromyalgia Impact Questionnaire; GAD-7, Generalised Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9; SD, standard deviation. Fibromyalgia Severity (FS) score was calculated as the sum of the Widespread Pain Index (WPI) and Symptom Severity Scale (SSS).

6.4.1 Primary spectral analysis

Across the repeated post-activity contrasts, cluster-based permutation testing demonstrated recurrent increase in alpha and beta-family power relative to baseline, with alpha effects becoming stronger across the session. Broad beta was the most consistent beta-family finding across activities, with low beta also showing reproducible significant effects. Theta and gamma changes were also observed in

several post-activity contrasts, but these were more variable across the session and less central to the overall pattern than the alpha and beta-family effects. Delta showed no consistent significant change.

The temporal pattern across the repeated contrasts suggested an evolving within-session spectral response. The earliest post-activity comparison (baseline vs post-activity 1) was characterised mainly by changes in the low and broad beta bands, together with additional high beta and gamma effects, indicating that spectral divergence from baseline was already detectable after the first VR exposure but remained relatively focused. By post-activity 2 this had broadened to include alpha and theta, and a similar pattern was seen post-activity 3. Post-activity 4 showed the clearest alpha effect of the session, occurring alongside continued beta-family effects and weaker theta and gamma findings. Overall, the repeated contrasts suggested a progressive within-session broadening of the EEG response, with alpha emerging as the dominant later-session feature while beta-family changes were present throughout. All findings were robust to sensitivity analyses.

Across the repeated post-activity contrasts, alpha showed a progressive spatial expansion across the session. No significant alpha cluster was identified after the first post-activity recording, whereas the later contrasts showed widespread positive clusters maximal over posterior regions with increasing extension into central, frontocentral and frontal electrodes. Broad beta and low beta were expressed predominantly as posterior-parietal positive clusters, with variable frontocentral extension that became more marked in the later post-activity contrasts. Low beta was generally the more spatially extensive sub-band effect, particularly in the final comparison. High beta and gamma were more posterior-focal in the earlier contrasts but showed broader central and frontocentral extension by the final post-activity comparison. Theta effects, when present, were smaller and less spatially consistent, being largely posterior or centroparietal with variable lateralisation. ROI-based effect-size estimates are summarised in Figure 6.3, with corresponding post-activity change topographies shown in Figure 6.4.

The final pre-intervention versus post-intervention comparison showed a more selective endpoint pattern than the repeated post-activity contrasts (Table 6.3 and Figure 6.1). Significant increase was observed in alpha, broad beta and low beta, all of which were robust to sensitivity analyses. The alpha effect was large (Tier 0 dz

0.75, 95% CI 0.44 to 1.23), with moderate effects observed for broad beta (Tier 0 dz 0.44, 95% CI 0.04 to 0.93) and moderate-large effects for low-beta (Tier 0 dz 0.62, 95% CI 0.27 to 1.03). Delta, theta, high beta and gamma were not significant (Table 6.3). This suggests that although several bands showed transient post-activity differences during the session, the most stable endpoint effects were concentrated in alpha and the lower beta range. Alpha was expressed as a widespread positive cluster, maximal over posterior regions and extending broadly into central and frontocentral sites, with a similar topography retained in the Tier 1 sensitivity analysis. Low beta was expressed as a broader positive cluster spanning frontocentral, central, parietal and posterior electrodes, with this topography largely retained in the Tier 1 sensitivity analysis. Broad beta was expressed as smaller positive clusters involving posterior-parietal and frontocentral sites, with a less spatially stable and more fragmented pattern in the Tier 1 sensitivity analysis. Pre-post change topographies for alpha, broad beta, low beta and high beta are shown in Figure 6.2.

Sensitivity analyses supported the same overall interpretation. Alpha remained the most robust and reproducible finding across tiers and tended to strengthen under the stricter quality-control threshold. Beta-family effects were also consistent and tended to strengthen in Tier 1. Theta and gamma effects remained broadly consistent with higher quality-control thresholds.

Complementary mixed-effects analyses of ROI-averaged post-activity band power indicated that alpha was primarily characterised by a progressive order-related increase across the session, whereas beta band differences were more activity-specific. Fronto-central alpha and posterior alpha both increased significantly with order in session (fronto-central: β 0.29, 95% CI 0.13 to 0.45, Holm p=0.003; posterior: β 0.39, 95% CI 0.20 to 0.57, Holm p=0.001), with no corrected order effects observed for beta band measures. No order × activity interaction survived correction for any outcome. Robust omnibus activity effects were observed for posterior low beta ($F(3,87)=5.82$, Holm p=0.009) and posterior broad beta ($F(3,87)=4.97$, Holm p=0.022). Holm-corrected pairwise contrasts indicated that posterior low-beta and broad-beta power were higher following the match-3 and multitasking activities than following the memory/sequence recall activity.

Table 6.3. Pre versus post-intervention spectral effect sizes and cluster-level p-values across primary and sensitivity quality-control tiers

Tier	n	Delta dz (p-value)	Theta dz (p-value)	Alpha dz (p-value)	Beta dz (p-value)	Low Beta dz (p-value)	High Beta dz (p-value)	Gamma dz (p-value)
Tier 0	23	-0.10 (0.375)	0.08 (0.175)	**0.75 (0.001)**	**0.44 (0.006)**	**0.62 (0.002)**	0.10 (0.124)	-0.29 (0.368)
Tier 1	21	-0.02 (0.129)	0.11 (0.214)	**0.76 (0.004)**	**0.58 (0.013)**	**0.65 (0.004)**	0.28 (0.085)	-0.18 (0.258)

Values show ROI-based Cohen's dz effect sizes and cluster-level p-values for within-subject rsEEG band-power differences pre versus post-intervention. n denotes the number of participants retained within each quality-control tier. Negative dz values indicate lower band power post-intervention. Significant values ($p<0.05$) are in bold. Frequency bands were defined as delta 1-3Hz, theta 4-7Hz, alpha 8-12Hz, broad beta 13-30Hz, low beta 13-20Hz, high beta 21-30Hz and gamma 31-40Hz.

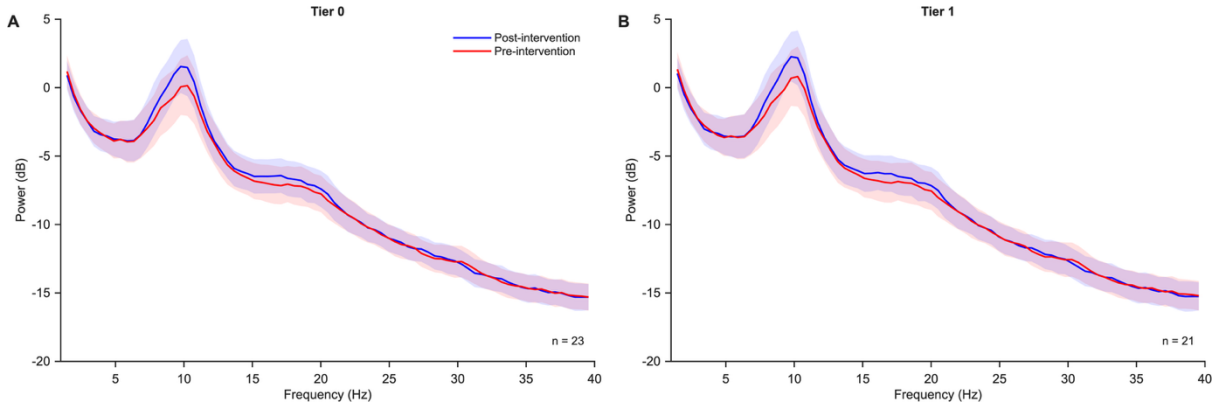


Figure 6.1. Pre and post-intervention resting-state power spectral density across quality-control tiers

Mean power spectral density (PSD), averaged across all channels, comparing pre and post-intervention. Shaded regions indicate 95% confidence intervals around the mean. Panels A and B show Tier 0 (full cohort) and Tier 1 (n=21) respectively.

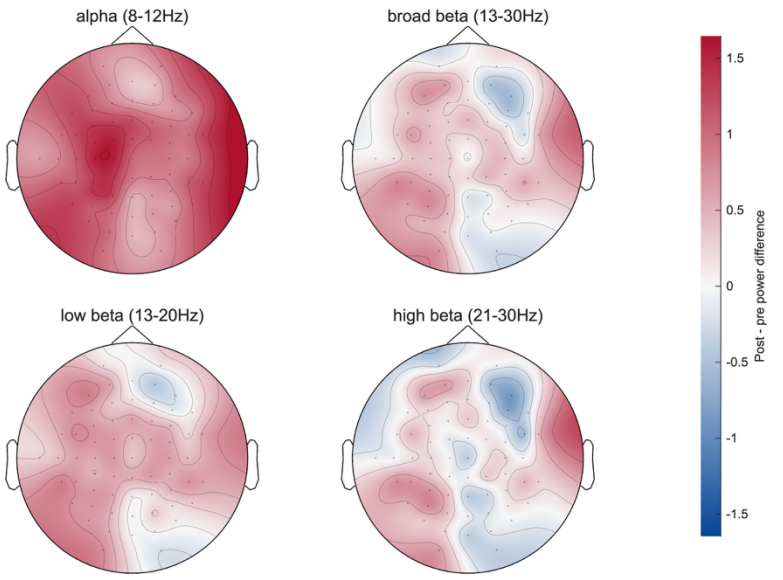


Figure 6.2. Pre-post intervention differences in alpha and beta band power topographies

Topographies show mean post-pre differences in conventional EEG band power across the scalp for alpha (8-12Hz), broad beta (13-30Hz), low beta (13-20Hz) and high beta (21-30Hz) in tier 0 (n=22). Red indicates higher band power following intervention relative to pre-intervention resting-state recordings, while blue indicates lower post-intervention band power. The maps are presented as descriptive spatial visualisations of the conventional spectral contrasts.

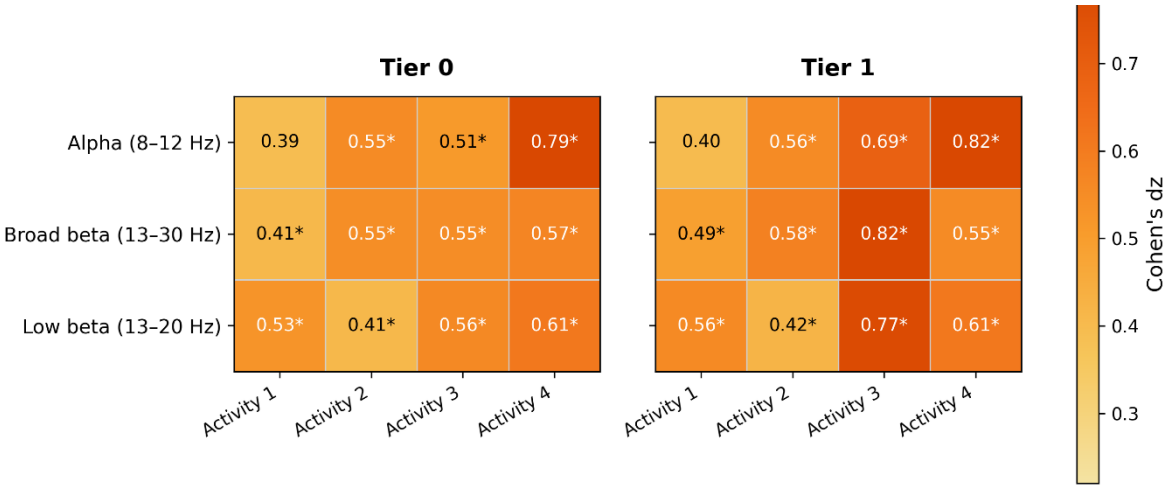


Figure 6.3. Repeated post-activity alpha, broad beta and low beta effect sizes across quality-control tiers

Heatmaps show within-subject Cohen's dz for resting-state spectral power comparing each post-activity recording to baseline across four sequential VR activities (Activities 1–4). Values represent ROI averaged effect sizes (Cohen's dz) corresponding to bands in which cluster-based permutation testing identified significant sensor-level effects. Asterisks denote cluster-level significance ($p<0.05$). Colour scale reflects effect size magnitude, truncated to the observed range.

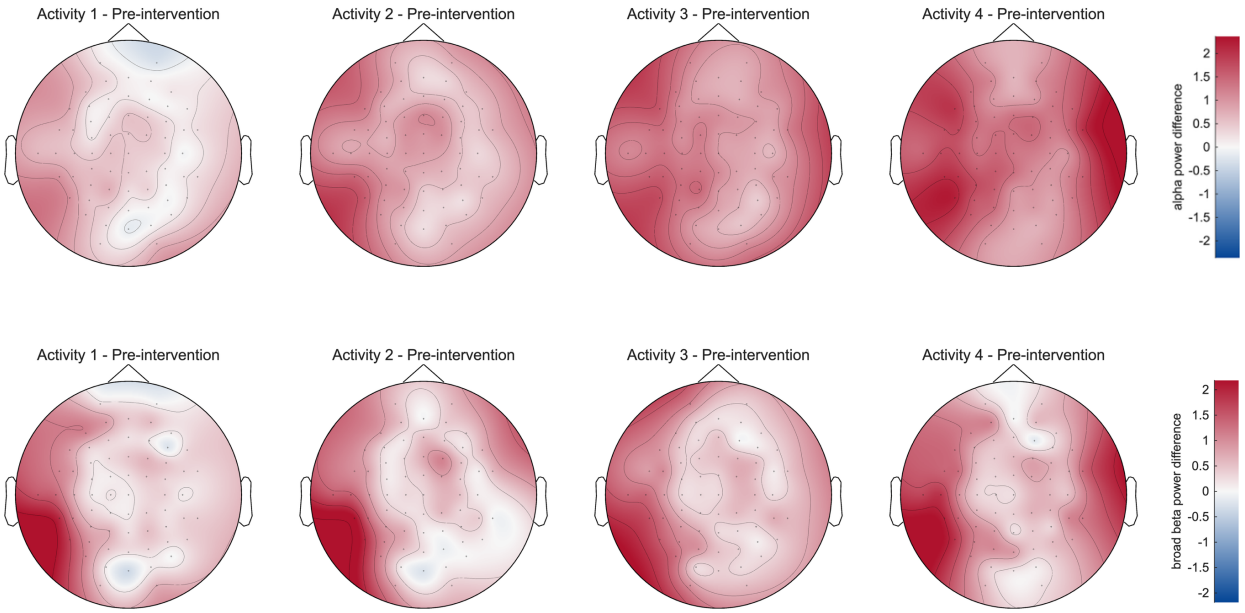


Figure 6.4. Alpha and beta band power topographies across sequential activity contrasts

Topographies show mean within-participant differences in conventional EEG band power for each activity relative to the pre-intervention resting-state recording. The upper row shows alpha-band power differences (8-12Hz), and the lower row shows broad beta band power differences (13-30Hz). Panels are ordered sequentially from Activity 1 to Activity 4. Red indicates higher band power following each activity relative to pre-intervention baseline, while blue indicates lower band power. The maps are presented as descriptive spatial visualisations of activity-related spectral changes.

6.4.2 Secondary aperiodic-periodic spectral parameterisation

In the pre- versus post-intervention comparison, neither the aperiodic exponent nor offset changed significantly (exponent: Holm-corrected p=1.000, dz 0.14, 95% CI -0.25 to 0.68; offset: Holm-corrected p=1.000, dz 0.13, 95% CI -0.29 to 0.69). Aperiodic-adjusted residual alpha power increased significantly (Holm-corrected p=0.006, dz 0.85, 95% CI 0.57 to 1.31, Figure 6.5). Figure 6.6 shows the corresponding group-average FOOOF decomposition, including the original and fitted spectra and the residual aperiodic-adjusted spectra used to quantify alpha-band power. Residual low beta also showed an increase at the uncorrected level (uncorrected p=0.012, dz 0.60, 95% CI 0.19 to 1.13), but this did not survive Holm correction (p=0.070). Residual broad beta showed a similar direction of effect without reaching statistical significance after correction (uncorrected p=0.021, dz 0.55, 95%

CI 0.10 to 1.27; Holm-corrected p=0.106). No significant change was observed in residual theta, high beta, or gamma.

Across the repeated post-activity contrasts, no consistent change was observed in aperiodic offset or exponent. Exponent showed small negative shifts across activities, with the largest effect after Activity 4 (uncorrected p=0.041, dz -0.48, 95% CI -1.22 to -0.03), but this did not survive correction (Holm-corrected p=0.081). Offset remained not statistically significant throughout. Residual oscillatory findings again pointed primarily toward later-session alpha-range enhancement. Residual alpha increased at the uncorrected level after Activity 3 (p=0.022, dz 0.54, 95% CI 0.21 to 0.91) and more strongly, surviving Holm-Bonferroni correction, after Activity 4 (uncorrected p=0.002, Holm-corrected p=0.020, dz 0.74, 95% CI 0.43 to 1.22). Other residual band effects were not statistically significant and inconsistent in direction and magnitude.

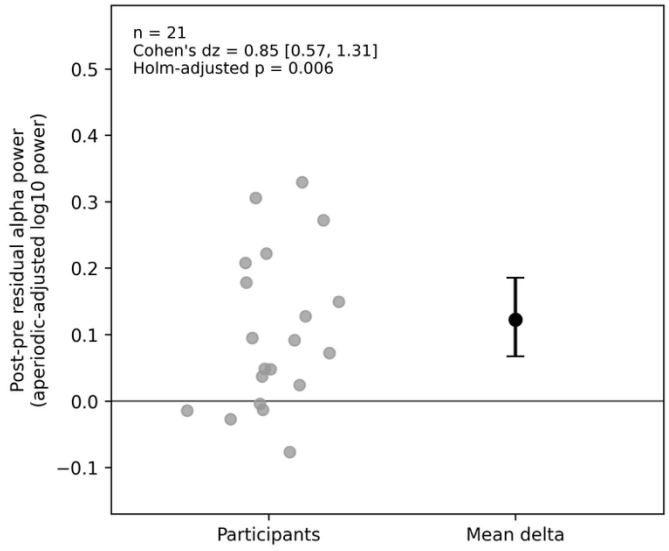


Figure 6.5. Pre-post increase in aperiodic-adjusted alpha power following VR intervention.

Grey points show individual participant post-pre differences in residual alpha power within the predefined frontocentral ROI. Residual alpha was defined as mean 8-12Hz log10 power after subtraction of the fitted aperiodic component. The black point shows the mean post-pre difference, with error bars representing the 95% bootstrap confidence interval. Positive values indicate higher aperiodic-adjusted alpha power after the intervention relative to the pre-intervention resting-state recording.

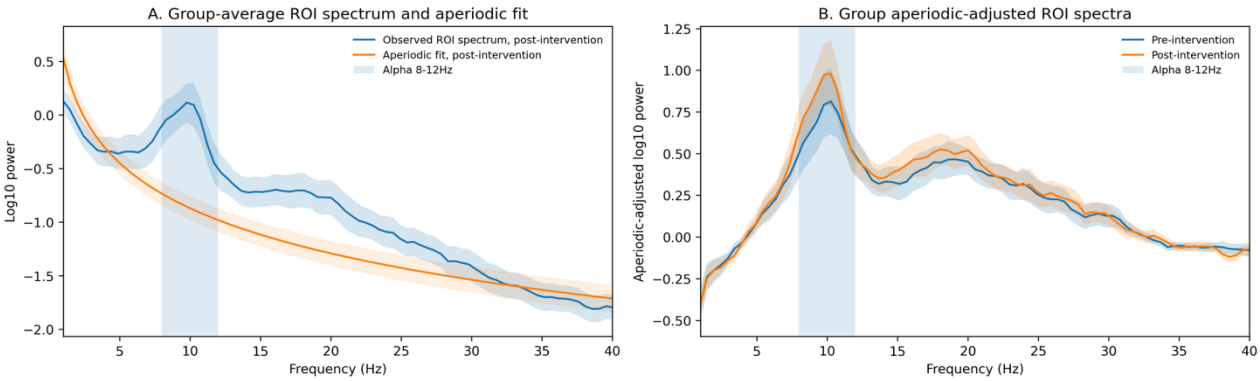


Figure 6.6. Aperiodic-adjusted frontocentral alpha power post-intervention

Mean frontocentral ROI spectra are shown before and after aperiodic adjustment. Panel A shows the mean ROI power spectrum and fitted aperiodic component, plotted in log10 power units. Panel B compares pre- and post-intervention spectra after subtraction of the fitted aperiodic component, shown as aperiodic-adjusted log10 power. The shaded vertical region indicates the predefined alpha band, 8-12Hz, used for residual alpha extraction. Shaded ribbons indicate 95% confidence intervals around the mean.

6.4.3 Secondary alpha peak analyses

For the occipital ROI, both IAF and alpha peak power increased following the intervention. In the Tier 0 pre versus post-intervention comparison, IAF showed a significant increase (mean difference 0.245Hz, p=0.018), while alpha peak power demonstrated a larger and more robust increase (mean difference 1.932dB, p=0.004). This pattern was robust to sensitivity analysis, with similar directional effects observed in Tier 1.

In contrast, the central ROI showed no evidence of a consistent change in IAF (Tier 0 mean difference 0.134Hz, p=0.352), indicating stability of peak frequency in sensorimotor regions. However, central alpha peak power increased significantly following the intervention (mean difference 1.654dB, p=0.002), with consistent direction and significance in Tier 1.

Across repeated post-activity contrasts, occipital IAF did not demonstrate consistent changes across activities, with all Tier 0 comparisons remaining not statistically significant. In contrast, occipital alpha peak power increased significantly following each activity (mean differences ranging from 1.485 to 2.554dB; p=0.010-0.001), with the largest effect observed following Activity 4. Central alpha peak power similarly

increased across all activities (mean differences 1.059-1.951dB; p=0.009-0.002), while central IAF again showed no consistent change.

6.4.4 EEG-clinical association analyses

In the primary repeated-measures models using current post-activity pain VAS as the outcome, fronto-central broad beta showed a small positive association with pain at the uncorrected level (β 0.24, 95% CI 0.00 to 0.48, p=0.048), whereas fronto-central low beta (β 0.16, 95% CI -0.10 to 0.41, p=0.220) and posterior alpha (β 0.09, 95% CI -0.15 to 0.32, p=0.475) were not significant. After Holm correction, none of the EEG predictors remained significant. Across these models, baseline pain VAS was a strong positive predictor of current post-activity pain (all $p<0.001$), whereas order in session and activity identity were not significant.

In lag-adjusted repeated-measures models additionally including previous post-activity pain VAS, no significant associations were observed between EEG measures and current pain ratings for any ROI-band outcome. Fronto-central broad beta and low beta showed small positive effect estimates, consistent in direction with the primary models, but these effects were attenuated and no longer approached statistical significance after adjustment for prior pain. Posterior alpha showed no evidence of association in either direction. Across all lag-adjusted models, previous pain VAS remained the strongest predictor of current pain ($p<0.001$), confirming substantial short-term autocorrelation within the repeated-measures pain series. Baseline pain VAS also remained a significant covariate, while neither order in session nor activity identity demonstrated significant effects after correction.

Participant-level pre-post models similarly provided no clear evidence that overall change in the principal EEG domains was associated with overall change in post-intervention pain VAS. Change in fronto-central broad beta (β 0.07, 95% CI -0.38 to 0.52, p=0.753), fronto-central low beta (β 0.07, 95% CI -0.38 to 0.52, p=0.761), and posterior alpha (β -0.26, 95% CI -0.72 to 0.19, p=0.240) were all not statistically significant, with no effect surviving Holm correction.

Similar null findings were observed for MPQ-SF total change for fronto-central broad beta (β 0.05, p=0.825) and fronto-central low beta (β -0.06, p=0.809). Posterior alpha showed the strongest participant-level association with MPQ-SF total change (β -0.42, 95% CI -0.85 to 0.01, p=0.057), but this did not survive Holm correction. In

additional exploratory posterior-alpha models, greater posterior alpha increase was associated with greater reduction in MPQ-SF sensory score at the uncorrected level (β -0.49, 95% CI -0.92 to -0.06, p=0.028), whereas no corresponding association was observed for MPQ-SF affective change (β -0.17, p=0.391).

6.5 Discussion

In this repeated-measures rsEEG study of participants with fibromyalgia undergoing an attention-based immersive VR intervention, the principal findings were increase in alpha and beta-family band power, particularly broad beta and low beta. Alpha showed the clearest progressive within-session pattern and the strongest support after aperiodic parameterisation, whereas broad beta and low beta showed consistent cluster-corrected conventional band-power increases but weaker residual evidence after separation of the aperiodic component. The final pre-post comparison demonstrated significant increases in alpha, broad beta and low beta, while delta, theta, high beta and gamma were not significant. Aperiodic exponent and offset did not change, arguing against a global shift in broadband spectral background. In the aperiodic-adjusted pre-post analysis, residual alpha remained significantly increased, whereas residual broad beta and low beta showed only nominal increases that did not survive correction. Alpha peak analyses further supported alpha enhancement, showing increased occipital and central alpha peak power with little evidence for a consistent shift in intrinsic alpha frequency. EEG-symptom association analyses did not identify robust corrected relationships between EEG signal change and pain, although posterior alpha showed a weak exploratory association with reduction in MPQ-SF sensory score. These findings suggest that immersive VR exposure in fibromyalgia is associated with measurable short-term modulation of resting-state spectral activity, with alpha providing the clearest evidence for a residual oscillatory change and beta-family effects representing consistent but less specifically resolved band-power changes.

This EEG secondary analysis extends findings from the parent VIPA feasibility study by showing that attention-based VR exposure was accompanied by measurable changes in resting-state neurophysiology. The parent study demonstrated that VR was acceptable and feasible in fibromyalgia, while also showing variability in response across systems, virtual environments and activities[8]. The present EEG analysis adds a mechanistic layer by demonstrating that resting-state spectral

features can change following a VR intervention designed for chronic pain. This is important because most therapeutic VR studies in chronic pain focus on feasibility, acceptability or symptom outcomes, with comparatively little evidence on whether immersive exposure is accompanied by changes in candidate neurophysiological domains[5, 6, 82]. The present results therefore provide preliminary indicators of EEG change to support future studies evaluating neurophysiological alteration with immersive therapeutics.

The present findings sit within a small but growing literature using EEG to study neurophysiological responses to VR. In spinal cord injury-related neuropathic pain, a small crossover study (n=15) recorded eyes-open EEG across no-task, 2D screen and 3D VR conditions and reported reduced relative alpha, increased low gamma and reduced self-reported pain[310]. In healthy participants, VR modulates alpha and beta responses during attention and emotional tasks, potentiating task-related alpha and beta desynchronisation relative to 2D presentation[311, 312]. Experimental pain work has similarly linked VR analgesia to oscillatory change during or around immersive exposure, including strategy-dependent gamma and beta effects[313]. The present study differs from much of this work by examining repeated eyes-closed rsEEG after VR activities in a fibromyalgia cohort, rather than acute task-state EEG during active immersion or experimental pain in healthy participants. This distinction matters because post-activity rsEEG may index residual or evolving neural state following immersive engagement rather than the immediate electrophysiological correlates of attention, immersion or analgesia during VR itself.

The alpha enhancement observed in this study is neurophysiologically meaningful because it was supported by convergent alpha peak findings and a spatial distribution consistent with established eyes-closed alpha activity. Alpha peak analyses showed increased alpha peak power across occipital and central ROIs with little evidence of a consistent shift in intrinsic alpha frequency, suggesting increased alpha prominence rather than a change in alpha timing. This is consistent with models in which alpha power reflects functional inhibition, sensory gating and regulation of cortical information flow rather than a passive resting rhythm[314, 315], and with evidence that alpha is sensitive to immersion, attentional direction and task difficulty in VR[311, 312]. In the transition from interactive, attentionally demanding VR tasks to eyes-closed rest, increased alpha may therefore reflect a short-term shift

toward reduced externally directed sensory engagement, altered attentional set or post-task recovery. The widespread but posterior-maximal alpha cluster is consistent with enhancement of typical eyes-closed alpha, while its central and frontocentral extension, together with the increased central alpha peak power, raises the possibility of sensorimotor alpha involvement, consistent with the mu rhythm as an alpha-range oscillation over sensorimotor cortex[316]. Scalp-level resting-state topography alone cannot, however, definitively separate mu activity from volume-conducted posterior alpha. These alpha effects are therefore best interpreted as short-term, state-responsive oscillatory modulation around VR exposure, not as evidence of disease normalisation, a direct marker of analgesia or a specific therapeutic mechanism.

Broad beta and low beta were also important spectral findings. Sensor-level cluster-based permutation testing showed significant positive broad beta and low beta clusters across all repeated post-activity contrasts and in the final pre-post comparison, retained in the Tier 1 sensitivity analysis, indicating that beta-family power changed consistently around VR exposure under conventional band-power analysis. This is biologically plausible because beta oscillations are linked to sensorimotor control, maintenance of the current cognitive or motor set, response selection and top-down regulation of ongoing processing[151] all of which are engaged by interactive VR tasks requiring visuospatial search, controller-based action, attentional updating and goal-directed responding. Beta is similarly context-sensitive in VR-EEG work[311-313], and more directly relevant to therapeutic VR, increased beta and gamma power have been reported during VR-guided meditation in cancer-related chronic pain, indicating beta sensitivity to the immediate immersive therapeutic state[305], and increases in alpha and beta power after VR-based bilateral arm training in upper-limb rehabilitation[317]. These findings support beta-family activity as a plausible intervention-sensitive signal, particularly in immersive tasks involving attention, action selection and sensorimotor engagement.

The absence of consistent aperiodic change is also important. Aperiodic exponent and offset did not change in the pre-post comparison or across repeated post-activity contrasts, indicating that the intervention was not associated with a global shift in broadband spectral background. This helps guard against overinterpreting conventional band-power changes that could otherwise arise from nonspecific

spectral slope or offset effects[157]. The persistence of residual alpha after aperiodic adjustment strengthens the interpretation that the alpha finding reflects true oscillatory activity rather than broadband change. By contrast, residual broad and low beta increases were attenuated and did not survive correction. This does not negate the consistency of beta band effects, but it does reduce confidence that these effects can be attributed specifically to isolated beta oscillatory modulation. The beta findings may therefore reflect a combination of oscillatory and broader spectral influences or may have been insufficiently powered to survive correction after spectral decomposition. This distinction reinforces the value of periodic-aperiodic separation in EEG biomarker studies, particularly when interpreting intervention-related changes in small repeated-measures datasets.

The temporal pattern across the VR session suggests that the neurophysiological response was not reducible to a single immediate post-task effect. Alpha increased progressively with order in session in both frontocentral and posterior ROIs, with the strongest post-activity effects emerging later in the session, particularly after the fourth activity. This pattern is compatible with literature linking alpha power to attentional gating and state regulation, as well as neurofeedback studies showing that alpha modulation can evolve across repeated training periods rather than only reflecting immediate task demand[314, 315, 318, 319]. Alpha enhancement may therefore reflect cumulative within-session exposure, post-task settling, or an evolving participant state, rather than the specific identity of the immediately preceding VR activity. By contrast, beta effects were present from the first activity and remained more consistent. Mixed-effects modelling of ROI-averaged post-activity power identified activity effects for posterior broad beta and posterior low beta, with higher power after the match-3 and multitasking activities than after the memory/sequence recall activity. This activity-sensitive pattern is consistent with literature linking beta activity to sustained visual attention, goal-directed processing, response selection, motor output and working-memory control[151, 320-323]. Match-3 and multitasking tasks plausibly required more continuous visuospatial search, controller-based sensorimotor coordination, response selection and updating of task set. Memory/sequence recall, although less movement-heavy, still plausibly recruited ordered encoding, anticipation and proactive control, but with less continuous visuomotor engagement[324, 325]. On this interpretation, beta-family activity, particularly

low beta, may be more sensitive to the attentional-sensorimotor implementation demands of specific VR activities, while alpha may better reflect a cumulative post-exposure shift in resting-state balance. This remains theoretical, because activity identity, order, fatigue, expectation, learning, novelty, relaxation and cumulative exposure cannot be fully disentangled in this small repeated-measures intervention study.

The final pre-post comparison should also be interpreted in relation to acquisition context. The four intermediate post-activity recordings were obtained with the headset in place and the display set to black, whereas the pre-intervention and final post-intervention recordings were both obtained without the headset. This makes the endpoint comparison useful as a session-level contrast, but not identical to the internally standardised headset-on post-activity sequence. The fact that alpha increased in both repeated post-activity contrasts and the final post-intervention contrast strengthens confidence that the alpha effect is not solely a headset-state artefact. Nevertheless, residual contextual differences remain possible. Future studies should consider including matched headset-on and headset-off resting baselines, or a non-VR control condition, to separate the effects of immersive exposure, headset context, task engagement and repeated rest more clearly.

The symptom association analyses show no clear associations between EEG and symptom reporting, including pain. Pain VAS and MPQ-SF scores decreased descriptively across the intervention session, but EEG changes were not robustly associated with pain outcomes after correction. In repeated-measures models, the weak association between frontocentral broad beta and current post-activity pain did not survive correction and was attenuated after adjustment for prior pain ratings, while previous pain VAS was the strongest predictor of current pain. Participant-level pre-post models similarly showed no corrected association between changes in rsEEG signals and changes in pain outcomes. The strongest exploratory signal was a negative association between posterior alpha increase and reduction in MPQ-SF sensory score, but this also did not survive correction. These findings should be interpreted in the context of a feasibility study with limited power for detecting robust EEG-clinical associations and relatively constrained symptom variability across individual activities. The absence of corrected EEG-symptom associations therefore does not exclude a relationship between neurophysiological modulation and clinical

response but indicates that such a relationship was not detectable in this dataset. Overall, the findings support physiological responsiveness to immersive VR exposure rather than proof of therapeutic mechanism or validation of rsEEG change as a surrogate endpoint for analgesia.

The study has several strengths. It used a within-subject repeated-measures design, allowing each participant to serve as their own control and increasing sensitivity to short-term neurophysiological change. EEG was acquired repeatedly across a structured VR session, enabling both sequential post-activity contrasts and a final pre-post comparison. The use of a 64-channel active gel-based system provided relatively high spatial sampling for a repeated-measures intervention study, and all recordings were processed using an automated reproducible pipeline aligned with the broader thesis methodology. The analysis also incorporated sensitivity testing, cluster-based permutation inference, ROI effect quantification, alpha peak analysis and aperiodic-periodic decomposition. These elements improve interpretability and reduce the risk that the main findings reflect isolated analytical choices.

Several limitations temper interpretation. The study was a secondary analysis of a feasibility dataset and did not include a non-VR control condition, sham intervention or randomised comparator. Of importance, the observed EEG changes cannot be attributed specifically to VR. The sample size was modest, particularly for mixed-effects activity modelling and EEG-clinical association analyses. The intervention session combined multiple activities, cumulative exposure and repeated measurement, making it difficult to isolate activity-specific mechanisms. Acquisition context differed between the intermediate headset-on post-activity recordings and the final headset-off post-intervention recording, which complicates direct interpretation of session-level endpoint effects. Retained post-clean EEG duration was shorter than in many conventional rsEEG studies, although quality-control procedures and sensitivity analyses were used to address this limitation. Finally, the cohort consisted of participants with fibromyalgia recruited into a feasibility study, and findings may not generalise to other chronic musculoskeletal pain conditions, more severe clinical populations or longer therapeutic programmes.

Future work should investigate whether the alpha and beta changes observed here can be replicated in larger controlled studies and whether they are specific to immersive VR rather than non-immersive attention tasks, relaxation, repeated rest or

expectancy effects. Designs including sham VR, non-immersive digital tasks and passive control conditions would help separate the contribution of immersion, interactivity and attentional engagement. Longer follow-up would clarify whether EEG changes are transient post-task phenomena or whether they persist with repeated therapeutic exposure. Future studies should also evaluate whether individual differences in baseline EEG, symptom burden, fatigue, cognitive state or VR tolerability predict neurophysiological response. For future closed-loop VR development, these findings are most useful as mechanistic context rather than as candidate real-time control signals. Alpha and beta changes may help characterise neurophysiological responsivity to immersive VR exposure, but further work is needed to determine their reliability, durability and relationship to therapeutic processes or clinical outcomes.

In conclusion, this study provides preliminary evidence that attention-based immersive VR exposure in fibromyalgia is associated with short-term modulation of rsEEG, principally involving increased alpha, broad beta and low beta power. Alpha showed the clearest evidence of a residual oscillatory change after aperiodic adjustment and was supported by increased alpha peak power across occipital and central regions. Broad beta and low beta were consistent conventional spectral findings across the VR session but were less robust after aperiodic separation and appeared more activity-sensitive than the progressive alpha response. There was no consistent evidence for a global aperiodic shift, and EEG changes were not robustly associated with pain change after correction. Overall, immersive VR exposure appears capable of altering candidate rsEEG domains in fibromyalgia, but these findings should be interpreted as preliminary evidence of neurophysiological responsivity rather than evidence of mechanism, durability, clinical efficacy or closed-loop control validity.

7. Study 3: The VR-FOCUS Study: Investigating eye tracking during a virtual reality N-back task as a predictor of cognitive load in chronic pain

This chapter presents the VR-FOCUS study, an experimental investigation of headset-integrated eye tracking during a graded VR N-back task in adults with and without chronic musculoskeletal pain. Ocular features are evaluated against

behavioural performance and subjective workload, and time-series machine learning classifiers are benchmarked against a conventional summary-feature baseline under subject-independent cross-validation. The study contributes the state-sensitive measurement and predictive inference components of the empirical work, with attention to whether cognitive load inference generalises across a clinically heterogeneous pain cohort.

Pupil diameter and blink rate increased systematically with task level and subjective workload in both groups under stable luminance. The principal finding concerned prediction. Time-series classifiers applied to continuous pupil signals discriminated low- from high-load task conditions well in healthy controls, with accuracies broadly comparable across classifier families, whereas the summary-feature baseline performed close to chance. Performance was attenuated and more variable in the chronic pain group, despite preserved group-level ocular responses to task demand, indicating that the temporal structure supporting inference was less consistent in this cohort. Pupil-derived signals were the most informative, with blink dynamics adding small reproducible gains only in the chronic pain group. The ROCKET classifier approached the performance of more complex ensembles at substantially lower computational cost. Cognitive load inference from embedded eye tracking was therefore feasible but cohort-sensitive, rather than uniformly generalisable across chronic pain and healthy control participants.

7.1 Introduction

Chronic musculoskeletal pain is heterogeneous between individuals and fluctuates within individuals across pain, fatigue, attentional capacity and cognitive effort[9-11, 36, 37, 61, 203]. A static immersive intervention is unlikely to remain optimally calibrated across this variation. Cognitive load is one candidate adaptive target. Excessive load may compromise tolerability, performance and engagement, whereas insufficient load may reduce attentional capture and limit therapeutic content delivery[6, 51, 55, 198, 199]. Inference of cognitive load during immersion would therefore offer a basis for dynamic adjustment of task demand, pacing or environmental complexity.

Eye tracking is an attractive candidate signal in this context. It is increasingly integrated within commercially available VR head-mounted displays and acquired continuously without additional wearable hardware[67, 166, 167]. Pupil size and blink

behaviour have well-established associations with cognitive effort in non-immersive paradigms[171, 173, 176, 180-183, 264], although their reliability and interpretability within VR, and within chronic pain populations in particular, remain insufficiently characterised. Chronic pain may alter attentional allocation, fatigue, medication exposure and ocular symptoms in ways that could affect both signal quality and the mapping between ocular features and inferred load[205-207]. Validation in the target population is therefore a precondition for use in future adaptive systems.

Conventional analyses of ocular workload signals rely on aggregate summary features such as mean pupil diameter or blink rate, which discard temporal structure and may underrepresent information available in continuously sampled streams. Time-series machine learning provides an alternative framework in which the temporal shape of the signal is treated as informative rather than as nuisance variation[211, 213]. Whether such methods offer incremental value over conventional approaches in deployable VR contexts remains underexplored, particularly in chronic pain populations where signal heterogeneity may degrade classifier performance. Subject-independent evaluation is essential, reflecting the operational reality that an adaptive system deployed in a new user is unlikely to have access to extensive within-user calibration data.

VR-FOCUS addresses these gaps using a consumer-grade VR system with embedded eye tracking and a bespoke graded N-back paradigm. Adults with and without chronic musculoskeletal pain completed the task under matched conditions, with eye tracking, behavioural performance and subjective workload acquired throughout. The N-back paradigm was chosen for its established use as a graded working-memory load manipulation under controlled environmental conditions[208-210]. Inclusion of a clinically heterogeneous chronic musculoskeletal pain cohort alongside healthy controls supported evaluation of whether the same inference framework remained informative in a population in which signal quality and feature stability cannot be assumed. Analysis combined a conventional aggregate-feature baseline with multiple time-series machine learning classifiers, evaluated under subject-independent cross-validation.

The objective of this study was therefore to determine whether a deployable embedded sensing modality, interpreted through time-series machine learning, can support inference of cognitive load during VR cognitively demanding tasks, and

whether that inference generalises uniformly or is cohort-sensitive across chronic pain and healthy control participants.

7.2 Study aims

The primary aim of this study was to determine whether eye tracking signals acquired during a VR N-back task could support inference of cognitive load in both healthy controls and individuals with chronic musculoskeletal pain.

A secondary aim was to compare conventional summary-feature approaches with temporally resolved machine learning approaches for the prediction of cognitive load from eye tracking data, to determine whether temporally sensitive models provide meaningful advantages for state inference in immersive environments.

An additional aim was to examine whether model performance differed between healthy controls and individuals with chronic musculoskeletal pain, thereby assessing whether cognitive load inference from eye tracking was equally generalisable across groups.

A further exploratory aim was to characterise the relative contribution of pupil-centred and blink-related features to cognitive load classification within VR, with particular interest in whether blink-derived features provided incremental value beyond pupil signals alone.

7.3 Methods

7.3.1 Study design and participants

This study employed an experimental cross-sectional design in which adults with and without chronic musculoskeletal pain completed a single face-to-face data collection session.

Participants were recruited into two groups: healthy controls and individuals living with chronic musculoskeletal pain. Healthy controls were recruited purposively from the UEA community, restricted to current students and staff within Norwich Medical School and the School of Health Sciences. Recruitment routes included standardised email invitations distributed through administrative channels, physical and digital advertisements, and brief promotion during teaching sessions or academic meetings with prior gatekeeper approval.

Participants with chronic musculoskeletal pain were recruited purposively through two existing UEA research databases: the Norfolk Arthritis Register (NOAR) and the VIPA study fibromyalgia database. For recruitment through NOAR, individuals likely to be living with chronic pain were first identified through database filtering procedures using pain and disability-related variables in addition to relevant diagnostic fields, in order to generate a list of potentially eligible participants for contact. These individuals were then approached by telephone to establish interest and to confirm they had experienced pain on most days for more than 3 months. Those expressing interest and fulfilling this criterion were directed to the VR-FOCUS study website for full screening, participant information, and consent procedures. For recruitment through the VIPA study database, individuals were contacted by email and directed to the study website. This staged approach allowed database-assisted identification of likely eligible chronic pain participants while preserving a common downstream screening and consent pathway across recruitment sources.

Initial screening and consent were conducted through the VR-FOCUS study website electronically. Interested individuals were directed to the online participant information sheet (Appendix 7), eligibility screening questions, and electronic consent form (Appendix 6). Eligible participants who completed online consent were asked to provide an email address, after which they were assigned a unique participant identifier and contacted by the research team to arrange an in-person study session. This process standardised screening and consent across participant groups while preserving anonymity within the research dataset.

General inclusion criteria for both groups were age 18 years or older, conversational English, and capacity to provide informed consent. For the chronic pain group, inclusion additionally required participants to have chronic pain characterised by pain experienced for 3 months or more. Exclusion criteria were aligned across groups and aimed to minimise both participant risk and interpretive confounding in ocular measurement. Participants were excluded if they had any co-morbid condition exacerbated by exposure to flashing lights or screens, current visual or hearing impairment preventing use of the VR system, diagnosed cognitive impairment, current facial injury or other condition making VR use uncomfortable or impractical, or any co-morbid condition known to significantly affect eye movements or pupil

reactivity. Healthy controls were additionally excluded if they reported acute or chronic pain at the time of recruitment.

Each participant attended a single in-person testing session located in a quiet, private testing room at the UEA. The two-group design served a specific translational purpose within the thesis. Inclusion of healthy controls allowed evaluation of whether eye tracking signals varied systematically with graded task demand under relatively unconstrained physiological conditions, whereas inclusion of a broader chronic musculoskeletal pain cohort enabled assessment of whether the same state-inference framework remained informative in a clinically heterogeneous population in whom pain, fatigue, comorbidity, medication use, and other sources of variability might plausibly affect signal quality or classifier performance. The design therefore supported both proof-of-principle testing of cognitive load inference and preliminary evaluation of its generalisability to a chronic pain context.

7.3.2 Ethics and governance

Ethical approval for the VR-FOCUS study was obtained from the University of East Anglia (UEA) Faculty of Medicine and Health Research Ethics Sub-committee prior to study initiation (REC reference: ETH2324-2498, Appendix 5). All study procedures, including participant recruitment, data collection, and analysis, were conducted in accordance with this approval and in line with the principles of the Declaration of Helsinki.

The study involved prospective collection of behavioural, ocular, and questionnaire data from human participants during immersive virtual reality exposure. Particular consideration was given to potential risks associated with VR use, including motion discomfort, visual strain, and sensory sensitivity. The protocol incorporated predefined exclusion criteria to minimise risk, and participants were monitored throughout testing with the option to pause or withdraw at any stage without consequence.

Recruitment through the VIPA and NOAR databases was limited to individuals who had previously provided formal consent to be re-contacted about future ethically approved research. This ensured that database-assisted recruitment was conducted within the scope of prior participant permission and institutional governance procedures.

All participants provided written informed consent prior to enrolment. Consent procedures were conducted electronically via the study website following provision of a detailed participant information sheet (Appendix 7), and were reconfirmed in person at the start of the testing session. Participants were assigned unique study identifiers to ensure anonymity, and no directly identifiable information was included in the analytical dataset.

Data governance procedures were designed to ensure secure handling of both identifiable and de-identified data. Identifiable information used for scheduling and communication was stored separately from research data and accessible only to authorised members of the research team. All research data, including eye tracking recordings, behavioural task outputs, and questionnaire responses, were stored on secure UEA systems in accordance with institutional data protection policies and General Data Protection Regulation (GDPR) requirements.

7.3.3 VR apparatus and implementation

The hardware, software and calibration framework for VR-FOCUS is described in Chapter 4 (Section 4.2). Within the present study, all data were collected using a single Pico Neo 3 Pro Eye headset, deployed offline with task and eye tracking data written directly to device storage and headset output mirrored to a research laptop via USB-C for real-time researcher observation. Participants remained seated throughout testing and used a handheld controller for response input. Before task commencement, headset fit and interpupillary distance were adjusted individually, after which the standardised five-point eye tracking calibration was performed using the Tobii Ocumen workflow until predefined bias and precision criteria were met (Chapter 4, Section 4.2.1).

7.3.4 VR N-back task design and virtual environment

A bespoke graded VR N-back task was developed with five conditions: Baseline, 1-Back, 2-Back, 3-Back, and 4-Back. The conceptual rationale for the N-back paradigm and its suitability as a working-memory load manipulation is described in Chapter 4 (4.2.2). The Baseline condition functioned as a minimal-load control without target-matching demands, while the 1- to 4-Back conditions provided progressively greater active working-memory challenge. A load range up to 4-Back was selected to provide a broad graded manipulation while remaining feasible and

interpretable within a brief VR protocol for a pragmatic chronic pain cohort. Although higher-load variants are reported in the experimental N-back literature, task demands increase substantially with increasing N, and the present design prioritised a balance between cognitive challenge, tolerability and stable repeated sampling.

Each task block presented 40 letter stimuli. This was a pragmatic design choice rather than a fixed N-back convention, intended to provide sufficient repeated observations for behavioural, pupillometric and time-series modelling analyses while helping to keep total VR exposure below 20 minutes. This duration cap was informed by prior feasibility work within the broader programme[8] and consistent with guidance suggesting that shorter VR sessions may help mitigate cybersickness and adaptation effects[326]. The use of letter stimuli supported a simple and familiar working-memory structure that did not require complex semantic interpretation or visuospatial navigation, allowed difficulty to be modulated primarily through memory demands rather than scene content, and supported repeated ocular sampling across many trials within each task block.

For each active N-back block, 10 target matches were inserted, corresponding to 25% of total trials. A target proportion of 25% has precedent in N-back work[262] and was selected here to provide enough target events for stable performance estimation without making the task overly predictable[327]. Lures were also incorporated into the higher-load conditions. A lure was defined as a letter appearing one step before a true match would ordinarily occur. The rationale for including lures was methodological rather than simply to increase difficulty. Lures increase interference and false-alarm pressure[328], allowing task performance to be characterised using signal-detection measures such as discrimination index (d′) and response bias (c) rather than relying on raw accuracy alone[327]. A modest lure proportion of approximately 15-20% of stimuli was used, aimed to increase interference without allowing lure events to dominate the sequence structure.

Letter sequences were generated algorithmically within the application rather than being fixed in advance. Sequence generation was random but constrained so that the required number of true N-back matches was guaranteed and lure stimuli were inserted according to predefined task logic. As a result, repeated completion of the same task level did not present an identical sequence, while preserving the intended target and lure counts for that condition. This was an intentional design choice so

that if a participant had to repeat a task, or if a repeated-task paradigm were created within future work, identical task sequences would not bias results.

Stimulus timing was selected as a compromise between established N-back practice, participant tolerability and downstream analytic utility. Each letter was displayed for 750ms, followed by a 2000ms inter-trial interval. In the broader N-back literature, stimuli are commonly presented for around 500ms[329] followed by a delay or interstimulus interval of one to several seconds, with 2000ms intervals also reported[330]. The slightly longer 750ms display duration used here preserved readability and comfort in-headset, particularly in a cohort that included people with chronic pain and other comorbidities, while remaining within the timing range used in prior N-back work.

Each task began with a 10s fixation cross. Pre-task fixation periods are methodologically important in pupillometry because they help stabilise gaze, reduce uncontrolled eye-position variability, and provide a local reference against which subsequent task-evoked pupil changes can be interpreted[174, 175]. In the present study, the fixation period also served a pragmatic purpose by providing a brief baseline immediately before each task block within a repeated multi-block VR design, allowing ocular activity to settle after task transitions while avoiding a substantially longer pre-task interval that might increase visual fatigue or reduce tolerability.

Participants responded using a single trigger button on a handheld controller when they detected a stimulus identified as a match. If a letter was present on screen at the time of response, that letter was underlined to provide input feedback, but no accuracy feedback was given. The absence of trial-by-trial accuracy feedback was intentional, because explicit performance feedback could influence subjective workload ratings, emotional response or task strategy across blocks, thereby introducing variance unrelated to the intended cognitive load manipulation.

Task ordering balanced standardisation and control of order effects. Baseline and 1-Back were always completed first, ensuring that participants initially encountered the least demanding conditions to support task understanding before progressing to more difficult tasks. The 2-, 3- and 4-Back conditions were then administered according to one of six pre-defined counterbalanced orders, reducing the risk that systematic sequencing effects would be confounded with task difficulty while

preserving a predictable early structure that supported familiarisation with the task and controller response format. Before each task block, participants reviewed a short in-headset video tutorial to ensure understanding of the current task rules and response requirements, minimising the risk that early behavioural or ocular responses would reflect misunderstanding rather than the intended cognitive load manipulation.

The task was delivered within a deliberately constrained virtual environment designed to minimise unnecessary visual variability across task blocks. The surrounding space was neutral grey with a darker floor, and all task instructions and letter stimuli were presented in dull white text on a darker grey rectangular panel positioned centrally in the participant's field of view. There was no requirement for locomotion or deliberate body movement within the scene. Participants remained seated, and the task display remained spatially stable throughout each block.

Frame-level luminance was logged using the standard linear RGB-to-luminance transform described in Chapter 4 (Section 4.2.3). Mean luminance, standard deviation and coefficient of variation were derived for each task block. Luminance remained highly stable across task conditions, supporting later interpretation of pupil-diameter changes as primarily task-related rather than driven by scene-level luminance differences.

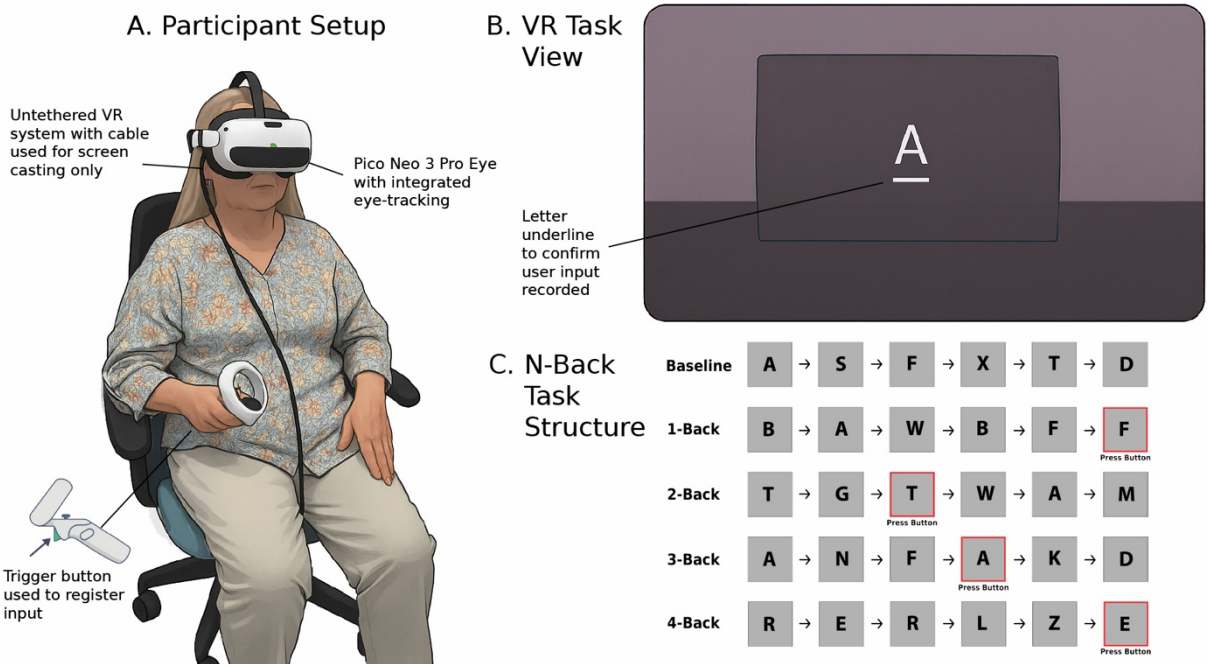


Figure 7.1. Overview of VR-FOCUS experimental setup, in-headset task display, and N-back procedure

C. The baseline task acted as a control condition with minimal load, the 1-Back task was the least difficult active task and the 4-Back task was the most difficult active task.

7.3.5 Data logging

Data capture and logging followed the framework described in Chapter 4. Within VR-FOCUS, raw eye tracking exports were retained in .ocumen format alongside behavioural task logs, calibration summaries, luminance records and questionnaire outputs. All files were written directly to headset storage and extracted after testing, with subsequent scripted generation of participant-level derivative datasets rather than manual consolidation.

7.3.6 Experimental procedure

Prior to VR use, participants completed the baseline questionnaires and were briefed on the equipment and tasks. This included being shown short tutorial videos explaining each task. Understanding was confirmed verbally before using VR.

After putting the headset on, participants first completed the standardised eye tracking calibration. They were given the option to refresh their memory of a task by watching the short video tutorial again in VR before starting. Each task began with a

10s fixation cross to collect that task's baseline eye tracking metrics and give a period of time for the pupil to stabilise. Participants pressed the controller trigger when they detected a match. If a letter was present on screen at the time, that letter underlined to provide input feedback (Figure 7.1, Panel B). No accuracy feedback was given.

Tasks were performed in a fixed order for the first two conditions (Baseline, 1-Back), followed by a pre-randomised allocation (counterbalanced, six pre-defined task orders) for the order in which they completed the 2-, 3-, and 4-Back tasks. This was to account for the risk of order effect on task difficulty, pupil metrics and cognitive load.

After each task, participants completed brief VR questionnaires using the handheld controller. Following all tasks, participants removed the VR system and completed the Virtual Reality Sickness Questionnaire (VRSQ) on paper prior to debrief.

7.3.7 Measures

A comprehensive set of subjective, behavioural and physiological measures was collected during a single in-person session to characterise baseline health status, task performance, cognitive load and ocular responses. All questionnaires were administered in English. Digital forms were used for all measures except the Pain Catastrophising Scale, which was completed on paper.

Participants first completed a baseline characteristics questionnaire that captured demographic information, previous exposure to technology, medical comorbidities and regular medications. Individuals with chronic musculoskeletal pain additionally completed validated patient-reported outcome measures to quantify symptom severity and functional impact. These comprised the PROMIS Pain Intensity (Short Form 3a)[331] and PROMIS Pain Interference (Short Form 8a)[331] instruments, providing structured assessments of pain severity and its disruption to day-to-day activity. They also completed the Pain Catastrophising Scale[332] and provided single-item numeric ratings of current pain and fatigue intensity on a 0–10 scale. Together, these assessments enabled detailed characterisation of the chronic pain cohort and provided data on potential covariates for subsequent analyses.

Task performance was recorded automatically for each N-back level except Baseline. For each stimulus sequence, the system logged the number of correctly

identified targets, those responded to that were incorrect, false alarms to lure stimuli, and reaction times. These raw metrics were used to derive discrimination index (d′; a participant's ability to distinguish targets from non-targets) and response bias (c; a participant's tendency to respond more liberally or conservatively) according to standard signal-detection conventions[210, 327]. These measures allowed quantification of working-memory performance and supported analyses of how behaviour related to subjective load and physiological responses.

Continuous eye tracking data were captured during every task block, with timestamped pupil diameter and gaze direction in three-dimensional space recorded at 60-90Hz for both eyes and written directly to device storage. Missing samples arising from blinks were processed according to predefined temporal criteria as described in the Data Preprocessing section. These ocular metrics comprised the primary physiological measures used to investigate cognitive load, evaluate relationships with subjective and behavioural metrics, and train time-series classifiers.

After completing each N-back block, participants remained in the VR environment and completed a brief set of subjective ratings using a continuous sliding scale. Cognitive workload was assessed using the NASA Task Load Index[333], which includes six dimensions rated from 0 to 100, and an adapted version of the one-item PAAS scale[334], recorded on a 0–9 scale in VR and subsequently rescaled to align with conventional scoring (1-9). Participants also rated current pain and fatigue intensity using blinded 0–100 visual analogue scales. A seven-item subjective experience questionnaire developed for this study assessed immersion, comfort, anxiety, ability to maintain attention, difficulty focusing, motivation and engagement. These additional measures provided contextual information about users' experiences and potential sources of variance in eye tracking or performance data.

Immediately after removal of the headset, participants completed the Virtual Reality Sickness Questionnaire[335]. This nine-item measure quantified the presence and severity of VR side effects, including disorientation, nausea and oculomotor discomfort. These data allowed assessment of tolerability of the VR protocol and examination of group-level differences in adverse effects.

Measure selection was guided by the need to characterise a clinically heterogeneous pain cohort, capture factors that could plausibly affect ocular signals or task performance, and keep in-headset assessment brief enough for repeated administration after each task block. PROMIS short forms were selected because they are calibrated using item response theory, are applicable across musculoskeletal diagnoses and generate T-scores referenced to the US general population, allowing participants with inflammatory arthritis, osteoarthritis and fibromyalgia to be described using a common metric with relatively low burden, while recognising that these reference norms may not fully reflect a UK population[331, 336]. Pain intensity and interference were both assessed because they represent distinct core outcome domains[297]. The PCS was included because catastrophising has been associated with pain-related attentional bias and poorer cognitive performance in the presence of pain[106], making it a plausible contributor to variation in both task performance and ocular responses. Current pain and fatigue were also recorded because both may influence cognitive performance and ocular behaviour at the time of testing independently of task demand.

Two complementary measures were used to assess cognitive workload. The NASA-TLX is a widely used multidimensional workload instrument that distinguishes mental demand from effort, perceived performance and frustration, while its unweighted form reduces administration burden and has shown comparable performance to the weighted version[333, 337]. The single-item PAAS scale provided a rapid measure of invested mental effort suitable for repeated administration and sensitive to variation in cognitive load[334, 338]. The VRSQ was selected in preference to the Simulator Sickness Questionnaire because it was developed specifically for VR-related symptoms[335], while its use in the parent feasibility study[8] also enabled tolerability to be considered consistently across studies. Bespoke experience items captured immersion, comfort and engagement, for which no single brief validated instrument was suited to repeated in-headset administration within this protocol. Standardised measures of depressive and anxiety symptoms were not included in order to limit pre-task burden, which restricts interpretation of mood as a potential contributor to group differences.

Appendix 2 provides further detail on all measures used.

7.3.8 Sample size and power

A formal a priori sample size calculation was not undertaken. Recruitment was determined pragmatically by feasibility within the available study period, while aiming to provide sufficient independent participants for subject-level machine learning evaluation. The final analysed sample comprised 84 participants, including 42 healthy controls and 42 individuals with chronic musculoskeletal pain.

Post hoc sensitivity estimates were calculated to contextualise, rather than justify, the final sample size. Under a simplified two-tailed paired-comparison framework with $\alpha=0.05$, a sample of 84 would provide 80% power to detect a within-subject effect of approximately Cohen's $dz=0.31$. Within each subgroup separately, a sample of 42 would provide 80% power to detect a within-subject effect of approximately $dz=0.45$. These estimates should be interpreted cautiously, because the main analytical framework was not limited to simple paired comparisons and included task-level mixed-effects models, exploratory group comparisons, and subject-independent machine learning analyses.

For the predictive analyses, the key constraint was the number of independent participants rather than the total number of repeated observations. Under grouped leave-one-subject-out cross-validation, all observations from a given participant were withheld together at each iteration, meaning that effective sample size for subject-independent generalisation was determined by participant number. The sample was therefore considered suitable for feasibility testing of cognitive-load classification and for detecting moderate repeated-measures effects, but not for providing definitive population-level effect estimates or precise estimates of out-of-sample classifier performance.

7.3.9 Data preprocessing and feature extraction

Eye tracking preprocessing followed the framework described in Chapter 4 (Section 4.2.4). All study-specific preprocessing parameters were applied as specified in Chapter 4, including 90Hz to 60Hz nearest-neighbour downsampling, blink duration criteria (50-500ms), the local SD-based interpolation mask with reference windows at 100ms before blink onset and 150ms after blink offset, and cubic spline reconstruction with linear fallback. Both summary-feature and time-series representations were generated, with summary features comprising mean pupil

diameter and blink rate at task-block level, and time-series representations retaining pupil trajectories together with a binary blink-indicator channel for machine learning analyses.

Additional derived variables were created from behavioural and questionnaire data. Unweighted NASA-TLX scores were used. Behavioural summary variables included discrimination index (d′) and response bias (c) calculated according to standard signal-detection conventions (Appendix 2)[210, 327]. These derived features were combined with the ocular data in analysis-ready datasets for conventional statistical modelling and time-series classification.

7.3.10 Processed eye tracking data quality

Quality of the eye tracking dataset was evaluated using structured audit outputs generated during preprocessing. These metrics were used to characterise eye tracking calibration quality, signal completeness, interpolation burden, and overall usability for downstream statistical and machine learning analyses.

Calibration metrics were available for all 84 participants, comprising 142 calibration attempts in total. Final calibration was achieved on the first attempt for 69% of participants (n=58) with a mean of 1.69 attempts per participant (median 1, range 1-10). Using the predefined in-study calibration criteria of bias ≤3°, precision ≤1°, and valid samples at all calibration points, 95.2% of participants (n=80) met threshold on their final attempt. On the final calibration, mean precision error across participants was 0.25° ±0.13, with a mean maximum precision of 0.61°, while mean bias error was 0.62° ±0.27, with a mean maximum bias of 1.21°. Here, precision reflects the consistency or repeatability of gaze samples, whereas bias reflects the average spatial offset from the intended calibration target. These values indicate generally stable binocular tracking quality within the standalone VR system. Additional descriptive review showed that 20.2% of participants (n=17) wore glasses during VR and a further 9.5% (n=8) were usual glasses wearers who completed VR without them. Wearing glasses during VR was not associated with more calibration attempts or poorer final precision, although mean final bias was modestly higher (mean bias 0.76° vs 0.58°; mean maximum bias 1.54° vs 1.13°). However, final calibration pass rates remained high in both those who wore glasses in VR (94.1%) and those who did not (95.5%). For the four participants who did not meet all predefined calibration

thresholds, calibration quality was reviewed visually at the point of acquisition prior to task commencement and judged acceptable overall, with deviation limited to a minimal domain that was not considered sufficient to compromise tracking quality for the study outcomes. These participants were therefore allowed to proceed.

Across the analysed dataset, 61.9% of participants (n=52) were recorded at 60Hz and 38.1% (n=32) at 90Hz, with 90Hz recordings downsampled to 60Hz prior to analysis as described in Chapter 4. Blink interpolation burden across participants was modest but variable, with a mean of 7.37% ± 4.27 of task time reconstructed by interpolation (median 6.00%, IQR 4.35-10.00, range 0.59-21.49). Longer missing-data segments exceeding 500ms were uncommon overall, contributing a mean of 0.14% ± 0.32 of total task duration per participant (median 0.00%, IQR 0.00-0.16, range 0.00-2.32), although 36.9% of participants (n=31) had at least one such segment. No participant data were excluded, and no task blocks were removed from the analysed dataset. Beyond blink-related interpolation and infrequent longer deleted missing-data segments, no additional recurring source of unusable pupil data was identified during preprocessing. Taken together, these findings indicate that both calibration quality and processed pupil-data quality were stable across the cohort, with expected pragmatic variability related primarily to repeated headset fitting and blink-associated loss rather than widespread failure of ocular signal acquisition. Visual inspection of preprocessed pupil time series confirmed preservation of physiologically plausible signal structure, including expected peri-stimulus dynamics and absence of sustained artefactual drift.

7.3.11 Inferential framework

The inferential strategy was structured around two complementary levels of analysis. First, conventional descriptive analyses were used to characterise behavioural, ocular, and subjective-workload measures across task-defined cognitive load conditions and between groups. Linear mixed-effects models were then used to evaluate task-related and subjective workload-related changes in eye tracking outcomes, while also assessing group effects and group-by-task interactions. In this framework, task level (Baseline, 1-Back, 2-Back, 3-Back, 4-Back) was treated as a within-subject factor and group (healthy control versus chronic pain) as a between-subject factor, with participant ID included as a random effect to account for repeated observations within individuals. Second, predictive analyses addressed a distinct

inferential question, namely whether cognitive load state could be classified in unseen participants from eye tracking data under subject-level separation. In line with the framework set out in Chapter 4, the primary predictive endpoint emphasised subject-independent binary low versus high-load contrasts, particularly 1-Back versus 4-Back, with broader multiclass analyses treated as secondary.

7.3.12 Statistical analysis

Conventional analyses were conducted in R (v4.4.3)[339]. Continuous variables were assessed for normality using the Shapiro–Wilk test and summarised as mean ± SD for normal or median [IQR] for non-normal data. Categorical variables were reported as frequencies and percentages. Associations between pupil diameter, blink rate, and cognitive-load scores (NASA-TLX, PAAS) were evaluated using linear mixed-effects models with participant ID as a random effect. Models adjusted for group, age, task order, sampling resolution, and fatigue, with exploratory analyses including anxiety, engagement, perceived difficulty, attention consistency, and performance indices (d′, c). Effect size was reported using β-coefficient with 95% confidence intervals. Model diagnostics were inspected visually. All tests of statistical significance were two-tailed with $\alpha=0.05$ and p-values reported where appropriate.

7.3.13 Predictive models

Predictive modelling followed the framework described in Chapter 4 (Section 4.3). Time-series classification was implemented in Python (v3.12)[340] using the AEON toolkit (v1.1)[213]. The same preprocessing pipeline was applied uniformly to both healthy controls and chronic pain participants to preserve methodological symmetry across groups, with all preprocessing parameters (winsorisation, z-scoring, mean imputation) estimated on training data only and applied unchanged to the held-out participant in each fold to avoid information leakage.

The logistic regression baseline used pre-specified task-level summary features comprising mean left pupil diameter and blink rate, with no automated feature selection, interaction terms or non-linear transformations applied. The primary multivariate feature set for time-series classifiers comprised left-eye pupil diameter and a binary blink-indicator channel. Default AEON hyperparameters were used, with stochastic components initialised using fixed random seeds for reproducibility. For

computational feasibility, the maximum training time for HIVE-COTE 2.0 was contracted to eight minutes per cross-validation fold.

Principal classification analyses focused on the binary 1-Back versus 4-Back contrast, conducted separately within healthy controls and chronic pain groups, with N-back levels used as class labels. Performance was reported using pooled out-of-fold metrics as specified in Chapter 4 (Section 4.3.3), including accuracy with Wilson 95% confidence intervals, alongside sensitivity, specificity and confusion matrices. ROC AUC was computed only for classifiers producing probabilistic outputs; ROCKET, which does not generate class probabilities, was summarised using accuracy, sensitivity, specificity and confusion matrices alone.

To assess whether age contributed to between-group differences in classifier performance, exploratory linear regression was used to model participant-level classification accuracy as a function of age and group, with an age-by-group interaction tested. Regression coefficients (β) and two-tailed p-values are reported.

Computational tractability was assessed pragmatically alongside predictive performance. Runtime was not prospectively benchmarked for all models; however, fold-level runtime logs were available for most classifiers. Where available, runtime was recorded as seconds per cross-validation fold and summarised across grouped LOSO folds, with reported runtimes averaged across the corresponding healthy control and chronic pain runs to provide a single comparative estimate per classifier. All timed analyses were run locally on a MacBook Pro with Apple M5 processor, 10-core CPU, 24 GB RAM, running macOS Tahoe 26.3.1. These runtime estimates were therefore treated as pragmatic local execution benchmarks rather than formal hardware-independent measures of computational complexity.

Exploratory time-series regression of continuous subjective workload used ROCKETRegressor[274] as described in Chapter 4 (Section 4.3.2). Each case represented one participant completing one task level, with the pupil time series for that task as the input and the corresponding NASA-TLX total score as the regression target. Models were run separately in healthy controls and chronic pain participants using grouped LOSO cross-validation, ensuring that all task-level observations from the held-out participant were excluded from training. The input feature used across all task levels and participants was left pupil diameter. Performance was summarised

using Pearson correlation between predicted and observed NASA-TLX scores and mean absolute error (MAE) in NASA-TLX units, both reported as mean ±SD across cross-validation folds.

7.4 Results

A total of 84 participants were included with equal numbers in the healthy control and chronic pain groups (n=42 per group). Table 7.1 summarises participant characteristics between groups. For those with previous VR exposure (n=56, 67%), most reported rare or one-off use (n=47, 84%).

Most participants in the chronic pain group had complex multimorbidity, with 93% (n=39) reporting two or more long-term conditions and 95% (n=40) taking at least two regular medications. The most common diagnoses were inflammatory arthritis such as rheumatoid or psoriatic Arthritis (n=35, 83%), osteoarthritis (n=21, 50%), fibromyalgia (n=12, 29%) and autoimmune connective tissue disease (n=6, 14%). Over a quarter of participants (26%, n=11) reported an eye condition such as dry eyes or Sjogren's syndrome. Opioid use was reported in 45% (oral morphine equivalent 2.11 ±6.49mg, range 1.5-40mg daily). Regular eye-drop use (n=9, 21%), medications with anticholinergic effects (n=12, 29%), and selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs) (n=13, 31%) were also frequent. In contrast, only 33% (n=14) of the healthy control group were diagnosed with a medical condition (none of which caused regular pain), with most on no regular medication or a single medication (Table 7.1).

Baseline symptom burden in the chronic pain group was moderate, with mean pain and fatigue scores of 3.9 ±1.8 and 5.6 ±2.4 respectively. Only four participants (10%) reported severe pain (NRS ≥7/10). PROMIS Pain Intensity T-scores were mildly elevated relative to population norms (51.7 ±5.9), whereas Pain Interference was in the moderate range (61.6 ±8.3). Pain catastrophising levels were modest overall, with mean PCS scores (20.1 ±13.5) below commonly used thresholds for high catastrophising (>30).

Table 7.1. Baseline participant characteristics for the VR-FOCUS cohort

Characteristic	Healthy controls (n=42)	Chronic pain (n=42)
Age in years, mean (SD)	35.3 (12.2)	57.8 (11.4)
Gender, n (%)		
Male	13 (31%)	8 (19%)
Female	29 (69%)	34 (81%)
Ethnicity, n (%)		
White	31 (74%)	41 (98%)
Non-white	11 (26%)	1 (2%)
Education level, n (%)		
Up to A level	9 (21%)	19 (45%)
University	33 (79%)	16 (38%)
Previous VR Exposure, n (%)	34 (81%)	22 (52%)
Number of Medical Conditions, median (IQR)	0 (0-1)	5 (3-6)
Number of Regular Medications, median (IQR)	0 (0-1)	5 (3-7)

Data are n (%) unless otherwise stated. IQR, interquartile range; SD, standard deviation; VR, virtual reality.

7.4.1 Descriptive task-level outcomes across VR N-back levels

7.4.1.1 Performance indices

Performance declined with increasing task level in both groups, with the steepest reductions occurring between 1-Back and 3-Back (Figure 7.2).

When comparing the groups, healthy controls achieved more correct responses (“hits”) at all task levels (Figure 7.2, Panel A). Ability to discriminate between true targets and false alarms (d′) showed a similar pattern, except at the 3-Back level (Figure 7.2, Panel B). Response bias (c) became more conservative with increasing

task load in both groups. The chronic pain group demonstrated more conservative bias across 1-Back through 3-Back, with higher bias at lower task load (Figure 7.2, Panel C).

Mean reaction times for both “hits” and “false alarms” were faster in the healthy control group (difference in means 36.7ms and 17.5ms, respectively), with mean “hit” reaction times being faster than with “false alarms” overall. Detailed task-level medians and exploratory between-group comparisons for hits, d′, c and mean hit reaction time are provided in Appendix 4, Table A4.1.

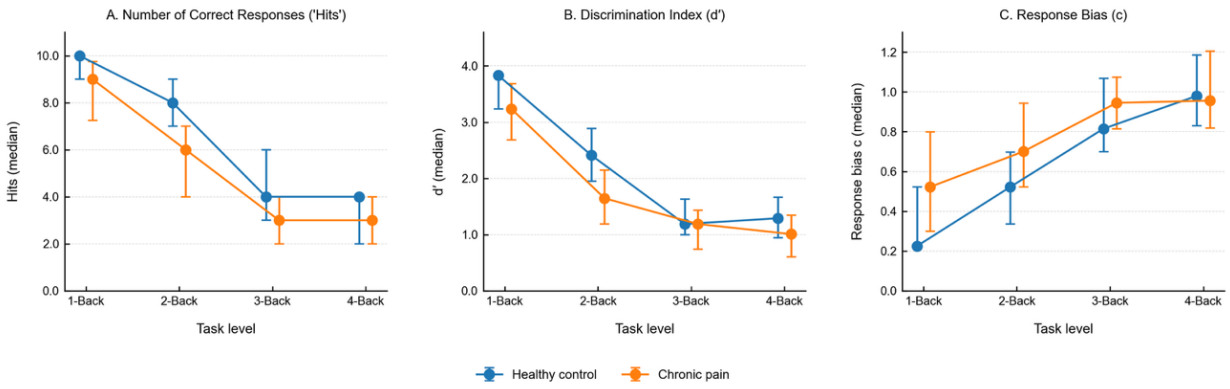


Figure 7.2. Performance indices across tasks and between groups

Task performance across increasing N-back load for healthy control and chronic pain participants. Panel A shows the median number of correct target responses (“hits”) at each task level. Panel B shows discrimination index (d′), with higher values indicating better discrimination between target and non-target stimuli. Panel C shows response bias (c), with higher positive values indicating a more conservative response tendency and lower values indicating a more liberal tendency to respond “target”. Points show group medians and error bars show variability around the median.

7.4.1.2 Participant-reported cognitive load and post-task measures

Similar patterns were observed across all self-reported post-task measures of cognitive load, with the NASA-TLX total, mental demand subscale and PAAS measures increasing sharply from Baseline through 3-Back before stabilising (Figure 7.3, Panels A-C).

Group differences were uniformly small across all levels and measures, with similar reporting of cognitive load in both the healthy control and chronic pain groups.

Perceived difficulty maintaining task focus increased with the task level. Clear between-group differences emerged from 3-Back onwards, with participants in the chronic pain group reporting greater difficulty with the higher N-back levels.

Motivation and engagement ratings remained high across all task levels, with median scores above 90/100 from 1-Back onwards in both groups. Additional participant-reported task-level summaries, including pain, fatigue, workload, focus difficulty, immersion, comfort, motivation and engagement, are provided in Appendix 4, Table A4.2.

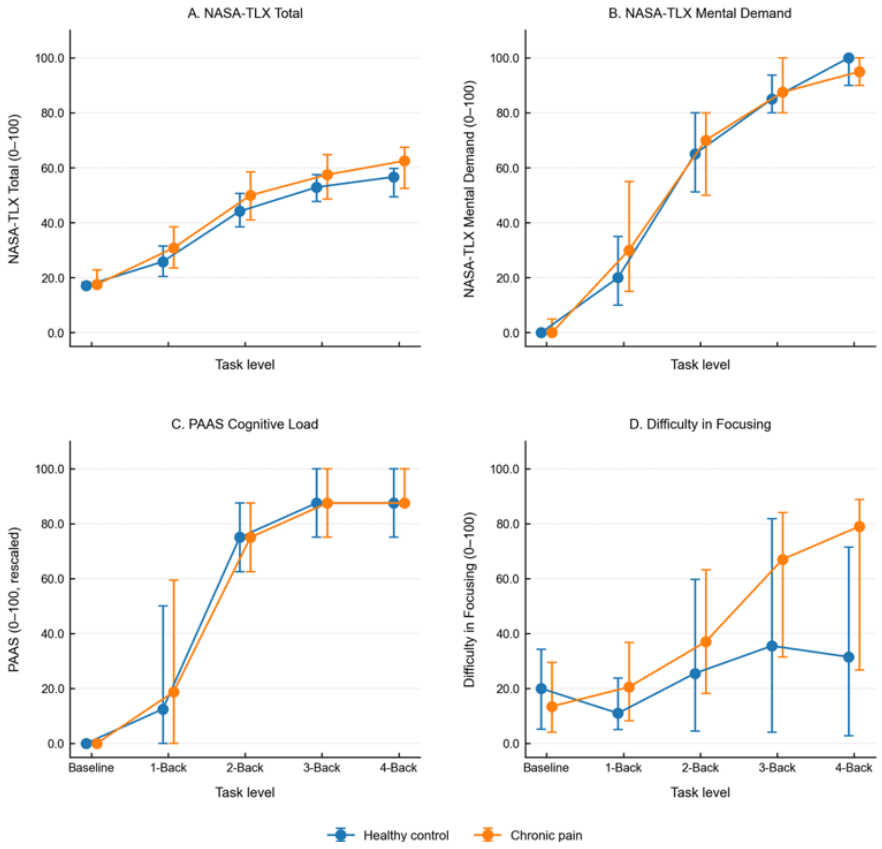


Figure 7.3. Participant-reported post-task measures across tasks and between groups

Subjective task ratings across increasing N-back load for healthy control and chronic pain participants. Panel A shows total NASA-TLX score, Panel B shows NASA-TLX mental demand sub-score, Panel C shows PAAS cognitive load ratings rescaled to 0-100 for comparability, and Panel D shows self-reported difficulty in focusing. Higher scores indicate greater perceived workload, mental demand, cognitive load or difficulty maintaining focus. Points show group medians and error bars show variability around the median.

7.4.2 Pupil diameter and blink rate

Mean pupil diameter increased from Baseline to 2-Back in both groups and then remained stable through 3-Back and 4-Back (Figure 7.4, Panels A-B). Blink rate showed a similar early increase in both groups, but whereas it continued to rise across higher task levels in the chronic pain group, it plateaued from 2-Back onwards in healthy controls (Figure 7.4, Panel C).

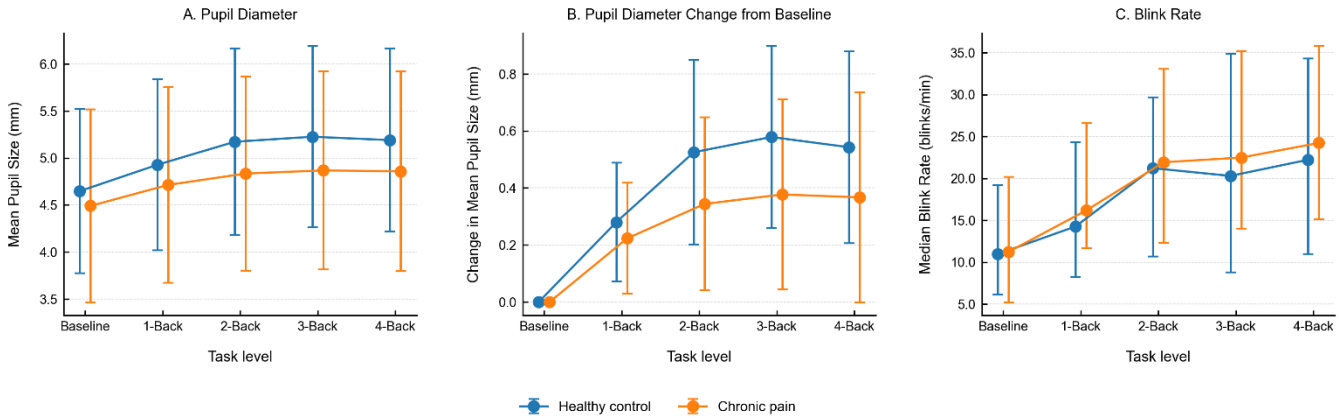


Figure 7.4. Trends in pupil diameter and blink rate across N-back levels and between groups

Eye tracking measures across baseline and increasing N-back levels for healthy control and chronic pain participants. Panel A shows mean pupil diameter in millimetres. Panel B shows change in mean pupil diameter relative to baseline. Panel C shows median blink rate in blinks per minute. Points show group summary values and error bars show variability around the summary estimate.

7.4.2.1 Relationship between eye tracking features and cognitive load

Mixed-effects models showed that higher subjective cognitive load predicted larger pupil dilation, with significant effects for NASA-TLX total (β-coefficient 0.009, 95% CI 0.009-0.010, p<0.001) and PAAS scores (β-coefficient 0.051, 95% CI 0.045-0.058, p<0.001). These effects were consistent across both groups, with no group difference in overall pupil change from Baseline to 4-Back (p=0.196). Higher cognitive load also predicted increased blink rate, with significant effects for NASA-TLX (β-coefficient 0.187, 95% CI 0.139-0.234, p<0.001) and PAAS (β-coefficient 1.081, 95% CI 0.827-1.335, p<0.001), again with no group-level differences.

Replacing subjective ratings with N-back level as the predictor produced the same overall pattern. Relative to Baseline, each task level was associated with greater

pupil dilation (model-estimated mean difference in pupil diameter: 1-Back +0.252mm; 2-Back +0.435mm; 3-Back +0.478mm; 4-Back +0.455mm; all $p<0.001$), with most of the increase occurring by 3-Back. Using 1-Back as the reference condition yielded analogous results, with pupil dilation significantly greater in the 2-, 3- and 4-Back tasks (all $p<0.001$), supporting the use of these contrasts in subsequent classification analyses. Blink rate demonstrated a similar increase with N-back level, with significantly higher blink rate at each task level compared with Baseline (all $p<0.001$). For both eye tracking features, no group effect or group-by-task interaction was observed after adjustment.

Age was positively associated with both pupil size ($p=0.019$) and blink rate ($p=0.015$), although including age as a covariate did not alter the task or group effects. Pain and fatigue scores were not associated with either pupil size or blink rate. However, several behavioural and oculomotor indicators showed associations with both pupil size and blink rate, including performance indices (d′ and c) and self-reported difficulty focusing on the task (all $p<0.001$).

7.4.3 Predictive models

7.4.3.1 Logistic regression

A logistic regression model using summary eye tracking features showed limited discriminative value for classifying low-load (1-Back) versus high-load (4-Back). Using mean pupil size alone, accuracy was 0.56 (95% CI 0.45-0.66, AUC 0.57) in healthy controls and 0.51 (95% CI 0.41-0.61, AUC 0.53) in the chronic pain group. Including blink rate provided only a small improvement, with accuracies reaching 0.60 (AUC 0.62-0.63). These values remain close to chance-level performance, indicating that simple aggregate pupil and blink metrics have limited discriminative value for this contrast.

7.4.3.2 Time-series classifiers

Evaluation of a diverse range of time-series classifiers (Table 7.2) showed consistently high accuracy for distinguishing 1-Back from 4-Back task levels in healthy controls using pupil-derived signals, with broadly comparable performance across classifiers. In contrast, classification performance in the chronic pain group was attenuated and showed greater inter-individual variability. Adding a blink-mask as a second time-series feature had negligible impact in healthy controls but

produced small, reproducible gains in the chronic pain group, suggesting additional discriminative information carried by blink dynamics in this group.

Accuracy across classifiers ranged between 0.71-0.87 in healthy controls and 0.56-0.66 in the chronic pain group (Table 7.2). A comparison of ROC curves between groups for classifiers that produce probabilistic estimates is shown in Figure 7.5. Given its computational efficiency, consistent behaviour across folds and performance, ROCKET was selected as the primary exemplar, with corresponding contingency tables presented in Figure 7.6. In healthy controls, ROCKET's pooled confusion matrix showed a clear diagonal structure with minimal misclassification, and both sensitivity and specificity on held-out participants were substantially above chance.

Including blinks in the form of a blink mask as an additional time-series feature made little difference in healthy controls but yielded small, consistent gains in the chronic pain group. ROCKET showed a modest improvement in accuracy from 0.64 to 0.70 in the 1-Back versus 4-Back comparison with similar effects observed across other classifiers. These findings suggest that blink dynamics may provide complementary discriminative information in chronic pain, possibly reflecting altered oculomotor stability or compensatory blink patterns under cognitive load.

Given that earlier mixed-effects models identified age as a potential confounder, exploratory linear regression was conducted to evaluate whether age contributed to classification accuracy. In a model including both age and group, age showed no meaningful association with accuracy (β-coefficient −0.0006, p=0.793), accounting for less than 0.1% of explained variance. In contrast, group membership was associated with accuracy after adjusting for age (β-coefficient +0.1882, p=0.018), indicating higher classification accuracy in healthy controls than in participants with chronic pain. No age-group interaction was observed, suggesting that age does not differentially affect accuracy across groups.

Computational tractability differed substantially between classifiers. QUANT was fastest (0.79s/fold; 33.1s total), followed by ROCKET (18.0s/fold; 12.6min total), whereas DrCIF was markedly slower (21.2min/fold; 14.9h total). HIVE-COTE 2.0 was unfeasible to run uncontracted on the local development machine and was therefore contracted to 8 minutes per fold. Estimates averaged over 30 prior

classification experiments conducted by aeon developers suggest that an uncontracted full run would be approximately 11.5 times slower than DrCIF (approximately 243.8min/fold; 171.4h total).

Table 7.2. Classifier accuracy metrics for 1-Back and 4-Back class labels with left pupil size as signal input

	Accuracy		**ROC AUC**		**Sensitivity**		**Specificity**	
Classifier	HC	CP	HC	CP	HC	CP	HC	CP
ROCKET	0.87	0.64	N/A	N/A	0.81	0.67	0.93	0.62
HIVE-COTE 2.0	0.83	0.61	0.91	0.66	0.79	0.57	0.88	0.64
DrCIF	0.85	0.66	0.92	0.65	0.81	0.64	0.88	0.67
Rotation Forest	0.82	0.60	0.90	0.72	0.76	0.50	0.88	0.69
QUANT	0.71	0.56	N/A	N/A	0.71	0.52	0.71	0.60

Values show classifier performance for binary classification of lower versus higher load (1-Back versus 4-Back) using continuous left pupil diameter time-series data. Accuracy represents the proportion of correctly classified cases. ROC AUC is reported only for classifiers generating probabilistic outputs. Sensitivity indicates the proportion of 4-back cases correctly classified, and specificity indicates the proportion of 1-back cases correctly classified. HC, healthy controls; CP, chronic pain; ROC AUC, receiver operating characteristic area under the curve; N/A, not available.

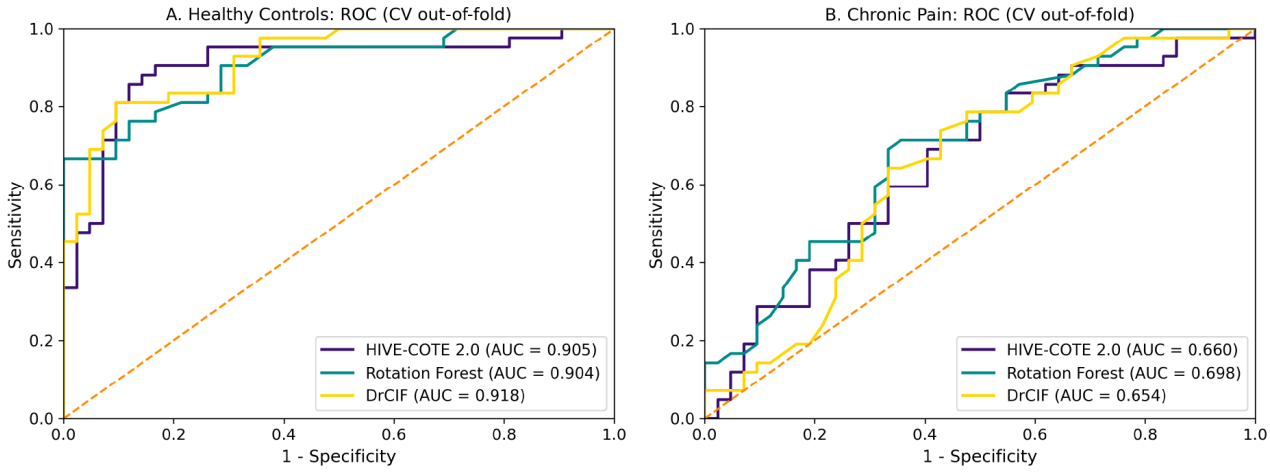


Figure 7.5: ROC curves for 1-Back vs 4-Back across HIVE-COTE 2.0, Rotation Forest and DrCIF classifiers (left pupil size only)

Receiver operating characteristic (ROC) curves for classifier performance in healthy control and chronic pain groups using continuous left pupil diameter as the input signal. Panel A shows cross-validated out-of-fold ROC curves for healthy controls. Panel B shows cross-validated out-of-fold ROC curves for participants with chronic pain. Curves are shown for HIVE-COTE 2.0, Rotation Forest and DrCIF classifiers. The dashed diagonal line indicates chance-level discrimination. Area under the curve (AUC) values are shown in the legend, with higher values indicating better separation of 1-back and 4-back task classes.

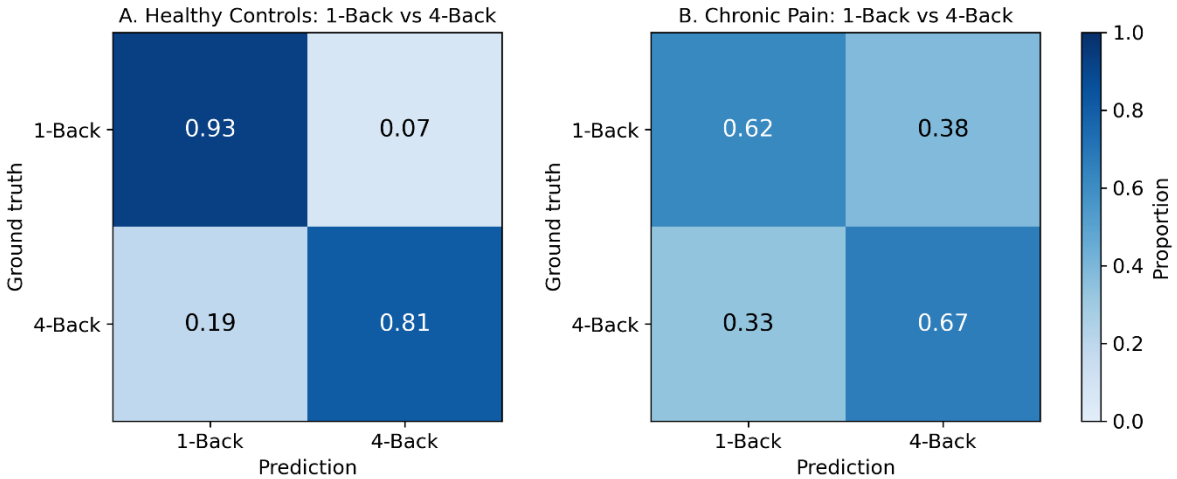


Figure 7.6: Normalised confusion matrices for 1-Back vs 4-Back ROCKET classification (left pupil size only)

Normalised confusion matrices for the ROCKET classifier using continuous left pupil diameter as the input signal. Panel A shows classification performance in healthy controls and Panel B shows classification performance in participants with chronic pain. Rows indicate the true task class and columns indicate the predicted class. Values represent the proportion of cases within each true class. Correct classifications are shown on the diagonal.

7.4.3.3 Exploratory time-series regression

Exploratory regression modelling of continuous NASA-TLX scores broadly paralleled the classification findings, with stronger performance in healthy controls than in participants with chronic pain. Using left pupil size across all task levels, ROCKET Regressor achieved moderate fold-level association with NASA-TLX in healthy controls (Pearson r 0.70 ±0.30; MAE 12.5 ±5.2). Performance was weaker and more variable in the chronic pain group (Pearson r 0.50 ±0.40; MAE 15.0 ±5.8).

7.4.4 Adverse effects and task acceptability

VR tasks were well tolerated, with overall adverse effects remaining low across participants. Median Virtual Reality Sickness Questionnaire (VRSQ) scores were 8.3 [4.2–15.8] out of 100, indicating minimal cybersickness symptoms during the protocol. Although participants with chronic pain reported higher VRSQ scores than healthy controls (12.5 [6.7–17.3] versus 4.2 [4.2–12.5], p=0.006, r=0.34), absolute symptom levels were still low in both groups. Pain and fatigue VAS scores were higher in participants with chronic pain but remained stable across N-back levels, with a slight downward trend in median pain scores between Baseline (18.0, [9.3-25.3] and 4-Back (10.5, [1.0-19.0]). Domain-level VRSQ tolerability scores are provided in Appendix 4, Table A4.3.

7.5 Discussion

This study investigated whether consumer-grade, VR-integrated eye tracking can provide objective markers of cognitive load that could inform future closed-loop pain therapeutics, in healthy adults and a pragmatic, clinically heterogeneous cohort with chronic musculoskeletal pain. Increasing N-back level led to poorer performance and higher subjective workload; in parallel, pupil diameter and blink rate changed systematically with task level and workload under stable luminance. Simple models based on aggregate pupil and blink features showed limited discrimination of low versus high load, whereas time-series classifiers achieved higher participant-level performance, particularly in healthy controls. Although mixed-effects models indicated broadly similar mean task-evoked ocular responses across groups, classification performance was lower and more variable in chronic pain. This variability did not appear attributable to age and underscores the additional challenges of modelling cognitive load in chronic pain groups, motivating cohort-

sensitive and participant-calibrated approaches. Overall, these findings support VR-embedded eye tracking as a feasible biosignal for cognitive-load estimation, while highlighting the need for personalised models for translation to closed-loop applications in chronic pain.

The present results extend a large evidence-base demonstrating task-evoked pupillary responses (TEPRs) as sensitive indices of cognitive effort in working-memory paradigms, including the N-back[176, 189]. TEPRs are widely interpreted as reflecting engagement of arousal and cognitive control systems, including the locus coeruleus–noradrenergic system, with pupil diameter typically increasing with cognitive load and stabilising or attenuating as demands approach capacity[176, 341]. In this N-back paradigm, mean pupil diameter increased from Baseline to 2-Back and then remained comparatively stable through 3-Back and 4-Back, broadly mirroring trajectories in performance and subjective workload, and consistent with a capacity-related plateau at higher levels.

The magnitude of pupillary change (approximately 0.2 to 0.6 mm) falls within ranges reported in laboratory pupillometry studies of mental workload[342]. Metrological work also indicates that video-based eye trackers can resolve sub-millimetre changes under controlled conditions, supporting the physiological interpretability of fractional-millimetre effects when luminance and data quality are appropriately managed[343]. In VR, additional measurement considerations apply (for example, gaze-angle dependent pupil-size artefacts), reinforcing the value of centrally presented stimuli and stabilised luminance across blocks[166, 344]. Together with prior VR studies indicating that pupil dynamics can track cognitive load when luminance is controlled or modelled[345, 346], these findings support the practicality of headset-integrated pupillometry for workload estimation in immersive applications.

Blink rate also increased with task level and subjective workload, but its interpretation is more context-dependent than pupil dilation. Blink suppression is frequently reported when continuous visual sampling demands are high[181, 184, 347], whereas other studies have shown increased blink rate with tasks requiring sustained mental effort and internally focused processing[182, 183]. In this task, blink rate increased systematically from Baseline to 2-Back, with a relative plateau thereafter. Notably, blink rate was higher overall in the chronic pain cohort. Exploratory clinical-context models did not identify a consistent pattern suggesting that pain severity,

pain interference, pain catastrophising, diagnosis, medication burden, opioid exposure, anticholinergic medication or SSRI/SNRI use explained pupil diameter or blink-rate outcomes (Appendix 4, Table A4.4). Despite this, higher blink rate in the chronic pain cohort may still reflect broader clinical factors and processes, including autonomic dysregulation, ocular surface symptoms, fatigue, medication exposure or unmeasured co-morbidity[185]. These results argue against blink counts as a standalone marker of cognitive load in VR but support their inclusion within multivariate models that account for task structure, viewing conditions, and individual differences.

The neurobiology of blinking helps to explain why blink behaviour may carry task-relevant information despite its limited specificity. Blinking is governed by partially overlapping brainstem, subcortical and cortical circuits that integrate ocular-surface input with central influences related to attention, arousal and motor control[348]. Spontaneous blink rate has also been linked to dopaminergic function through pharmacological, clinical and cognitive studies[349], although human PET and pharmacological studies have not demonstrated a simple relationship with striatal dopamine receptor availability or dopaminergic stimulation[350]. Blink rate should therefore not be interpreted as a direct proxy for central dopamine. More consistently, the timing of blinks varies with information-processing demands. Blinks tend to occur before and after periods of sustained processing[351] and can be adapted strategically according to when relevant visual information is expected[352]. Blink-related increases in default-mode network activity and decreases in dorsal attention network activity have also been reported during attentive viewing, consistent with transient attentional disengagement[353]. As blinking briefly interrupts visual input and is accompanied by neural suppression of visual sensitivity[354], its timing may therefore provide information about how attention is organised across a task rather than indexing a single underlying neurophysiological process.

This framework provides a plausible interpretation of the present findings without attributing changes in blink rate to a single mechanism. The 750ms stimulus followed by a 2000ms interstimulus interval created a predictable temporal structure in which blinks could be suppressed during stimulus presentation and occur during periods in which visual information was less critical. Previous working-memory studies have reported increased event-related blinking during memory updating and switching

between maintenance and updating states[355], although these processes were isolated using paradigms specifically designed to manipulate working-memory gating and cannot be mapped directly onto N-back level. In the present task, increasing N-back level increased working-memory and monitoring demands without changing the frequency of stimulus presentation. The accompanying increase in blink rate is therefore more cautiously interpreted as reflecting changes in task-related processing and the temporal organisation of attention.

The chronic pain cohort showed marginally poorer task performance, slower reaction times and greater subjective difficulty maintaining focus at higher N-back levels, despite broadly similar group-level ocular responses. This is consistent with evidence that chronic pain can be associated with attentional dysregulation and impairment in working memory and executive function, particularly in older adults and those with higher pain interference[204, 205]. Experimental studies also indicate that pain can disrupt N-back performance and increase false alarms, consistent with competition between nociceptive and task-related processing for limited cognitive resources[356].

Physiologically, the absence of marked group differences in mean task-related pupillary response suggests that chronic pain did not simply blunt pupillary responsivity to cognitive load. Instead, the machine learning analyses highlight more nuanced alterations in how ocular dynamics relate to task demands at the individual level. Time-series classifiers discriminated low versus high load more reliably in healthy controls than in the chronic pain group, with increased off-diagonal misclassifications under leave-one-subject-out cross-validation. Despite age differences between groups, linear regression showed no significant association between age and classifier performance, providing reassurance that this pattern is not solely age-driven. A plausible interpretation is that chronic pain is associated with increased variability in the temporal coupling between cognitive demand and ocular responses, rather than a uniform shift in mean physiology. This supports the view that chronic pain does not represent a single cognitive phenotype, and motivates participant-calibrated and cohort-sensitive modelling approaches for future biomarker-informed or adaptive systems.

Immersive VR is increasingly used as a non-pharmacological adjunct for both acute and chronic pain, typically through reallocation of attentional resources away from

pain (attentional modulation), movement-based activities, skills training or psychoeducational approaches[5]. Systematic reviews and recent randomised controlled trials report improvement in pain-related outcomes across multiple clinical contexts, although effect sizes are variable and uncertainty remains around dosing, mechanisms and target populations[68, 75, 80, 85]. Most deployed or evaluated VR interventions remain open-loop, with fixed content that does not adapt to patients' cognitive load, symptom burden or fatigue.

The present findings suggest a pathway toward adaptive VR therapeutics that explicitly monitor and regulate cognitive load. A closed-loop system could combine pupillometry, blink dynamics and performance features to estimate whether a user is under-challenged, optimally engaged or overloaded, then adjust task difficulty, interaction demands, or sensory richness in real time. When load is below a personalised engagement target or "flow range", the system might increase goal-directed interaction to sustain attention. When overload is detected, it could simplify tasks or transition to calmer content to limit fatigue and disengagement. In chronic pain populations, where baseline cognitive strain and symptoms fluctuate, the observed variability supports the design of algorithms that accommodate heterogeneity rather than assuming a fixed mapping between ocular dynamics and cognitive load.

Computational tractability is also central to the translational pathway. Future closed-loop VR systems would require models that generate reliable state estimates with low latency and limited on-device computational burden, rather than simply maximising offline classification performance. In the present analyses, higher-capacity classifiers such as DrCIF and HIVE-COTE 2.0 were useful offline comparators but imposed substantially greater runtime demands, whereas ROCKET provided a more favourable balance between discriminative performance and practical feasibility. This supports the view that model selection for adaptive VR therapeutics should consider accuracy, generalisability and deployability together.

To improve prediction accuracy and real-time feasibility further, cognitive-load estimation may benefit from integrating complementary signals that capture different aspects of task engagement. Prior work in non-clinical domains suggests that no single physiological measure generalises across all tasks and contexts, and that combining complementary features may improve robustness or interpretability of

workload estimation in certain settings[188, 189]. Multimodal approaches, including complementary biosignals such as heart-rate variability or EEG where feasible, may help disentangle cognitive effort from arousal and affective responses, though their incremental benefit should be weighed against added complexity. Critically, cognitive-load adaptation must be aligned with therapeutic intent: some interventions may aim to sustain moderate-high demand and engagement (for example, attention-based and CBT modules), while others may target reduced cognitive load (for example, mindfulness-oriented VR). Closed-loop control should therefore be guided by theory-driven therapeutic targets, not engagement alone.

Several methodological components strengthen the interpretation of this work. The study used a consumer-available VR headset with integrated binocular eye tracking, demonstrating that informative pupil and blink signals can be obtained within a realistic deployment setting outside tightly constrained laboratory conditions. Cognitive load was manipulated with a well-characterised N-back task and quantified using complementary measures, including validated self-report instruments (NASA-TLX, PAAS) and signal-detection performance metrics. Eye tracking was analysed as continuous time series across task blocks to preserve temporal ocular dynamics, luminance was logged and controlled at the block level, and preprocessing and feature extraction were embedded within cross-validation folds to reduce leakage. Performance was evaluated across multiple open-access, state-of-the-art time-series classifiers under leave-one-subject-out cross-validation, providing a conservative estimate of participant-level generalisation and supporting reproducibility.

Limitations do remain in this study. The single-session design precludes conclusions regarding within-person stability across days and fluctuating symptom states. The chronic pain sample was clinically representative but heterogeneous, comprising multiple diagnoses, comorbidities, and medication exposures that influence cognition and ocular physiology. Larger studies with stratified sampling and richer phenotyping are needed. The absence of a demographically matched control group limits interpretation of between-group comparisons. Pain severity and impact were generally low-to-moderate, limiting generalisability to more severe or disabling pain. The N-back task provides a controlled probe of cognitive load but does not capture the full complexity of therapeutic VR experiences. While luminance was controlled here, therapeutic VR will typically involve greater visual and luminance variability that

will need to be explicitly modelled to support future real-world closed-loop use. The acquired data were also heterogeneous at the signal level, combining two headset sampling rates and requiring interpolation across blink-related gaps. This was handled within preprocessing but made the classification problem more demanding than a uniformly standardised laboratory dataset would present. The analytical workflow was narrowed pragmatically toward pupil-centred inputs and binary load contrasts to obtain stable estimates, so the classifier comparisons are best interpreted as evidence on the feasibility and relative suitability of different model families rather than as a definitive ranking of classifier performance. Finally, all machine-learning models were trained and evaluated using internal cross-validation, and no independent external dataset was available for validation. As such, the generalisability of the classification results beyond the present sample remains to be established. Moreover, all analyses were conducted offline on desktop hardware; translation to real-time implementation on standalone VR devices will require additional engineering, model optimisation and prospective validation.

Future research should prioritise replication in larger and more diverse chronic pain cohorts, longitudinal studies across varying pain and fatigue states, and study of alternative VR hardware and eye tracking pipelines. Prospective trials should test whether cognitive-load-adaptive control improves engagement, tolerability, and clinical outcomes compared with open-loop VR within ecologically valid therapeutic applications. Translational progress will require close collaboration between clinicians, VR developers, and machine-learning researchers to ensure robustness to device heterogeneity and missing or noisy data, alongside acceptability, interpretability, and strong data governance.

In conclusion, this study provides initial evidence that consumer-grade VR-integrated eye tracking captures task-evoked changes consistent with cognitive load during a VR N-back task in adults with and without chronic musculoskeletal pain. While summary pupil and blink features offered limited discrimination of low versus high load, time-series machine learning classifiers that utilise temporal dynamics achieved higher discriminative performance, with performance attenuated and more variable in the chronic pain group. These findings support VR-embedded eye tracking as a practical route to estimating cognitive load in immersive settings, and

motivate participant-calibrated, externally validated approaches as next steps toward safe, clinically meaningful closed-loop VR interventions.

8. Synthesis

This chapter synthesises findings from the three empirical studies and considers their joint implications for the development of future closed-loop VR therapeutics in chronic musculoskeletal pain. Points of convergence and divergence between the EEG and eye tracking work are drawn out, alongside their respective implications for sensor selection, adaptation targets and decision logic in future adaptive systems. Methodological contributions are summarised, cumulative limitations spanning sample composition, temporal stability and multimodal integration are considered, and priorities for the next phase of research are set out.

8.1 Outcomes from the empirical studies

Across the three studies, the central message is that mechanistic biomarker investigation, embedded state-sensitive sensing and predictive modelling serve different translational purposes. PainWaves and VIPA-EEG contribute mainly to mechanistic phenotyping, where the priority is to identify candidate neurophysiological features that may act as future stratification tools or targets for closed-loop therapeutics, rather than to develop immediately deployable control signals. VR-FOCUS contributes to embedded sensing and inference, where the priority is to test whether headset-derived signals can support prediction of user state in a form that could eventually inform adaptation. Together, the studies identify beta band activity, alpha state-responsiveness and ocular cognitive-load inference as candidate components for further development. This chapter synthesises the findings by clarifying what each component contributes, what role it might play in future adaptive VR, and what evidence is still needed before these foundations could be translated into clinically useful immersive therapeutics for chronic musculoskeletal pain.

8.1.1 Beta band rsEEG activity

Beta band activity recurs across the EEG studies, although its meaning differs between them. In PainWaves, frontocentral beta power was lower in twins with CWP with effects strengthening under stricter quality-control thresholds and surviving in MZ pairs at high beta. In VIPA-EEG, broad beta and low beta increased after VR

exposure in fibromyalgia, with the effect present from the first post-activity recording and partly activity-sensitive at posterior sites. The directional contrast between reduced beta in CWP and increased beta after immersive exposure in fibromyalgia is suggestive of a normalisation pattern, with VR shifting beta band activity in the opposite direction to the CWP-associated reduction observed at trait level. This interpretation is plausible but requires caution.

The two cohorts differed in important ways. PainWaves recruited older participants from a population-based twin registry, TwinsUK, and selected twin pairs discordant for a questionnaire-derived CWP phenotype, with relatively mild overall symptom burden. In contrast, VIPA-EEG recruited a clinical fibromyalgia cohort with high symptom burden, moderate depressive symptoms and greater age variability, with a lower mean age overall (mean age 46.4 ±13.8 years, versus 69.9 ±8.1 years in PainWaves). The acquisition systems also differed, with PainWaves using a 32-channel sponge-based passive electrode system and VIPA-EEG a 64-channel gel-based active electrode system. PainWaves examined cross-sectional within-pair differences whereas VIPA-EEG examined within-subject pre-post change. VIPA-EEG had no non-VR control condition, which limits attribution of the observed change to immersive exposure specifically rather than to repeated rest, expectancy or VR-independent task effects. Beta in VIPA-EEG was also more sensitive to aperiodic adjustment than alpha, attenuating the strength of the residual oscillatory claim.

Despite these constraints, the recurrence of beta across two methodologically independent EEG datasets in chronic musculoskeletal pain populations adds incremental support to the broader literature implicating beta band alterations in chronic pain pathophysiology, including work linking beta to sensorimotor regulation, top-down control and indirect indices of inhibitory function[151, 152, 156]. Beta-range physiology emerges as the most consistent candidate domain for future mechanistic work in chronic musculoskeletal pain across the empirical studies. PainWaves provides more confidence that the trait-level signal is indeed related to CWP rather than being influenced by genetic or environmental factors, and VIPA-EEG provides preliminary evidence that beta band activity is responsive to immersive intervention in a fibromyalgia cohort. Neither study supports a strong mechanistic claim in isolation, but the recurrence of the same frequency band alteration in two independent datasets is more informative than either dataset alone.

8.1.2 Alpha as a more state-responsive rhythm

Alpha occupied a different position across the EEG studies. In PainWaves, alpha alterations were secondary, with lower effect sizes and a more central topographic focus. In VIPA-EEG, alpha was the strongest and most robust spectral effect of VR exposure, with the clearest evidence for residual oscillatory change after aperiodic separation, a progressive within-session trajectory, and convergent alpha peak power increases across occipital and central ROIs. These findings suggest that alpha may be more state-responsive than trait-defining in chronic musculoskeletal pain, at least within the cohorts and acquisition contexts examined here. This is broadly consistent with established models in which alpha is linked to functional inhibition, sensory gating and regulation of cortical information flow rather than representing a passive resting rhythm[314, 315].

The interpretation has translational implications. If alpha is more sensitive to within-session context than to long-term pain status, it may be useful as an index of immersive engagement, post-task settling or attentional reallocation rather than as a candidate diagnostic biomarker for chronic pain. The within-session alpha increase observed in VIPA-EEG is notable in this respect, as it dissociates alpha from immediate task-specific responses and aligns it more closely with cumulative within-session processes. Alpha may therefore be more appropriate as a candidate process indicator within future adaptive systems, useful for monitoring whether an intervention is producing the intended state shift, rather than as a target of mechanistic engagement itself.

8.1.3 VR-embedded eye tracking as a signal for state inference

VR-FOCUS demonstrated that ocular signals captured through VR-embedded eye tracking respond systematically to graded cognitive demand. Mean pupil diameter and blink rate increased with N-back level under stable luminance in both groups, mixed-effects models linked these features to subjective workload, and continuous pupil signals supported subject-independent discrimination of cognitive load conditions. These patterns are broadly consistent with the established cognitive load literature, in which task-evoked pupillary dilation, blink dynamics and fixation behaviour have been shown to index working memory load, attentional allocation and arousal[171, 176, 264, 357]. What VR-FOCUS contributes is the demonstration that the

same broad pattern is recoverable from a consumer-grade head-mounted VR system, during completion of cognitively challenging tasks in a VR environment, under pragmatic conditions consistent with eventual deployment in a therapeutic VR context, in both healthy controls and people with chronic musculoskeletal pain.

Despite this, there are several translational limitations that need to be addressed before integration into a consumer-available VR therapeutic would be possible. Acquisition was confined to a single VR eye tracking system (Pico Neo 3 Pro Eye with Tobii binocular eye tracking) under controlled luminance, with seated participants and standardised ambient conditions. Pupil responses are sensitive to light, gaze direction and arousal states that may be unrelated to cognitive load, so deployment beyond this constrained setting would need to account for these confounds. This includes testing in more realistic, chronic pain-relevant VR applications, where cognitive demand is likely to occur alongside visual stimulation, movement and affective engagement. Longitudinal within-person stability across days, symptom states, medication exposure and time of day was not characterised, with only a cross-sectional measurement per participant collected. Generalisation to different VR systems, different usage environments and over longer or more variable session structures has not been tested.

The clearest population-level observation was that mean ocular responses to graded task demand were preserved across groups, but classifier performance dropped in chronic pain participants in a way that age did not explain. The moment-to-moment coupling between cognitive demand and ocular dynamics is therefore more individually variable in the chronic musculoskeletal pain population, consistent with the wider literature on heterogeneity in chronic pain phenotypes, treatment response and within-person symptom fluctuation[9-11, 36]. VR-embedded eye tracking is therefore a useful candidate signal for state inference in future adaptive VR systems, but its utility in any given system will be conditioned on the cohort it is built for and on accommodating the inter- and intra-individual variability of that cohort.

8.1.4 Machine learning in the VR-FOCUS study

Machine learning was used in VR-FOCUS as an analytical tool to evaluate whether continuous ocular signals captured through VR-embedded eye tracking carried enough information to discriminate between cognitive load conditions in previously

unseen participants. It was not used as a way of explaining how cognitive load is generated, nor was it intended to differentiate between different types of cognitive load. The main outcome was that time-series classifiers trained on continuous data outperformed logistic regression using summary features for binary cognitive load discrimination in healthy controls and produced more modest but still informative discrimination in chronic pain. VR-FOCUS investigated several time-series classifier families, including convolution kernel-based, interval-based, vector-based, and a meta-ensemble incorporating further dictionary-based and shapelet-based methods. These methods extract discriminative information from the pupil time series in different ways. Broadly comparable performance suggests that the discriminative information was not confined to one very specific representation of the data. Instead, it suggests that the pupil time series contained task-related temporal information that could be detected using different modelling approaches.

ROCKET in particular offered a parsimonious, computationally efficient solution that approached the performance of more complex ensemble approaches, suggesting that strong predictive accuracy and plausible deployment efficiency are not necessarily in conflict in this context. Although it produced lower accuracy in both groups, QUANT still outperformed logistic regression and was 23 times faster per fold than ROCKET when run on a local machine. This is encouraging for future closed-loop VR systems that may require offline, on-device classification to support data privacy.

Some limitations should be kept in mind when interpreting this work. Classification of N-back level is not equivalent to direct estimation of an internal latent cognitive state. The labels used in these analyses were operational task labels, meaning the models discriminated between task conditions associated with differing cognitive demand rather than measuring cognitive load itself. All VR-FOCUS analyses were also performed offline on desktop hardware. Translation to real-time implementation will require additional engineering, model adaptation and prospective validation. For future closed-loop interventions, it will be necessary to evaluate on-device inference latency, computational burden and robustness to missing or noisy data. This includes determining whether model compression or TinyML-style deployment could support efficient local execution on headset processors or associated embedded hardware, and whether ‘on-chip’, edge-based or cloud-based architectures offer the best

balance between latency, power consumption, privacy, reliability and predictive performance. Even with these limitations, the machine learning work in VR-FOCUS provides a foundational methodological starting point for state inference to inform future adaptive or closed-loop VR work.

8.2 Implications for the design of future closed-loop VR therapeutics

8.2.1 Informing future system design requirements

A useful adaptive VR system for chronic pain requires three things working together: a sense of what is biologically worth targeting, a means of measuring user state during the intervention, and a rule for adjusting the intervention in response. PainWaves and VIPA-EEG address the first of these requirements, VR-FOCUS addresses the second, and partly the third. None of the completed studies address the rule that links a state estimate to a change in the intervention itself, and the work does not implement a closed-loop system. What it provides is groundwork for the components that sit underneath the eventual decision logic.

This framing is broadly consistent with closed-loop and just-in-time adaptive intervention models in which decision logic, tailoring variables and modifiable intervention components are treated as distinct elements[124, 229, 230]. It is also consistent with current MRC guidance on complex intervention development, which emphasises programme theory, feasibility, context, implementation and iterative refinement as components of evaluation rather than as preparatory stages prior to a single definitive trial[125].

8.2.2 Sensor selection in light of the present findings

The sensor selection rationale developed in Chapter 2 anticipated several of the trade-offs that emerged across the empirical work. PainWaves demonstrated that informative rsEEG features can be acquired at scale using a pragmatic 32-channel sponge-based system, applied by non-specialist operators, within a broader research workflow. This provides reassurance that mechanistic EEG investigation does not have to be confined to high-density gel-based laboratory acquisition and supports the use of pragmatic EEG acquisition approaches for collecting data at scale in larger participant cohorts. Limitations of the pragmatic system were also evident, particularly in the moderately variable retained data length and quality across participants. VIPA-EEG complemented this by showing that a 64-channel gel-

based system can be used effectively in a repeated-measures intervention context, while also illustrating how VR hardware and session design can constrain EEG acquisition. The combination of an EEG cap, VR-HMD, repeated post-activity recordings and participant-burden considerations meant that post-clean retained durations were modest relative to conventional resting-state datasets.

For closed-loop VR specifically, neither EEG acquisition approach is well suited to routine use as an in-session control signal in consumer-facing systems. Use for this purpose would require affordable, comfortable dry-electrode hardware with reliable scalp contact, stable signal quality and sufficient robustness to movement and headset-related artefact. These requirements are not yet available for routine closed-loop use in consumer VR. The PainWaves and VIPA-EEG findings therefore support the position advanced in Chapter 2 that rsEEG is most useful as a mechanistic measure that can inform what an adaptive intervention should seek to engage or modify, rather than as a practical real-time control signal for current headset-based VR systems. VR-FOCUS supports the complementary positioning of eye tracking as the principal candidate state-sensitive signal for current adaptive VR development. Eye tracking is increasingly integrated into commercially available VR headsets, and as shown in VR-FOCUS, signals can be acquired continuously during the same task that delivers the intervention to support discrimination of task-defined cognitive load in pragmatic conditions[67, 166, 167]. Given these empirical findings, future closed-loop VR interventions in chronic musculoskeletal pain should plan sensor deployment according to translational role. In the near term, EEG is best positioned as a tool for mechanistic phenotyping and intervention development outside of the VR headset, whereas eye tracking is better suited as the principal embedded, state-sensitive signal stream for within-headset adaptation.

8.2.3 Adaptation targets and decision logic

The findings from VR-FOCUS support cognitive load as a plausible adaptive target within future closed-loop VR systems, while clarifying its limits. Cognitive load is theoretically linked to the established attention-pain interaction, empirically modifiable through VR task parameters, and methodologically feasible using embedded headset sensors. The mixed-effects analyses confirmed that pupil and blink features varied with both task level and subjective workload, and the predictive analyses confirmed that informative state structure was recoverable from continuous

ocular signals using time-series classifiers. The implication is that an adaptive system could in principle use ocular state estimates to detect underload or overload during immersive task performance and respond by adjusting task difficulty, pacing, sensory complexity or therapeutic content. This is closer to the just-in-time adaptive intervention framework of maintaining a therapeutic operating range than to a simple goal of pain reduction at every moment[124].

The contrast between the regression and classification analyses in VR-FOCUS is informative for what could drive that adaptation. Time-series classification of continuous pupil signals achieved usable accuracy in healthy controls and attenuated but still informative accuracy in chronic pain, whereas exploratory time-series regression predicting continuous NASA-TLX scores was weaker and more variable, particularly in the chronic pain group. The pattern indicates that discrete state classification from objective signals provides a more reliable foundation for adaptation than continuous prediction of subjective load. Self-report may still add value in combination with classification, for example as ground truth during system calibration, for periodic recalibration in use, or for resolving uncertainty in low-confidence classifier outputs. The broader principle is that adaptation logic in chronic pain VR is best driven by automated, passively acquired signals rather than by user-facing input. Use of VR already brings an element of technological burden, particularly when considering home-based delivery. Frequent in-session self-report prompts within future closed-loop VR therapeutics risks compounding that burden and contributing to further implementation barriers[358-360]. Therefore, future closed-loop systems should be designed as automated and user-friendly by default.

Adaptation logic is best treated as theory-aligned rather than engagement-driven. Different therapeutic content has different cognitive demands and different intended mechanisms, and a system that maximises engagement without reference to therapeutic intent risks pushing users toward absorbing experiences that do not deliver the intended therapeutic process. Behavioural skills training, attention-based modulation, relaxation-oriented interventions, embodied movement work and graded exposure each impose different cognitive and affective profiles, and adaptation rules should reflect these differences. Decision rules also benefit from being probabilistic and conservative. The ocular signals in VR-FOCUS were informative but not specific, and the chronic pain group showed greater variability in the temporal

coupling between task demand and ocular dynamics. State estimates are therefore better treated as graded probabilities that inform the magnitude and frequency of intervention adjustment, rather than as binary triggers for abrupt change.

8.2.4 Personalisation and the limits of population-level inference

At the group level, both cohorts showed task-related changes in pupil diameter and blink rate, with no significant group-by-task interaction in the mixed-effects models. However, these aggregate similarities did not translate into equivalent classification performance in healthy controls and participants with chronic pain. This suggests that the temporal structure of the ocular signal was less consistently discriminative in the chronic pain group, potentially reflecting greater inter-individual variability, noisier state expression or more heterogeneous physiological responses to task demand. Practically, this means that a pre-trained classifier developed primarily in people without pain may perform less accurately when deployed in people with chronic pain. In a closed-loop system, inaccurate state inference could lead the system to make inappropriate adaptations, such as increasing task demand when the user is already overloaded or reducing engagement when additional challenge would be appropriate. This creates a potential route to reduced efficacy, poor tolerability or harm if cohort-specific variation is not considered during model development and validation.

This is a real concern that requires emphasis. Although limited, evidence for VR eye tracking inference of cognitive load across training and simulation paradigms has largely come from healthy, student or other non-clinical samples[14, 361, 362]. Separately, time-series classifiers such as ROCKET, HIVE-COTE 2.0 and related methods have been benchmarked mainly on general time-series datasets rather than VR eye tracking or chronic pain cohorts[211-213, 273], and the bulk of non-VR cognitive load studies reviewed in Chapter 2 used healthy participants. A multimorbid chronic pain population brings medication exposure, fatigue, sleep disturbance, ocular symptoms and mood-related variation that can affect ocular behaviour in ways that are not captured by healthy-participant training data. More broadly, inter-individual variability in treatment response is a recognised challenge across chronic pain intervention research, supporting the need for stratified, personalised or adaptive approaches[35, 363].

The practical implication is that future closed-loop VR for chronic musculoskeletal pain will probably need to combine population-level pre-training with some form of per-user adaptation. The exact form of personalisation is an open question. Plausible options include short calibration sessions to anchor the user's baseline ocular characteristics, hybrid models that pair a population prior with online updates from the individual user, or fully personalised models trained on each user's data. Microrandomised trial methodology may help evaluate which adaptation strategies improve proximal outcomes and for which users[230]. The key point is that personalisation should be built in from the start rather than added later as a refinement.

8.3 Methodological contributions

8.3.1 Automated reproducible EEG pipelines

A consistent methodological approach across the EEG studies was the use of an automated reproducible preprocessing pipeline adapted from DISCOVER-EEG[234], with structured audit logs, tiered sensitivity analyses and consistent inferential testing across datasets. This approach supported comparability between PainWaves and VIPA-EEG while allowing for differences in acquisition context and data quality. It also reduced the risk of subjectivity in preprocessing decisions and aligned with the broader translational aim of developing scalable, reproducible EEG workflows. Methodological heterogeneity has been a recurring barrier to replication in chronic pain rsEEG[44, 127, 147], and the use of a common automated framework across two independent datasets within the same programme provides one example of how that variability can be reduced.

The translational case for automated pipelines extends beyond comparability between studies. Future closed-loop and accessible clinical applications of EEG should not rely on manual or even semi-automated preprocessing approaches. Real-time or near-real-time inference, whether for adaptive intervention or for clinical decision support, requires preprocessing that is deterministic, auditable and capable of running without human intervention. Even where deployment is offline, the same constraints apply for reproducible biomarker estimation across sites, devices and operators. The pipeline used across PainWaves and VIPA-EEG was not built for

real-time use, but its design principles are consistent with what an eventual deployment pipeline would need to provide.

8.3.2 Pragmatic acquisition approaches

The PainWaves data collection and preprocessing approach shows that mechanistic rsEEG investigation can be embedded within broader research workflows, conducted by non-specialist operators, and analysed through automated pipelines without forfeiting interpretive validity. The 32-channel sponge-based R-Net system achieved acceptable signal quality across most participants, supported the principal spectral findings, and produced effects that strengthened under stricter quality-control thresholds. This is translationally relevant. If rsEEG biomarker work in chronic pain is to scale, future studies are likely to need acquisition models that fit within existing research and clinical pathways rather than requiring dedicated specialist EEG environments. The PainWaves pipeline provides one example of how this can be achieved while maintaining methodological standards.

VR-FOCUS extended this pragmatic approach to a different modality and a different deployment scenario. The Pico Neo 3 Pro Eye is a standalone consumer-grade headset that does not require a tethered PC, and the study was designed around on-device acquisition with a straightforward data export and storage workflow rather than around laboratory-bound recording infrastructure. Data could be acquired, stored and exported by a single operator in a non-specialist setting, and the consumer availability of the headset means that this approach is replicable without bespoke hardware. Across both studies, methodological choices were made with translation in view, accepting some compromise in raw signal density or acquisition flexibility in return for workflows that more closely resemble what eventual real-world deployment would require.

8.3.3 Periodic-aperiodic separation as a routine standard

The use of FOOOF-based parameterisation across both EEG studies provided convergent evidence that the principal spectral findings were not driven by global aperiodic shifts. In PainWaves, the absence of aperiodic differences strengthened the interpretation of beta as an oscillatory rather than broadband phenomenon, even though the residual broad beta effect did not survive correction. In VIPA-EEG, the persistence of residual alpha after aperiodic adjustment strengthened the

interpretation of the alpha effect as oscillatory, while the attenuation of residual beta-family effects appropriately tempered the strength of the beta claim. Conventional band-power analyses in chronic pain rsEEG remain vulnerable to confounding between oscillatory and broadband contributions, and routine periodic-aperiodic decomposition is a useful safeguard against that conflation[157, 158]. Although parameterisation requires sufficient spectral quality and adequate data stability, it is readily implementable using widely available software packages and provides important interpretive value. Future chronic pain EEG studies should therefore plan periodic-aperiodic decomposition at the design stage, particularly where band-power findings are intended to support biomarker claims, mechanistic interpretation or intervention-related modulation.

8.3.4 Subject-independent evaluation in pragmatic machine learning

The use of grouped leave-one-subject-out cross-validation in VR-FOCUS was deliberate. Random resampling at trial level would have allowed participant-specific structure to leak between training and test folds, producing optimistic accuracy estimates that would not reflect realistic deployment performance. The reported classifier performance is therefore conservative and approximates what could be expected for a previously unseen user. The implication for the synthesis is that the observed group difference between healthy controls and chronic pain participants is more likely to reflect a substantive characteristic of the data than an artefact of overfitting. The cost is that some classifiers, particularly larger ensembles such as HIVE-COTE 2.0, required pragmatic adaptations such as contracted folds to remain computationally tractable. These adaptations were managed transparently and do not undermine the core finding, but they reinforce the broader principle that future deployment-relevant machine learning evaluation should weigh predictive performance, computational tractability and validation rigour as connected rather than independent considerations.

8.4 Cumulative limitations of the empirical work

The individual study chapters address study-specific methodological limitations in detail. When considering the studies as an integrated sequence of work, several limitations apply and are worth consideration here.

The studies provide early-stage evidence rather than definitive validation of biomarkers, intervention mechanisms or adaptive control signals. PainWaves used a discordant co-twin design, allowing CWP-related EEG differences to be examined with partial control for shared genetic and environmental influences. However, the analysed sample remained modest for zygosity-stratified, network-level and multivariable analyses. VIPA-EEG was sufficiently powered to detect only moderate-to-large within-subject EEG changes, given it was a secondary analysis of a feasibility dataset. VR-FOCUS supported subject-independent model evaluation but lacked an independent external validation dataset. Across all three studies, therefore, the primary findings are best interpreted as candidate signals and methodological foundations rather than validated clinical biomarkers or deployable adaptive system components.

The cohorts were not fully representative of the populations to whom future adaptive immersive therapeutics would ultimately be offered, with limited phenotyping depth. Chronic pain cohorts across the three studies included predominantly White British, middle to older age females with variable clinical characteristics and symptom severity. This limits generalisation to more diverse populations, high-impact pain cohorts, men, younger adults and people with different chronic musculoskeletal pain phenotypes. PainWaves illustrates the trade-off between pragmatic cohort-based research and detailed clinical phenotyping at the time of assessment. Medication and co-morbidity data improved characterisation, but timing, active disease status, medication indication, adherence and use at EEG acquisition were not fully verified. More broadly, acquired co-morbidity, medication exposure, sleep, fatigue, mood and symptom fluctuation remain individual-level factors that may influence EEG or ocular physiology and are not fully accounted for within the three study designs.

The empirical work provides limited evidence about temporal stability and causal specificity. PainWaves was cross-sectional and therefore cannot determine whether reduced beta power precedes CWP, is a consequence of longstanding pain, tracks symptom fluctuation or reflects treatment-related or co-morbidity-related factors. VIPA-EEG demonstrated short-term EEG change around immersive VR exposure, but without a sham or non-VR comparator these changes cannot be attributed specifically to immersion, interaction, attentional engagement, expectancy, or non-VR task-related effects. VR-FOCUS tested ocular responses during a single

controlled session, so it cannot determine whether cognitive-load inference remains stable across days, symptom states, medication changes, fatigue, sleep variation or repeated home use.

The studies deliberately prioritised pragmatic and translationally plausible acquisition, but this constrained measurement depth. The 32-channel sponge-based EEG system in PainWaves supported scalable acquisition but limited spatial precision and source-level interpretation. VIPA-EEG used a higher-density gel-based EEG system but was constrained by considerations related to study length and participant burden, with shorter retained post-clean recording durations as a result. VR-FOCUS used an integrated consumer-grade VR eye tracking system under controlled luminance and task conditions, whereas real therapeutic VR environments will involve richer visual stimulation, movement, affective engagement and less predictable use contexts. These choices strengthen deployment relevance, but they also mean that the findings require replication under both more controlled mechanistic conditions and more naturalistic therapeutic conditions.

Research conducted did not directly integrate EEG and eye tracking within the same participants or implement a closed-loop adaptive system. The studies were intentionally complementary, with EEG used for mechanistic phenotyping and intervention-related neurophysiological response, and eye tracking used for embedded state-sensitive inference. This supports the conceptual framework, but multimodal integration remains theoretical rather than empirically demonstrated. Similarly, no study tested whether sensor-derived estimates can be used prospectively to adapt VR content and improve engagement, tolerability, pain or function. The work therefore establishes candidate components for future closed-loop VR therapeutics, but not the full adaptive system itself.

Finally, equity and implementation readiness remain underexplored. Digital exclusion and home connectivity are recognised drivers of health inequalities in the United Kingdom[121, 122], and adaptive immersive therapeutics could exacerbate inequalities if access, support and implementation barriers are not considered from the outset. Future deployment would need to account for device availability, digital literacy, sensory tolerance, usability across multiple levels of disability, private space for home use, internet connectivity, and the capacity to engage with repeated sessions without specialist supervision. These factors were not directly evaluated in the

empirical studies, but they are central to whether closed-loop VR therapeutics could be implemented safely, accessibly and without widening existing inequalities.

8.5 Priorities for future research

Several priorities follow the completed empirical work. Replication of the PainWaves beta finding in larger co-twin control studies, with denser EEG montages, source-space modelling and dynamic analyses such as beta-burst quantification is a priority for chronic pain mechanistic research. Combining rsEEG with functional and structural MRI would also be valuable, given the limited evidence linking electrophysiological alterations to brain network and structural changes in chronic pain. Longitudinal designs would clarify whether reduced frontocentral beta tracks symptom persistence, fatigue fluctuation, transition between pain states or treatment response. Future studies should prioritise collection of detailed information related to potential confounders including medication exposure, relevant co-morbidities, sleep, mood, and fatigue to aid interpretation of candidate EEG biomarkers.

Replication of the VIPA-EEG alpha and beta findings in sham-controlled designs would help determine whether the observed modulation is specific to components within the VR intervention including immersion and interaction, or reflects more general post-task or expectancy effects. Further work is also needed to evaluate whether candidate rsEEG features can be measured reliably in repeated home-based or clinical settings, given that scalable longitudinal phenotyping will be necessary for any biomarker-informed adaptive therapeutic.

External validation of the VR-FOCUS classifiers on independent datasets is a priority for the state-sensitive and inferential components of future closed-loop work. The internal cross-validation framework provides a conservative estimate of generalisation but does not establish robustness to changes in headset hardware, eye tracking pipeline, task structure or population. Replication across different chronic musculoskeletal pain phenotypes, including more severe pain populations and stratified subgroups, is similarly needed. Studies that examine the within-person stability of cognitive load classification across time and symptom states would help determine whether classifiers trained at a single timepoint can be expected to remain valid in repeated longitudinal real-world use. Multimodal sensing studies that combine eye tracking with movement, autonomic and where feasible cortical signals

would test whether multimodal integration improves robustness or interpretability of state estimation in chronic pain contexts.

Personalised and hybrid modelling strategies are a related priority. The VR-FOCUS finding that classifier performance was lower in the chronic pain group despite similar mean ocular responses supports the case that population-level models are likely to need to be supplemented with user-specific adaptation. Methodological work is needed to determine the most efficient calibration procedures, the trade-offs between transfer learning and full personalisation, and the implications for governance and reproducibility of models that adapt to individual users.

Prospective studies are also needed to evaluate whether adaptive closed-loop systems improve proximal outcomes such as engagement, tolerability and pain compared with open-loop VR. Systems of this kind belong to the broader family of just-in-time adaptive interventions developed in the mobile health literature[124], in which the timing and content of intervention components are dynamically contingent on the user's changing internal state or context. The microrandomised trial provides an evaluation methodology designed for this class of intervention, estimating the proximal effects of adaptive components and characterising how those effects vary across time, state and context[230]. The framework advanced in the present work maps directly onto the structure such a trial requires. Parallel implementation research is required to address real-time on-device processing, robustness to noisy and missing data, acceptability across diverse populations, and the digital infrastructure needed to support adaptive control outside specialist environments. This work should also test whether adaptive VR can be delivered equitably, evaluating not only technical performance but also who can realistically access, tolerate and sustain use of the system when device availability, digital confidence, connectivity, disability, home environment and remote support vary across users. Closed-loop immersive therapeutics will ultimately require collaboration across clinical pain medicine, neurophysiology, human-computer interaction, machine learning, implementation and regulatory science.

9. Conclusions

9.1 How the empirical findings addressed the aims of this work

The completed empirical work examined EEG and VR eye tracking measures that may be relevant to future adaptive VR interventions for chronic musculoskeletal pain. Chapter 1 identified the three competencies any such system requires: a target, a state measure, and an inference step. The three empirical aims mapped onto this framework.

PainWaves and VIPA-EEG addressed what might be worth targeting, examining rsEEG features at trait level in CWP and in response to immersive VR exposure in people with fibromyalgia. VR-FOCUS addressed the measurement and inference elements, examining whether headset-derived eye tracking can carry usable information about task-defined cognitive load and whether time-series classifiers can predict task level in users the model has not seen before. Each aim was addressed within its own design, with qualifications and limitations set out in the synthesis. The individual components do not in themselves form a working closed-loop system, although findings from the three studies lay the foundations for future work evaluating such systems.

9.2 How I would have approached the research differently

Looking back, the main design limitation with the empirical work was that it did not move earlier towards a more integrated multimodal study design. The studies were conducted within the practical constraints of a PhD programme, including cohort access, equipment availability, governance requirements, data sharing agreements, participant burden, funding and time. Within those constraints, the work was feasible and designed to be methodologically rigorous, but the translational logic became clearer as the studies progressed. The central challenge was not EEG or eye tracking in isolation, but how mechanistic phenotyping and state inference through embedded sensing might eventually combine to form an adaptive therapeutic pathway.

The clearest example is that EEG and eye tracking were collected in separate studies. This was pragmatic, but concurrent or closely paired acquisition in the same participants would have provided a stronger basis for examining whether rsEEG

features, VR-related EEG modulation and ocular markers of cognitive load relate to one another within individuals.

PainWaves also illustrates the trade-offs involved in working within an externally embedded collaborative study. The project was conducted with a team at another institution and integrated within an existing TwinsUK neuroimaging research study, meaning that recruitment, assessment schedules, data access and clinical phenotyping were shaped by wider governance and logistical requirements. More detailed information on medication use, co-morbidities, sleep, current pain distribution and symptom fluctuation would have strengthened interpretation, particularly given the mild on-the-day symptom profile and the importance of zygosity-stratified findings.

In VIPA-EEG, a sham or active-control comparator would have strengthened causal interpretation by helping distinguish VR-specific effects from expectancy, repeated rest or non-specific task effects. In VR-FOCUS, the N-back task provided a controlled way to elicit different levels of cognitive load, but cognitive load was still inferred primarily from task level rather than measured more directly. Including a broader range of VR tasks, together with additional physiological signals such as heart rate variability, electrodermal activity or movement features, could have allowed cognitive load to be characterised more comprehensively and tested in contexts closer to therapeutic VR use.

Finally, greater emphasis on diversity, longitudinal signal stability and implementation would have strengthened the work. The chronic pain cohorts were not entirely representative of the wider population, most measurements were single-session, and all machine learning analyses were offline without consideration of processing closer to acquisition as would be required by a closed-loop system. These additions were beyond the realistic scope of the current PhD, but they define clear priorities for the next stage of work, whilst also considering key advancements in the field that have been realised since this work began.

9.3 How the field has changed

Since the start of the PhD in 2023, several related technologies have developed substantially, particularly in the field of artificial intelligence following the public launch of large language models in late 2022.

Closed-loop technologies have started to move from research to clinical application. An example is the Medtronic BrainSense™ adaptive system, which adjusts deep brain stimulation in real time from beta band local field potentials in Parkinson's disease[364], gaining CE Mark approvals in 2025. Akili's EndeavorOTC adaptive cognitive training game received FDA approval for ADHD in 2024, using performance data captured whilst completing a mobile-based game to adjust difficulty level in real time and provide feedback to the user through a 'focus score'[365]. In the research domain VR closed-loop systems are now emerging, with the first real-time adaptive cognitive load intervention designed for cognitive training reported in 2025[14].

Consumer immersive hardware has also evolved considerably. The Pico Neo 3 Pro Eye used in VR-FOCUS has been followed by newer standalone devices with higher resolution displays, mixed-reality passthrough, improved onboard processing and more mature integrated tracking[366-368]. Lightweight smart-glasses and augmented reality display glasses have moved into the mainstream, including Meta Ray-Ban Display, Snap Spectacles and XREAL devices, in form factors closer to ordinary eyewear[220, 369, 370]. Peripheral technologies to improve user interaction have also emerged, including Meta's EMG-based neural wristband for gesture and micro-gesture interpretation[220, 221], which points toward physiological signals contributing to both intent inference and state estimation from outside the headset itself.

Above all, artificial intelligence has seen an unprecedented shift. Since the release of ChatGPT in late 2022[371, 372], large language models have moved through multimodal variants that process image, audio and numerical data alongside text, reasoning-focused models that plan over multi-step problems before responding, and agentic frameworks in which models call tools, query external systems and act over extended tasks rather than producing single replies[373]. EEG foundation models such as LaBraM[374] and Neuro-GPT[375] now offer fine-tuning routes for the small clinical datasets typical of chronic pain research, time-series foundation models including Chronos[376] are beginning to compete with bespoke classifiers, and efficient ROCKET variants such as HYDRA[377] have narrowed the gap between accuracy and on-device inference. The next year is likely to bring further maturation of biosignal foundation models and agentic systems, both of which would directly support the kind of architecture a closed-loop VR system for chronic pain would need.

9.4 The next phase of research: towards clinically deployable closed-loop VR therapeutics

Moving this work from foundational evidence toward clinical deployment requires a staged programme rather than a single next step.

The first stage is to strengthen each empirical component within its own remit. The validation and replication requirements for PainWaves, VIPA-EEG and VR-FOCUS were set out in Chapter 8 (Section 8.5). The purpose at this stage of the translational pathway is to bring each signal stream to a level of robustness sufficient to determine its eventual role in a closed-loop pathway, rather than to integrate the components prematurely.

The next priority would be to build and evaluate a closed-loop prototype. This requires resolving how trait-level and within-session state measurement fit together in a potential treatment pathway. A plausible architecture would acquire rsEEG at baseline to characterise the user, deliver an immersive intervention with within-session adaptation driven by headset-derived ocular signals, and reacquire rsEEG at defined intervals to test whether VR produces measurable longer-term neuroplastic change. Microrandomised trial methodology[230] or similar approaches would then be used to evaluate the proximal effects of adaptive decisions across time, state and context before any definitive clinical trial is conducted against open-loop VR.

Following closed-loop system development, evaluation of clinical viability and safety is essential. Closed-loop systems raise safety considerations beyond those of conventional VR interventions, particularly where adaptation is driven by physiological estimates that may be miscalibrated or transiently unreliable. Home-based immersive therapeutics also carry risks of worsening health inequalities driven by factors including digital exclusion[121, 122]. Implementation will also require the digital infrastructure and interdisciplinary collaboration needed to support adaptive control outside specialist environments. Therefore, clinical translation will require co-design with patients and clinicians, early regulatory engagement, and equity built into deployment from the start rather than added in retrospect.

9.5 Principal conclusions

Across the completed work, beta band activity emerged as the most consistent candidate rsEEG domain for further investigation in chronic musculoskeletal pain. In

PainWaves, reduced frontocentral beta-family power was observed in a co-twin control design, suggesting that at least part of the beta-range phenotype is associated with CWP rather than being explained only by shared genetic or environmental background. In VIPA-EEG, beta-family power also changed around immersive VR exposure in fibromyalgia, although the aperiodic-adjusted analyses appropriately temper the strength of the claim that this reflects a specifically periodic beta oscillatory effect. These findings do not establish beta as a diagnostic biomarker, but they support further mechanistic study.

Alpha activity showed a different pattern. Rather than emerging as a stable trait-level marker of pain, alpha was most clearly modulated in VIPA-EEG following VR exposure and appeared more consistent with short-term within-session state change. This suggests that alpha may be better understood as a marker of post-exposure neurophysiological state or intervention responsiveness than as a trait-defining pain biomarker in the contexts studied here.

VR-FOCUS showed that headset-integrated eye tracking can support subject-independent classification of task-defined cognitive load in previously unseen users. However, performance was lower and more variable in the chronic pain group, despite broadly preserved group-level ocular responses to task demand. This indicates that aggregate physiological responses may appear similar across groups while the temporal structure required for prediction is less stable in chronic pain. Future closed-loop VR systems should therefore treat cohort-specific and user-specific variability as a core design problem, not as a secondary refinement.

Empirical study findings support a central thesis framework in which future closed-loop VR for chronic musculoskeletal pain will require different forms of evidence for different components of the system. rsEEG is best positioned as a mechanistic guide for closed-loop therapeutics, providing information that can support patient stratification and neurophysiological targets for interventional approaches. Embedded eye tracking is better suited to real-time state-sensitive sensing, informing personalised adaptive systems that respond to the user. Machine learning provides the analytical layer that takes sensor signals, predicts cognitive state, and drives adaptive logic.

This work therefore provides preliminary empirical evidence and a methodological framework for components that may contribute to future closed-loop VR therapeutics in chronic musculoskeletal pain. It does not provide a working adaptive therapeutic system, nor does it establish any signal as ready for clinical deployment. Translation will require concurrent multimodal acquisition, more diverse and longitudinal cohorts, real-time implementation studies, and prospective evaluation of adaptive control under realistic deployment conditions. The central contribution of this work is to clarify the signal roles, methodological requirements and evidential gaps that future adaptive immersive therapeutics will need to address.

10. References

1. Arthritis UK. The State of Musculoskeletal Health 2025: Arthritis and Other Musculoskeletal Conditions in Numbers. Chesterfield, UK: Arthritis UK; 2025.
2. Fayaz A, Croft P, Langford RM, Donaldson LJ, Jones GT. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. BMJ Open. 2016;6(6):e010364.
3. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204-22.
4. National Institute for Health and Care Excellence. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. National Institute for Health and Care Excellence (NICE); 2021.
5. Tsigarides J, Sabu BK, Al-Amri M, Hughes S, Fuggle N. Therapeutic Virtual Reality in Rheumatology: A Narrative Review. touchREVIEWS in RMD. 2026;5(1):26-33.
6. Tsigarides J, Shenker NG, MacGregor A. Ready Patient One: the role of therapeutic virtual reality in the future management of chronic pain. Rheumatology (Oxford). 2022;61(2):482-3.
7. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res. 2019;12:2053-85.
8. Tsigarides J, Grove V, Chipping J, Dainty J, Miles S, Shenker N, et al. Feasibility of attention-based virtual reality interventions in fibromyalgia syndrome: comparing systems, virtual environments and activities. Br J Pain. 2025;19(2):71-85.

9. Fillingim RB. Individual differences in pain: understanding the mosaic that makes pain personal. Pain. 2017;158 Suppl 1(Suppl 1):S11-s8.
10. Hutchings A, Calloway M, Choy E, Hooper M, Hunter DJ, Jordan JM, et al. The Longitudinal Examination of Arthritis Pain (LEAP) study: relationships between weekly fluctuations in patient-rated joint pain and other health outcomes. J Rheumatol. 2007;34(11):2291-300.
11. Toussaint L, Vincent A, McAllister SJ, Whipple M. Intra- and Inter-Patient Symptom Variability in Fibromyalgia: Results of a 90-Day Assessment. Musculoskeletal Care. 2015;13(2):93-100.
12. Rice DA, Ozolins C, Biswas R, Almesfer F, Zeng I, Parikh A, et al. Home-based EEG Neurofeedback for the Treatment of Chronic Pain: A Randomized Controlled Clinical Trial. The Journal of Pain. 2024;25(11):104651.
13. Schuurman BB, Lousberg RL, Schreiber JU, van Amelsvoort T, Vossen CJ. A Scoping Review of the Effect of EEG Neurofeedback on Pain Complaints in Adults with Chronic Pain. J Clin Med. 2024;13(10).
14. Szczepaniak D, Harvey M, Deligianni F. Your Eyes Controlled the Game: Real-Time Cognitive Training Adaptation based on Eye-Tracking and Physiological Data in Virtual Reality. arXiv. 2025.
15. Zhang S, Yoshida W, Mano H, Yanagisawa T, Mancini F, Shibata K, et al. Pain Control by Co-adaptive Learning in a Brain-Machine Interface. Current Biology. 2020;30(20):3935-44.e7.
16. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain. 2015;156(6):1003-7.
17. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163-96.
18. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet. 2021;397(10289):2082-97.
19. Blyth FM, March LM, Brnabic AJ, Cousins MJ. Chronic pain and frequent use of health care. Pain. 2004;111(1-2):51-8.

20. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskeletal Disorders. 2010;11(1):144.
21. Ojeda B, Salazar A, Dueñas M, Torres LM, Micó JA, Failde I. The impact of chronic pain: the perspective of patients, relatives, and caregivers. Fam Syst Health. 2014;32(4):399-407.
22. Smith T, Mansfield M, Hanson S, Welsh A, Khoury R, Clark A, et al. Caregiving for older people living with chronic pain: analysis of the English longitudinal study of ageing and health survey for England. British Journal of Pain. 2023;17(2):166-81.
23. Tsuji H, Tetsunaga T, Tetsunaga T, Misawa H, Oda Y, Takao S, et al. Factors influencing caregiver burden in chronic pain patients: A retrospective study. Medicine. 2022;101(39).
24. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for Chronic Noncancer Pain: A Systematic Review and Meta-analysis. JAMA. 2018;320(23):2448-60.
25. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database of Systematic Reviews. 2017(4).
26. Williams ACdC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database of Systematic Reviews. 2020(8).
27. Enthoven WTM, Roelofs P, Deyo RA, van Tulder MW, Koes BW. Non-steroidal anti-inflammatory drugs for chronic low back pain. Cochrane Database of Systematic Reviews. 2016(2).
28. Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Häuser W, Fluß E, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis. 2017;76(2):318-28.
29. Chou R, Turner JA, Devine EB, Hansen RN, Sullivan SD, Blazina I, et al. The Effectiveness and Risks of Long-Term Opioid Therapy for Chronic Pain: A Systematic Review for a National Institutes of Health Pathways to Prevention Workshop. Annals of Internal Medicine. 2015;162(4):276-86.
30. Els C, Jackson TD, Kunyk D, Lappi VG, Sonnenberg B, Hagtvedt R, et al. Adverse events associated with medium- and long-term use of opioids for chronic non-cancer pain: an overview of Cochrane Reviews. Cochrane Database of Systematic Reviews. 2017(10).

31. Becker WC, Dorflinger L, Edmond SN, Islam L, Heapy AA, Fraenkel L. Barriers and facilitators to use of non-pharmacological treatments in chronic pain. BMC Family Practice. 2017;18(1):41.

32. Goldsmith ES, Miller WA, Koffel E, Ullman K, Landsteiner A, Stroebel B, et al. Barriers and Facilitators of Evidence-Based Psychotherapies for Chronic Pain in Adults: A Systematic Review. J Pain. 2023;24(5):742-69.

33. Ashaye T, Hounsome N, Carnes D, Taylor SJC, Homer K, Eldridge S, et al. Opioid prescribing for chronic musculoskeletal pain in UK primary care: results from a cohort analysis of the COPERS trial. BMJ Open. 2018;8(6):e019491.

34. Scott IC, Bailey J, White CR, Mallen CD, Muller S. Analgesic prescribing in patients with inflammatory arthritis in England: an observational study using electronic healthcare record data. Rheumatology (Oxford). 2022;61(8):3201-11.

35. Watson JA, Ryan CG, Atkinson G, Williamson P, Ellington D, Whittle R, et al. Inter-Individual Differences in the Responses to Pain Neuroscience Education in Adults With Chronic Musculoskeletal Pain: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. J Pain. 2021;22(1):9-20.

36. Whibley D, Williams DA, Clauw DJ, Sliwinski MJ, Kratz AL. Within-day rhythms of pain and cognitive function in people with and without fibromyalgia: synchronous or syncopated? Pain. 2022;163(3):474-82.

37. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull. 2007;133(4):581-624.

38. Engel GL. The Need for a New Medical Model: A Challenge for Biomedicine. Science. 1977;196(4286):129-36.

39. Vlaeyen JWS, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. PAIN. 2000;85(3).

40. Kosek E, Cohen M, Baron R, Gebhart GF, Mico JA, Rice ASC, et al. Do we need a third mechanistic descriptor for chronic pain states? Pain. 2016;157(7):1382-6.

41. Clauw DJ. Fibromyalgia: A Clinical Review. JAMA. 2014;311(15):1547-55.

42. Kelleher EM, Lange F, Wanigasekera V, Rathod-Mistry T, Nichols T, Seymour B, et al. Brain signatures of nociplastic pain: Fibromyalgia Index and descending modulation at population level. Brain. 2026;149(4):1365-80.

43. Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. Neuron. 2007;55(3):377-91.
44. Zebhauser PT, Hohn VD, Ploner M. Resting-state electroencephalography and magnetoencephalography as biomarkers of chronic pain: a systematic review. Pain. 2023;164(6):1200-21.
45. Seymour B. Pain: A Precision Signal for Reinforcement Learning and Control. Neuron. 2019;101(6):1029-41.
46. Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, et al. Temporal difference models describe higher-order learning in humans. Nature. 2004;429(6992):664-7.
47. Quartana PJ, Campbell CM, Edwards RR. Pain catastrophizing: a critical review. Expert Review of Neurotherapeutics. 2009;9(5):745-58.
48. Karimi R, Mallah N, Scherer R, Rodríguez-Cano R, Takkouche B. Sleep quality as a mediator of the relation between depression and chronic pain: a systematic review and meta-analysis. British Journal of Anaesthesia. 2023;130(6):747-62.
49. Werneck AO, Stubbs B. Bidirectional relationship between chronic pain and depressive symptoms in middle-aged and older adults. General Hospital Psychiatry. 2024;89:49-54.
50. Bushnell MC, Čeko M, Low LA. Cognitive and emotional control of pain and its disruption in chronic pain. Nature Reviews Neuroscience. 2013;14(7):502-11.
51. Eccleston C, Crombez G. Pain demands attention: a cognitive-affective model of the interruptive function of pain. Psychol Bull. 1999;125(3):356-66.
52. Wiech K. Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. Science. 2016;354(6312):584-7.
53. Melzack R, Wall PD. Pain mechanisms: a new theory. Science. 1965;150(3699):971-9.
54. Millan MJ. Descending control of pain. Prog Neurobiol. 2002;66(6):355-474.
55. Bantick SJ, Wise RG, Ploghaus A, Clare S, Smith SM, Tracey I. Imaging how attention modulates pain in humans using functional MRI. Brain. 2002;125(Pt 2):310-9.
56. Legrain V, Crombez G, Mouraux A. Controlling attention to nociceptive stimuli with working memory. PLoS One. 2011;6(6):e20926.
57. Sprenger C, Eippert F, Finsterbusch J, Bingel U, Rose M, Büchel C. Attention modulates spinal cord responses to pain. Curr Biol. 2012;22(11):1019-22.

58. Wiech K, Ploner M, Tracey I. Neurocognitive aspects of pain perception. Trends in Cognitive Sciences. 2008;12(8):306-13.
59. Seminowicz DA, Davis KD. Pain Enhances Functional Connectivity of a Brain Network Evoked by Performance of a Cognitive Task. Journal of Neurophysiology. 2007;97(5):3651-9.
60. Seminowicz DA, Davis KD. Interactions of Pain Intensity and Cognitive Load: The Brain Stays on Task. Cerebral Cortex. 2007;17(6):1412-22.
61. Moriarty O, McGuire BE, Finn DP. The effect of pain on cognitive function: a review of clinical and preclinical research. Prog Neurobiol. 2011;93(3):385-404.
62. Bingel U, Tracey I. Imaging CNS modulation of pain in humans. Physiology (Bethesda). 2008;23:371-80.
63. Ossipov MH, Morimura K, Porreca F. Descending pain modulation and chronification of pain. Curr Opin Support Palliat Care. 2014;8(2):143-51.
64. Atlas LY, Wager TD. How expectations shape pain. Neurosci Lett. 2012;520(2):140-8.
65. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci. 2009;364(1535):3549-57.
66. Slater M, Sanchez-Vives MV. Enhancing Our Lives with Immersive Virtual Reality. Frontiers in Robotics and AI. 2016;Volume 3 - 2016.
67. Clay V, König P, König S. Eye Tracking in Virtual Reality. J Eye Mov Res. 2019;12(1).
68. Eccleston C, Fisher E, Liikkanen S, Sarapohja T, Stenfors C, Jääskeläinen SK, et al. A prospective, double-blind, pilot, randomized, controlled trial of an "embodied" virtual reality intervention for adults with low back pain. Pain. 2022;163(9):1700-15.
69. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-Administered Skills-Based Virtual Reality Intervention for Chronic Pain: Randomized Controlled Pilot Study. JMIR Form Res. 2020;4(7):e17293.
70. Keefe FJ, Huling DA, Coggins MJ, Keefe DF, Rosenthal ZM, Herr NR, et al. Virtual reality for persistent pain: a new direction for behavioral pain management. Pain. 2012;153(11):2163-6.
71. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. JMIR Ment Health. 2019;6(1):e11973.

72. Hoffman HG, Richards TL, Coda B, Bills AR, Blough D, Richards AL, et al. Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI. Neuroreport. 2004;15(8):1245-8.
73. Lier EJ, Oosterman JM, Assmann R, de Vries M, van Goor H. The effect of Virtual Reality on evoked potentials following painful electrical stimuli and subjective pain. Scientific reports. 2020;10(1):9067.
74. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. PLoS One. 2019;14(8):e0219115.
75. Dreesmann NJ, Su H, Thompson HJ. A Systematic Review of Virtual Reality Therapeutics for Acute Pain Management. Pain Manag Nurs. 2022;23(5):672-81.
76. Teh JJ, Pascoe DJ, Hafeji S, Parchure R, Koczoski A, Rimmer MP, et al. Efficacy of virtual reality for pain relief in medical procedures: a systematic review and meta-analysis. BMC medicine. 2024;22(1):64.
77. Jones T, Moore T, Choo J. The Impact of Virtual Reality on Chronic Pain. PloS one. 2016;11(12):e0167523.
78. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality Program for Chronic Low Back Pain: Double-Blind, Randomized, Placebo-Controlled Trial Conducted During COVID-19. J Med Internet Res. 2021;23(2):e26292.
79. Maddox T, Oldstone L, Sackman J, Maddox R, Adair T, Ffrench K, et al. Twelve-month results for a randomized sham-controlled effectiveness trial of an in-home skills-based virtual reality program for chronic low back pain. Pain Rep. 2024;9(5):e1182.
80. Maddox T, Sackman J, Bonakdar R, Maddox R, Darnall BD. Skill-Based Virtual Reality Therapy in High-Impact Chronic Pain: 2-Year Follow-Up Results From a Secondary Analysis of a Randomized Controlled Trial. J Med Internet Res. 2026;28:e90688.
81. Maddox T, Sparks C, Oldstone L, Maddox R, Ffrench K, Garcia H, et al. Durable chronic low back pain reductions up to 24 months after treatment for an accessible, 8-week, in-home behavioral skills-based virtual reality program: a randomized controlled trial. Pain medicine (Malden, Mass). 2023;24(10):1200-3.

82. Goudman L, Jansen J, Billot M, Vets N, De Smedt A, Roulaud M, et al. Virtual Reality Applications in Chronic Pain Management: Systematic Review and Meta-analysis. JMIR Serious Games. 2022;10(2):e34402.
83. Medina S, Clarke S, Hughes S. Virtual reality-based analgesia: towards a novel framework for the biopsychosocial management of chronic pain. British Journal of Anaesthesia. 2024;133(3):486-90.
84. Spiegel BMR, Eberlein SA, Persky S, Ishimori ML, Venuturupalli S, Tu J, et al. Randomized-controlled trial of skills-based vr vs. distraction vr vs. sham VR for chronic low back pain. NPJ Digit Med. 2026;9(1).
85. Lier EJ, de Vries M, Steggink EM, Ten Broek RPG, van Goor H. Effect modifiers of virtual reality in pain management: a systematic review and meta-regression analysis. Pain. 2023;164(8):1658-65.
86. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: current trends and future directions. Pain Manag. 2011;1(2):147-57.
87. Deng X, Jian C, Yang Q, Jiang N, Huang Z, Zhao S. The analgesic effect of different interactive modes of virtual reality: A prospective functional near-infrared spectroscopy (fNIRS) study. Front Neurosci. 2022;16:1033155.
88. Law EF, Dahlquist LM, Sil S, Weiss KE, Herbert LJ, Wohlheiter K, et al. Videogame distraction using virtual reality technology for children experiencing cold pressor pain: the role of cognitive processing. J Pediatr Psychol. 2011;36(1):84-94.
89. Wender R, Hoffman HG, Hunner HH, Seibel EJ, Patterson DR, Sharar SR. INTERACTIVITY INFLUENCES THE MAGNITUDE OF VIRTUAL REALITY ANALGESIA. J Cyber Ther Rehabil. 2009;2(1):27-33.
90. Xiang H, Shen J, Wheeler KK, Patterson J, Lever K, Armstrong M, et al. Efficacy of Smartphone Active and Passive Virtual Reality Distraction vs Standard Care on Burn Pain Among Pediatric Patients: A Randomized Clinical Trial. JAMA Network Open. 2021;4(6):e2112082-e.
91. Raghuraman N, Shafir R, Colloca G, Kier C, Brawn B, Varshney A, et al. A critical role of affective content in the analgesic effect of virtual reality: a cross-sectional within-subject study. Lancet Reg Health Am. 2026;55:101385.
92. Petkova VI, Ehrsson HH. If I were you: perceptual illusion of body swapping. PLoS One. 2008;3(12):e3832.

93. Slater M, Spanlang B, Sanchez-Vives MV, Blanke O. First person experience of body transfer in virtual reality. PLoS One. 2010;5(5):e10564.
94. Kilteni K, Groten R, Slater M. The sense of embodiment in virtual reality. Presence: Teleoper Virtual Environ. 2012;21(4):373–87.
95. Lewis JS, Kersten P, McCabe CS, McPherson KM, Blake DR. Body perception disturbance: a contribution to pain in complex regional pain syndrome (CRPS). Pain. 2007;133(1-3):111-9.
96. Matamala-Gomez M, Diaz Gonzalez AM, Slater M, Sanchez-Vives MV. Decreasing Pain Ratings in Chronic Arm Pain Through Changing a Virtual Body: Different Strategies for Different Pain Types. The journal of pain. 2019;20(6):685-97.
97. Solca M, Ronchi R, Bello-Ruiz J, Schmidlin T, Herbelin B, Luthi F, et al. Heartbeat-enhanced immersive virtual reality to treat complex regional pain syndrome. Neurology. 2018;91(5):e479-e89.
98. Leeuw M, Goossens ME, Linton SJ, Crombez G, Boersma K, Vlaeyen JW. The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. J Behav Med. 2007;30(1):77-94.
99. Groenveld TD, Smits ML, Knoop J, Kallewaard JW, Staal JB, de Vries M, et al. Effect of a Behavioral Therapy-Based Virtual Reality Application on Quality of Life in Chronic Low Back Pain. The Clinical journal of pain. 2023;39(6):278-85.
100. Sarig Bahat H, Takasaki H, Chen X, Bet-Or Y, Treleaven J. Cervical kinematic training with and without interactive VR training for chronic neck pain - a randomized clinical trial. Manual therapy. 2015;20(1):68-78.
101. Slatman S, Ostelo R, van Goor H, Staal JB, Knoop J. Physiotherapy with integrated virtual reality for patients with complex chronic low back pain: protocol for a pragmatic cluster randomized controlled trial (VARIETY study). BMC musculoskeletal disorders. 2023;24(1):132.
102. Tejera DM, Beltran-Alacreu H, Cano-de-la-Cuerda R, Leon Hernandez JV, Martin-Pintado-Zugasti A, Calvo-Lobo C, et al. Effects of Virtual Reality versus Exercise on Pain, Functional, Somatosensory and Psychosocial Outcomes in Patients with Non-specific Chronic Neck Pain: A Randomized Clinical Trial. International journal of environmental research and public health. 2020;17(16).

103. Yilmaz Yelvar GD, Cirak Y, Dalkilinc M, Parlak Demir Y, Guner Z, Boydak A. Is physiotherapy integrated virtual walking effective on pain, function, and kinesiophobia in patients with non-specific low-back pain? Randomised controlled trial. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2017;26(2):538-45.
104. Garcia LM, Birckhead BJ, Krishnamurthy P, Mackey I, Sackman J, Salmasi V, et al. Three-Month Follow-Up Results of a Double-Blind, Randomized Placebo-Controlled Trial of 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality (VR) for Chronic Low Back Pain. J Pain. 2022;23(5):822-40.
105. Vogel TA, Falk CF, Otto AR, Roy M. Distraction from pain depends on task demands and motivation. Pain Rep. 2022;7(6):e1041.
106. Tabry V, Vogel TA, Lussier M, Brouillard P, Buhle J, Rainville P, et al. Inter-individual predictors of pain inhibition during performance of a competing cognitive task. Scientific Reports. 2020;10(1):21785.
107. Slatman S, Harvie D, van der Heijden M, Bulle L, Dejaco B, van Goor H, et al. Reporting Guideline for Therapeutic Immersive Virtual Reality Interventions: International eDelphi Study (INVIRTUE). Journal of Medical Extended Reality. 2025;2(1):29941520251404744.
108. Giacomelli L, Martin Sölch C, Ledermann K. The effect of virtual reality interventions on reducing pain intensity in chronic pain patients: a systematic review. Virtual Reality. 2024;28(3):126.
109. Vlake JH, van Bommel J, Riva G, Wiederhold BK, Cipresso P, Rizzo AS, et al. Reporting the early stage clinical evaluation of virtual-reality-based intervention trials: RATE-VR. Nature Medicine. 2023;29(1):12-3.
110. United States Food and Drug Administration (FDA). De Novo Classification Request for EaseVRx (DEN210014) Decision Summary. Silver Spring, MD: U.S. Food and Drug Administration; 2021.
111. Caserman P, Garcia-Agundez A, Gámez Zerban A, Göbel S. Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. Virtual Reality. 2021;25(4):1153-70.

112. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. Front Hum Neurosci. 2020;14:96.
113. Riccio GE, Stoffregen TA. An ecological Theory of Motion Sickness and Postural Instability. Ecol Psychol. 1991;3(3):195-240.
114. Stanney K, Lawson BD, Rokers B, Dennison M, Fidopiastis C, Stoffregen T, et al. Identifying Causes of and Solutions for Cybersickness in Immersive Technology: Reformulation of a Research and Development Agenda. International Journal of Human–Computer Interaction. 2020;36(19):1783-803.
115. Hoffman DM, Girshick AR, Akeley K, Banks MS. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Journal of Vision. 2008;8(3):33.
116. Knight JF, Baber C. Effect of Head-Mounted Displays on Posture. Human Factors. 2007;49(5):797-807.
117. Aardema F, O'Connor K, Côté S, Taillon A. Virtual Reality Induces Dissociation and Lowers Sense of Presence in Objective Reality. Cyberpsychol Behav Soc Netw. 2010;13(4):429-35.
118. Astek A, Sparkes V, Sheeran L. Exploring the use of immersive virtual reality in adults with chronic primary pain: A scoping review. DIGITAL HEALTH. 2024;10.
119. Kouijzer M, Kip H, Bouman YHA, Kelders SM. Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. Implement Sci Commun. 2023;4(1):67.
120. Dy M, Olazo K, Lisker S, Brown E, Saha A, Weinberg J, et al. Virtual Reality for Chronic Pain Management Among Historically Marginalized Populations: Systematic Review of Usability Studies. J Med Internet Res. 2023;25:e40044.
121. NHS England. Inclusive digital healthcare: a framework for NHS action on digital inclusion. London: NHS England; 2023.
122. Ofcom. Adults' media use and attitudes report 2024. London: Ofcom; 2024.
123. Mirza KB, Golden CT, Nikolic K, Toumazou C. Closed-Loop Implantable Therapeutic Neuromodulation Systems Based on Neurochemical Monitoring. Frontiers in Neuroscience. 2019;Volume 13 - 2019.

124. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med. 2018;52(6):446-62.
125. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061.
126. Medical Research Council. The MRC Framework for the Development, Design and Analysis of Stratified Medicine Research. Swindon: Medical Research Council; 2018. Report No.: 978-0-903730-28-0.
127. Pinheiro ES, de Queirós FC, Montoya P, Santos CL, do Nascimento MA, Ito CH, et al. Electroencephalographic Patterns in Chronic Pain: A Systematic Review of the Literature. PLoS One. 2016;11(2):e0149085.
128. Bott FS, Zebhauser PT, Hohn VD, Turgut Ö, May ES, Tiemann L, et al. Exploring electroencephalographic chronic pain biomarkers: a mega-analysis. EBioMedicine. 2025;120:105955.
129. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nat Rev Neurol. 2020;16(7):381-400.
130. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD) and Bethesda (MD): Food and Drug Administration (US) and National Institutes of Health (US); 2016.
131. Collins LM, Murphy SA, Bierman KL. A Conceptual Framework for Adaptive Preventive Interventions. Prev Sci. 2004;5(3):185-96.
132. Murphy SA. An experimental design for the development of adaptive treatment strategies. Stat Med. 2005;24(10):1455-81.
133. Kyle BN, McNeil DW. Autonomic arousal and experimentally induced pain: a critical review of the literature. Pain Res Manag. 2014;19(3):159-67.
134. Hopstaken JF, van der Linden D, Bakker AB, Kompier MAJ. The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. Biological Psychology. 2015;110:100-6.
135. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. 2021;2(3):160.

136. Čeko M, Frangos E, Gracely J, Richards E, Wang B, Schweinhardt P, et al. Default mode network changes in fibromyalgia patients are largely dependent on current clinical pain. Neuroimage. 2020;216:116877.
137. Ichesco E, Schmidt-Wilcke T, Bhavsar R, Clauw DJ, Peltier SJ, Kim J, et al. Altered resting state connectivity of the insular cortex in individuals with fibromyalgia. J Pain. 2014;15(8):815-26.e1.
138. Napadow V, Kim J, Clauw DJ, Harris RE. Brief Report: Decreased intrinsic brain connectivity is associated with reduced clinical pain in fibromyalgia. Arthritis & Rheumatism. 2012;64(7):2398-403.
139. Napadow V, LaCount L, Park K, As-Sanie S, Clauw DJ, Harris RE. Intrinsic brain connectivity in fibromyalgia is associated with chronic pain intensity. Arthritis Rheum. 2010;62(8):2545-55.
140. López-Solà M, Woo CW, Pujol J, Deus J, Harrison BJ, Monfort J, et al. Towards a neurophysiological signature for fibromyalgia. Pain. 2017;158(1):34-47.
141. Medina S, Hughes SW. Immersion in nature through virtual reality attenuates the development and spread of mechanical secondary hyperalgesia: a role for insulo-thalamic effective connectivity. PAIN. 2025;166(9).
142. Ploner M, Sorg C, Gross J. Brain Rhythms of Pain. Trends Cogn Sci. 2017;21(2):100-10.
143. Vanneste S, De Ridder D. Chronic pain as a brain imbalance between pain input and pain suppression. Brain Commun. 2021;3(1):fcab014.
144. Choe MK, Lim M, Kim JS, Lee DS, Chung CK. Disrupted Resting State Network of Fibromyalgia in Theta frequency. Scientific Reports. 2018;8(1):2064.
145. Fallon N, Chiu Y, Nurmikko T, Stancak A. Altered theta oscillations in resting EEG of fibromyalgia syndrome patients. Eur J Pain. 2018;22(1):49-57.
146. Sarnthein J, Stern J, Aufenberg C, Rousson V, Jeanmonod D. Increased EEG power and slowed dominant frequency in patients with neurogenic pain. Brain. 2006;129(Pt 1):55-64.
147. Silva-Passadouro B, Tamasauskas A, Khoja O, Casson AJ, Delis I, Brown C, et al. A systematic review of quantitative EEG findings in Fibromyalgia, Chronic Fatigue Syndrome and Long COVID. Clin Neurophysiol. 2024;163:209-22.
148. Furman AJ, Thapa T, Summers SJ, Cavaleri R, Fogarty JS, Steiner GZ, et al. Cerebral peak alpha frequency reflects average pain severity in a human model of sustained, musculoskeletal pain. J Neurophysiol. 2019;122(4):1784-93.

149. Ahn S, Prim JH, Alexander ML, McCulloch KL, Fröhlich F. Identifying and Engaging Neuronal Oscillations by Transcranial Alternating Current Stimulation in Patients With Chronic Low Back Pain: A Randomized, Crossover, Double-Blind, Sham-Controlled Pilot Study. J Pain. 2019;20(3):277.e1-.e11.
150. Halpin SJ, Xing L, Greenwood DC, Tang N, Trujillo-Barreto NJ, Brown CA, et al. Pre-sleep alpha brain entrainment by audio or visual stimulation for chronic widespread pain and sleep disturbance: A randomised crossover feasibility trial. J Pain. 2025;31:105393.
151. Barone J, Rossiter HE. Understanding the Role of Sensorimotor Beta Oscillations. Front Syst Neurosci. 2021;15:655886.
152. Engel AK, Fries P. Beta-band oscillations--signalling the status quo? Curr Opin Neurobiol. 2010;20(2):156-65.
153. Alves RL, Zortea M, Serrano PV, Brugnera Tomedi R, Pereira de Almeida R, Torres ILS, et al. High-beta oscillations at EEG resting state and hyperconnectivity of pain circuitry in fibromyalgia: an exploratory cross-sectional study. Front Neurosci. 2023;17:1233979.
154. González-Villar AJ, Triñanes Y, Gómez-Perretta C, Carrillo-de-la-Peña MT. Patients with fibromyalgia show increased beta connectivity across distant networks and microstates alterations in resting-state electroencephalogram. Neuroimage. 2020;223:117266.
155. Camargo L, Pacheco-Barrios K, Marques LM, Caumo W, Fregni F. Adaptive and Compensatory Neural Signatures in Fibromyalgia: An Analysis of Resting-State and Stimulus-Evoked EEG Oscillations. Biomedicines. 2024;12(7).
156. Teixeira M, Mancini C, Wicht CA, Maestretti G, Kuntzer T, Cazzoli D, et al. Beta Electroencephalographic Oscillation Is a Potential GABAergic Biomarker of Chronic Peripheral Neuropathic Pain. Front Neurosci. 2021;15:594536.
157. Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, et al. Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci. 2020;23(12):1655-65.
158. Gil Avila C, May ES, Bott FS, Tiemann L, Hohn V, Heitmann H, et al. Assessing the balance between excitation and inhibition in chronic pain through the aperiodic component of EEG. Elife. 2025;13.
159. Malone SM, Burwell SJ, Vaidyanathan U, Miller MB, McGue M, Iacono WG. Heritability and molecular-genetic basis of resting EEG activity: a genome-wide association study. Psychophysiology. 2014;51(12):1225-45.

160. Schutte NM, Hansell NK, de Geus EJ, Martin NG, Wright MJ, Smit DJ. Heritability of resting state EEG functional connectivity patterns. Twin Res Hum Genet. 2013;16(5):962-9.
161. van Beijsterveldt CE, Molenaar PC, de Geus EJ, Boomsma DI. Heritability of human brain functioning as assessed by electroencephalography. Am J Hum Genet. 1996;58(3):562-73.
162. van Beijsterveldt CE, van Baal GC. Twin and family studies of the human electroencephalogram: a review and a meta-analysis. Biol Psychol. 2002;61(1-2):111-38.
163. Kato K, Sullivan PF, Evengård B, Pedersen NL. Importance of genetic influences on chronic widespread pain. Arthritis Rheum. 2006;54(5):1682-6.
164. Allen-Brady K, Fyer AJ, Weissman M. The multi-generational familial aggregation of interstitial cystitis, other chronic nociplastic pain disorders, depression, and panic disorder. Psychological Medicine. 2023;53(16):7847-56.
165. Gonggrijp BMA, van de Weijer SGA, Bijleveld CCJH, van Dongen J, Boomsma DI. The Co-Twin Control Design: Implementation and Methodological Considerations. Twin Research and Human Genetics. 2023;26(4-5):249-56.
166. Adhanom IB, MacNeilage P, Folmer E. Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. Virtual Real. 2023;27(2):1481-505.
167. Hou BJ, Abdrabou Y, Weidner F, Gellersen H, editors. Unveiling Variations: A Comparative Study of VR Headsets Regarding Eye Tracking Volume, Gaze Accuracy, and Precision. 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); 2024 16-21 March 2024.
168. Duchowski AT. Gaze-based interaction: A 30 year retrospective. Comput Graph. 2018;73(C):59–69.
169. Patney A, Salvi M, Kim J, Kaplanyan A, Wyman C, Benty N, et al. Towards foveated rendering for gaze-tracked virtual reality. ACM Trans Graph. 2016;35(6):Article 179.
170. Wang L, Shi X, Liu Y. Foveated rendering: A state-of-the-art survey. Computational Visual Media. 2023;9(2):195-228.
171. Kahneman D, Beatty J. Pupil diameter and load on memory. Science. 1966;154(3756):1583-5.
172. Gabay S, Pertzov Y, Henik A. Orienting of attention, pupil size, and the norepinephrine system. Atten Percept Psychophys. 2011;73(1):123-9.

173. Granholm E, Steinhauer SR. Pupillometric measures of cognitive and emotional processes. Int J Psychophysiol. 2004;52(1):1-6.
174. Mathôt S, Vilotijević A. Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. Behav Res Methods. 2023;55(6):3055-77.
175. Steinhauer SR, Bradley MM, Siegle GJ, Roecklein KA, Dix A. Publication guidelines and recommendations for pupillary measurement in psychophysiological studies. Psychophysiology. 2022;59(4):e14035.
176. van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: A review. Psychon Bull Rev. 2018;25(6):2005-15.
177. Lee JY, de Jong N, Donkers J, Jarodzka H, van Merriënboer JJG. Measuring Cognitive Load in Virtual Reality Training via Pupillometry. IEEE Transactions on Learning Technologies. 2024;17:704-10.
178. Gollan B, Haslgrübler M, Ferscha A. Demonstrator for extracting cognitive load from pupil dilation for attention management services. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; Heidelberg, Germany: Association for Computing Machinery; 2016. p. 1566–71.
179. Stolte M, Gollan B, Ansorge U. Tracking visual search demands and memory load through pupil dilation. J Vis. 2020;20(6):21.
180. Holland MK, Tarlow G. Blinking and mental load. Psychol Rep. 1972;31(1):119-27.
181. Stern JA, Walrath LC, Goldstein R. The endogenous eyeblink. Psychophysiology. 1984;21(1):22-33.
182. Biondi FN, Saberi B, Graf F, Cort J, Pillai P, Balasingam B. Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. Appl Ergon. 2023;106:103867.
183. Magliacano A, Fiorenza S, Estraneo A, Trojano L. Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. Neurosci Lett. 2020;736:135293.
184. Kim J, Hwang L, Kwon S, Lee S. Change in Blink Rate in the Metaverse VR HMD and AR Glasses Environment. International Journal of Environmental Research and Public Health. 2022;19(14):8551.
185. Okazaki Y, Yokoi N. Time-course observation of tear film dynamics during VR headset use. Scientific Reports. 2025;15(1):33082.

186. Ousler GW, 3rd, Abelson MB, Johnston PR, Rodriguez J, Lane K, Smith LM. Blink patterns and lid-contact times in dry-eye and normal subjects. Clin Ophthalmol. 2014;8:869-74.
187. Gorin H, Patel J, Qiu Q, Merians A, Adamovich S, Fluet G. A Review of the Use of Gaze and Pupil Metrics to Assess Mental Workload in Gamified and Simulated Sensorimotor Tasks. Sensors (Basel). 2024;24(6).
188. Charles RL, Nixon J. Measuring mental workload using physiological measures: A systematic review. Applied Ergonomics. 2019;74:221-32.
189. Hogervorst MA, Brouwer AM, van Erp JB. Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. Front Neurosci. 2014;8:322.
190. Ellis CJ. The pupillary light reflex in normal subjects. Br J Ophthalmol. 1981;65(11):754-9.
191. Watson AB, Yellott JI. A unified formula for light-adapted pupil size. J Vis. 2012;12(10):12.
192. Seeber BU, Wackler T. Measuring auditory attention effort in virtual audio-visual environments using pupillometry. Proceedings of the International Congress on Acoustics. 2022.
193. López de Audícana-Jimenez de Aberasturi Y, Vallejo-De la Cueva A, Aretxabala-Cortajarena N, Rodriguez-Nuñez C, Pelegrin-Gaspar PM, Gil-Garcia ZI, et al. The pupillary dilation reflex to a nociceptive stimulus as a tool for analgesia management: A diagnostic study. Aust Crit Care. 2024;37(2):230-5.
194. Mischkowski D, Palacios-Barrios EE, Banker L, Dildine TC, Atlas LY. Pain or nociception? Subjective experience mediates the effects of acute noxious heat on autonomic responses - corrected and republished. Pain. 2019;160(6):1469-81.
195. Robles E, Nieto I, Navas JF, Vázquez C. Attentional biases towards emotional information in chronic pain: A multilevel meta-analysis of eye-tracking studies. Acta Psychol (Amst). 2024;250:104555.
196. Gieshoff AC, Heeb AH. Cognitive load and cognitive effort: Probing the psychological reality of a conceptual difference. Translation, Cognition & Behavior. 2023;6(1):3-28.

197. Longo L, Barrett S, editors. A Computational Analysis of Cognitive Effort. Intelligent Information and Database Systems; 2010 2010//; Berlin, Heidelberg: Springer Berlin Heidelberg.
198. Sweller J. Cognitive load during problem solving: Effects on learning. Cognitive science. 1988;12(2):257-85.
199. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. 1990.
200. Gromakovskis V, Gūtmanis MG. Flow-Induced Analgesia: A Conceptual Review of Biopsychosocial and Neurobiological Mechanisms of Pain Modulation. J Pain Res. 2025;18:6723-43.
201. Barcatta K, Holl E, Battistutta L, van der Meulen M, Rischer KM. When Less Is More: Investigating Factors Influencing the Distraction Effect of Virtual Reality From Pain. Front Pain Res (Lausanne). 2022;2:800258.
202. Gaultney WM, Dahlquist LM, Quiton RL. Cognitive load and the effectiveness of distraction for acute pain in children. Eur J Pain. 2021;25(7):1568-82.
203. Berryman C, Stanton TR, Bowering KJ, Tabor A, McFarlane A, Moseley GL. Evidence for working memory deficits in chronic pain: a systematic review and meta-analysis. Pain. 2013;154(8):1181-96.
204. Chen J, Wang X, Xu Z. The Relationship Between Chronic Pain and Cognitive Impairment in the Elderly: A Review of Current Evidence. J Pain Res. 2023;16:2309-19.
205. Zhang X, Gao R, Zhang C, Chen H, Wang R, Zhao Q, et al. Evidence for Cognitive Decline in Chronic Pain: A Systematic Review and Meta-Analysis. Front Neurosci. 2021;15:737874.
206. Gil-Ugidos A, Rodríguez-Salgado D, Pidal-Miranda M, Samartin-Veiga N, Fernández-Prieto M, Carrillo-de-la-Peña MT. Working Memory Performance, Pain and Associated Clinical Variables in Women With Fibromyalgia. Frontiers in Psychology. 2021;12.
207. Jacobsen HB, Brun A, Stubhaug A, Reme SE. Stress specifically deteriorates working memory in peripheral neuropathic pain and fibromyalgia. Brain Commun. 2023;5(4):fcad194.
208. Kirchner WK. Age differences in short-term retention of rapidly changing information. J Exp Psychol. 1958;55(4):352-8.
209. Owen AM, McMillan KM, Laird AR, Bullmore E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp. 2005;25(1):46-59.

210. Kane MJ, Conway ARA, Miura TK, Colflesh GJH. Working memory, attention control, and the N-back task: a question of construct validity. J Exp Psychol Learn Mem Cogn. 2007;33(3):615-22.
211. Middlehurst M, Schäfer P, Bagnall A. Bake off redux: a review and experimental evaluation of recent time series classification algorithms. Data Mining and Knowledge Discovery. 2024;38(4):1958-2031.
212. Dempster A, Petitjean F, Webb GI. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Mining and Knowledge Discovery. 2020;34(5):1454-95.
213. Middlehurst M, Ismail-Fawaz A, Guillaume A, Holder C, Guijo-Rubio D, Bulatova G, et al. aeon: a Python Toolkit for Learning from Time Series. Journal of Machine Learning Research. 2024;25(289):1-10.
214. Nilsson EJ, Bärgman J, Ljung Aust M, Matthews G, Svanberg B. Let Complexity Bring Clarity: A Multidimensional Assessment of Cognitive Load Using Physiological Measures. Front Neuroergon. 2022;3:787295.
215. Tao D, Tan H, Wang H, Zhang X, Qu X, Zhang T. A Systematic Review of Physiological Measures of Mental Workload. Int J Environ Res Public Health. 2019;16(15).
216. Fishburn FA, Norr ME, Medvedev AV, Vaidya CJ. Sensitivity of fNIRS to cognitive state and load. Front Hum Neurosci. 2014;8:76.
217. Putze F, Herff C, Tremmel C, Schultz T, Krusienski D, editors. Decoding Mental Workload in Virtual Environments: A fNIRS Study using an Immersive n-back Task. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019 23-27 July 2019.
218. Hu X-S, Nascimento TD, DaSilva AF. Shedding light on pain for the clinic: a comprehensive review of using functional near-infrared spectroscopy to monitor its process in the brain. PAIN. 2021;162(12):2805-20.
219. Lathlean TJH, Ramachandran AK, Sim S, Whittle IR. The clinical utility and reliability of surface electromyography in individuals with chronic low back pain: A systematic review. Journal of Clinical Neuroscience. 2024;129.
220. Meta. Meta Ray-Ban Display: AI Glasses With an EMG Wristband. 2025.

221. Kaifosh P, Reardon TR, Allen BD, Anderson C, Arnoud S, Arora R, et al. A generic non-invasive neuromotor interface for human-computer interaction. Nature. 2025;645(8081):702-11.
222. Boucsein W, Fowles DC, Grimnes S, Ben-Shakhar G, roth WT, Dawson ME, et al. Publication recommendations for electrodermal measurements. Psychophysiology. 2012;49(8):1017-34.
223. Posada-Quintero HF, Chon KH. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. Sensors (Basel). 2020;20(2).
224. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart Rate Variability. Circulation. 1996;93(5):1043-65.
225. Kristoffersson A, Lindén M. A Systematic Review of Wearable Sensors for Monitoring Physical Activity. Sensors (Basel). 2022;22(2).
226. Garofano M, Del Sorbo R, Calabrese M, Giordano M, Di Palo MP, Bartolomeo M, et al. Remote Rehabilitation and Virtual Reality Interventions Using Motion Sensors for Chronic Low Back Pain: A Systematic Review of Biomechanical, Pain, Quality of Life, and Adherence Outcomes. Technologies [Internet]. 2025; 13(5).
227. Grunberg VA, Greenberg J, Mace RA, Bakhshaie J, Choi KW, Vranceanu AM. Fitbit Activity, Quota-Based Pacing, and Physical and Emotional Functioning Among Adults With Chronic Pain. J Pain. 2022;23(11):1933-44.
228. Bell KM, Alfikri Z, Anderst W, Clark WW, Cook HA, Darwin J, et al. In-Field Ecological Momentary Assessment From Wearable Motion Sensors and Self-Report in a Chronic Low Back Pain Cohort. JOR Spine. 2026;9(1):e70156.
229. Zahabi M, Abdul Razak AM. Adaptive virtual reality-based training: a systematic literature review and framework. Virtual Reality. 2020;24(4):725-52.
230. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol. 2015;34s(0):1220-8.
231. Brain Products GmbH. LiveAmp amplifier and BrainVision Recorder Professional documentation. Gilching, Germany: Brain Products GmbH.
232. Brain Products GmbH. R-Net: sponge-based electrode system for research applications Gilching, Germany: Brain Products GmbH; [Available from: https://brainvision.com/product/r-net-the-eeg-net-from-brain-products/.

233. Brain Products GmbH. actiCAP slim/snap.

234. Gil Ávila C, Bott FS, Tiemann L, Hohn VD, May ES, Nickel MM, et al. DISCOVER-EEG: an open, fully automated EEG pipeline for biomarker discovery in clinical neuroscience. Sci Data. 2023;10(1):613.

235. The MathWorks Inc. MATLAB R2025b. Natick, Massachusetts: The MathWorks Inc.; 2025.

236. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134(1):9-21.

237. Oostenveld R, Fries P, Maris E, Schoffelen J-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience. 2011;2011(1):156869.

238. Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, et al. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. Sci Data. 2019;6(1):103.

239. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. Front Neurosci. 2013;7:267.

240. Appelhoff S, Sanderson M, Brooks TL, van Vliet M, Quentin R, Holdgraf C, et al. MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. J Open Source Softw. 2019;4(44).

241. Mullen T. CleanLine EEGLAB plugin. Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC). 2012.

242. Lee TW, Girolami M, Sejnowski TJ. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput. 1999;11(2):417-41.

243. Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. Neuroimage. 2019;198:181-97.

244. Perrin F, Pernier J, Bertrand O, Echallier JF. Spherical splines for scalp potential and current density mapping. Electroencephalogr Clin Neurophysiol. 1989;72(2):184-7.

245. Chang CY, Hsu SH, Pion-Tonachini L, Jung TP. Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Trans Biomed Eng. 2020;67(4):1114-21.
246. Chu CJ, Kramer MA, Pathmanathan J, Bianchi MT, Westover MB, Wizon L, et al. Emergence of stable functional networks in long-term human electroencephalography. J Neurosci. 2012;32(8):2703-13.
247. Ta Dinh S, Nickel MM, Tiemann L, May ES, Heitmann H, Hohn VD, et al. Brain dysfunction in chronic pain patients assessed by resting-state electroencephalography. Pain. 2019;160(12):2751-65.
248. Delorme A. EEG is better left alone. Sci Rep. 2023;13(1):2372.
249. Cohen MX. Analyzing Neural Time Series Data: Theory and Practice: The MIT Press; 2014.
250. Chertic D, Dăbală V, Livinț-Popa L, Balea M, Drăghici NC, Strilciuc Ș, et al. EEG Spectral Analysis in Chronic Pain During Rest and Cognitive Reasoning. Sensors (Basel). 2025;25(19).
251. González-Roldán AM, Cifre I, Sitges C, Montoya P. Altered Dynamic of EEG Oscillations in Fibromyalgia Patients at Rest. Pain Med. 2016;17(6):1058-68.
252. Simis M, Imamura M, Pacheco-Barrios K, Marduy A, de Melo PS, Mendes AJ, et al. EEG theta and beta bands as brain oscillations for different knee osteoarthritis phenotypes according to disease severity. Sci Rep. 2022;12(1):1480.
253. Grandy TH, Werkle-Bergner M, Chicherio C, Schmiedek F, Lövdén M, Lindenberger U. Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults. Psychophysiology. 2013;50(6):570-82.
254. Furman AJ, Prokhorenko M, Keaser ML, Zhang J, Chen S, Mazaheri A, et al. Sensorimotor Peak Alpha Frequency Is a Reliable Biomarker of Prolonged Pain Sensitivity. Cereb Cortex. 2020;30(12):6069-82.
255. He BJ. Scale-free brain activity: past, present, and future. Trends Cogn Sci. 2014;18(9):480-7.
256. Donoghue T, Dominguez J, Voytek B. Electrophysiological Frequency Band Ratio Measures Conflate Periodic and Aperiodic Neural Activity. eNeuro. 2020;7(6).
257. Pico Interactive. Pico Neo 3 Pro Eye virtual reality headset. San Francisco, USA: Pico Interactive; 2021.

258. Tobii. VR headset & Tobii eye tracking | Pico Neo 3 Pro Eye 2023 [Available from: https://www.tobii.com/products/integration/xr-headsets/device-integrations/pico-neo-3-pro-eye/.

259. Unity Technologies. Unity Real-Time Development Platform, version 2021.3.45f1. San Francisco, USA: Unity Technologies; 2024.

260. Tobii AB. Tobii Ocumen SDK: Overview 2025 [updated 2025. Available from: https://developer.tobii.com/xr/ocumen/overview/.

261. Mathôt S. Pupillometry: Psychology, Physiology, and Function. Journal of Cognition. 2018.

262. Karatekin C, Bingham C, White T. Regulation of cognitive resources during an n-back task in youth-onset psychosis and attention-deficit/hyperactivity disorder (ADHD). Int J Psychophysiol. 2009;73(3):294-307.

263. Mitre-Hernandez H, Covarrubias Carrillo R, Lara-Alvarez C. Pupillary Responses for Cognitive Load Measurement to Classify Difficulty Levels in an Educational Video Game: Empirical Study. JMIR Serious Games. 2021;9(1):e21620.

264. Beatty J. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol Bull. 1982;91(2):276-92.

265. Beatty J, Lucero-Wagoner B. The pupillary system. Handbook of psychophysiology, 2nd ed. New York: Cambridge University Press; 2000. p. 142-62.

266. Nyström M, Andersson R, Niehorster DC, Hessels RS, Hooge ITC. What is a blink? Classifying and characterizing blinks in eye openness signals. Behavior Research Methods. 2024;56(4):3280-99.

267. Kret ME, Sjak-Shie EE. Preprocessing pupil size data: Guidelines and code. Behav Res Methods. 2019;51(3):1336-42.

268. Mathôt S, Fabius J, Van Heusden E, Van der Stigchel S. Safe and sensible preprocessing and baseline correction of pupil-size data. Behav Res Methods. 2018;50(1):94-106.

269. Winn MB, Wendt D, Koelewijn T, Kuchinsky SE. Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. Trends Hear. 2018;22:2331216518800869.

270. Middlehurst M, Large J, Bagnall A, editors. The Canonical Interval Forest (CIF) Classifier for Time Series Classification. 2020 IEEE International Conference on Big Data (Big Data); 2020 10-13 Dec. 2020.

271. Dempster A, Schmidt DF, Webb GI. quant: a minimalist interval method for time series classification. Data Mining and Knowledge Discovery. 2024;38(4):2377-402.

272. Rodriguez J, Kuncheva L, Alonso C. Rotation Forest: A New Classifier Ensemble Method. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006;28(10):1619-30.

273. Middlehurst M, Large J, Flynn M, Lines J, Bostrom A, Bagnall A. HIVE-COTE 2.0: a new meta ensemble for time series classification. Machine Learning. 2021;110(11):3211-43.

274. The aeon developers. RocketRegressor [Available from: https://www.aeon-toolkit.org/en/v0.5.0/api_reference/auto_generated/aeon.regression.convolution_based.RocketRegressor.html.

275. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. New York: Springer; 2009.

276. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. NeuroImage. 2018;180:68-77.

277. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. Gigascience. 2017;6(5):1-9.

278. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research. 2010;11:2079-107.

279. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns (N Y). 2023;4(9):100804.

280. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7(1):91.

281. Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction. JAMA Psychiatry. 2020;77(5):534-40.

282. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Häuser W, Katz RL, et al. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum. 2016;46(3):319-29.

283. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken). 2010;62(5):600-10.
284. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum. 1990;33(2):160-72.
285. Verdi S, Abbasian G, Bowyer RCE, Lachance G, Yarand D, Christofidou P, et al. TwinsUK: The UK Adult Twin Registry Update. Twin Res Hum Genet. 2019;22(6):523-9.
286. Jarrar ZA, Ward KJ, Mangino M, Cherkas LF, Gill R, Gillham-Nasenya I, et al. Definitive Zygosity Scores in the Peas in the Pod Questionnaire is a Sensitive and Accurate Assessment of the Zygosity of Adult Twins. Twin Research and Human Genetics. 2018;21(2):146-54.
287. White KP, Harth M, Speechley M, Ostbye T. Testing an instrument to screen for fibromyalgia syndrome in general population studies: the London Fibromyalgia Epidemiology Study Screening Questionnaire. J Rheumatol. 1999;26(4):880-4.
288. Livshits G, Macgregor AJ, Gieger C, Malkin I, Moayyeri A, Grallert H, et al. An omics investigation into chronic widespread musculoskeletal pain reveals epiandrosterone sulfate as a potential biomarker. Pain. 2015;156(10):1845-51.
289. Livshits G, Ni Lochlainn M, Malkin I, Bowyer R, Verdi S, Steves CJ, et al. Shared genetic influence on frailty and chronic widespread pain: a study from TwinsUK. Age Ageing. 2018;47(1):119-25.
290. Leuchs L, Pizzamigilo S, Bellomo E. How to decide which EEG electrode technology best fits your research Gilching, Germany: Brain Products GmbH; 2022 [updated 2022/08/25. Available from: https://pressrelease.brainproducts.com/choosing-your-electrode-technology/.
291. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care. 2003;41(11):1284-92.
292. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-7.
293. Wolfe F. Fibromyalgianess. Arthritis and Rheumatism. 2009;61(6):715-6.
294. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS),

Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care and Research. 2011;63(S11):S240-S52.
295. Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. Journal of Rheumatology. 2004;31(10):1896-902.
296. Cajochen C, Brunner DP, Krauchi K, Graw P, Wirz-Justice A. Power Density in Theta/Alpha Frequencies of the Waking EEG Progressively Increases During Sustained Wakefulness. Sleep. 1995;18(10):890-4.
297. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1):9-19.
298. Kleinert T, Koenig T, Nash K, Wascher E. On the Reliability of the EEG Microstate Approach. Brain Topogr. 2024;37(2):271-86.
299. Li N, Yang J, Long C, Lei X. Test-Retest Reliability of EEG Aperiodic Components in Resting and Mental Task States. Brain Topogr. 2024;37(6):961-71.
300. Wiesman AI, da Silva Castanheira J, Baillet S. Stability of spectral estimates in resting-state magnetoencephalography: Recommendations for minimal data duration with neuroanatomical specificity. Neuroimage. 2022;247:118823.
301. Rolle CE, Narayan M, Wu W, Toll R, Johnson N, Caudle T, et al. Functional connectivity using high density EEG shows competitive reliability and agreement across test/retest sessions. J Neurosci Methods. 2022;367:109424.
302. Tenke CE, Kayser J. Generator localization by current source density (CSD): implications of volume conduction and field closure at intracranial and scalp resolutions. Clin Neurophysiol. 2012;123(12):2328-45.
303. Michel CM, Koenig T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. Neuroimage. 2018;180(Pt B):577-93.
304. Schmidt S, Naranjo JR, Brenneisen C, Gundlach J, Schultz C, Kaube H, et al. Pain ratings, psychological functioning and quantitative EEG in a controlled study of chronic back pain patients. PLoS One. 2012;7(3):e31138.
305. Fu H, Garrett B, Tao G, Cordingley E, Ofoghi Z, Taverner T, et al. Virtual Reality-Guided Meditation for Chronic Pain in Patients With Cancer: Exploratory Analysis of Electroencephalograph Activity. JMIR Biomed Eng. 2021;6(2):e26332.

306. Emmerling T. actiCAP active Electrodes Walkthrough Gilching, Germany: Brain Products GmbH; 2017 [updated 2017/04/21. Available from: https://pressrelease.brainproducts.com/active-electrodes-walkthrough/.

307. Melzack R. The short-form McGill pain questionnaire. Pain. 1987;30(2):191-7.

308. Bennett RM, Friend R, Jones KD, Ward R, Han BK, Ross RL. The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties. Arthritis Research & Therapy. 2009;11(4):R120.

309. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine. 2001;16(9):606-13.

310. Tran Y, Austin P, Lo C, Craig A, Middleton JW, Wrigley PJ, et al. An Exploratory EEG Analysis on the Effects of Virtual Reality in People with Neuropathic Pain Following Spinal Cord Injury. Sensors [Internet]. 2022; 22(7).

311. Magosso E, De Crescenzio F, Ricci G, Piastra S, Ursino M. EEG Alpha Power Is Modulated by Attentional Changes during Cognitive Tasks and Virtual Reality Immersion. Comput Intell Neurosci. 2019;2019:7051079.

312. Schubring D, Kraus M, Stolz C, Weiler N, Keim DA, Schupp H. Virtual Reality Potentiates Emotion and Task Effects of Alpha/Beta Brain Oscillations. Brain Sci. 2020;10(8).

313. Li J, Yang H, Xiao Y, Liu X, Ma B, Ma K, et al. The analgesic effects and neural oscillatory mechanisms of virtual reality scenes based on distraction and mindfulness strategies in human volunteers. Br J Anaesth. 2023;131(6):1082-92.

314. Jensen O, Mazaheri A. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci. 2010;4:186.

315. Pfurtscheller G, Lopes da Silva FH. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol. 1999;110(11):1842-57.

316. Pineda JA. The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". Brain Res Brain Res Rev. 2005;50(1):57-68.

317. Song YH, Lee HM. Effect of Immersive Virtual Reality-Based Bilateral Arm Training in Patients with Chronic Stroke. Brain Sci. 2021;11(8).

318. Berger AM, Davelaar EJ. Frontal Alpha Oscillations and Attentional Control: A Virtual Reality Neurofeedback Study. Neuroscience. 2018;378:189-97.

319. Dekker MK, Sitskoorn MM, Denissen AJ, van Boxtel GJ. The time-course of alpha neurofeedback training effects in healthy participants. Biol Psychol. 2014;95:70-3.

320. Gola M, Magnuski M, Szumska I, Wróbel A. EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects. Int J Psychophysiol. 2013;89(3):334-41.
321. Lundqvist M, Miller EK, Nordmark J, Liljefors J, Herman P. Beta: bursts of cognition. Trends Cogn Sci. 2024;28(7):662-76.
322. Nougaret S, López-Galdo L, Caytan E, Poitreau J, Barthélemy FV, Kilavik BE. Low and high beta rhythms have different motor cortical sources and distinct roles in movement control and spatiotemporal attention. PLoS Biol. 2024;22(6):e3002670.
323. Spitzer B, Haegens S. Beyond the Status Quo: A Role for Beta Oscillations in Endogenous Content (Re)Activation. eNeuro. 2017;4(4).
324. Liljefors J, Almeida R, Rane G, Lundström JN, Herman P, Lundqvist M. Distinct functions for beta and alpha bursts in gating of human working memory. Nat Commun. 2024;15(1):8950.
325. Tian S, Cheng YA, Luo H. Rhythm Facilitates Auditory Working Memory via Beta-Band Encoding and Theta-Band Maintenance. Neurosci Bull. 2025;41(2):195-210.
326. Department for Business Energy & Industrial Strategy. The safety of domestic virtual reality systems: a literature review. Government report. London, UK: Department for Business, Energy & Industrial Strategy; 2020 2020/09. Report No.: 2020/038.
327. Haatveit BC, Sundet K, Hugdahl K, Ueland T, Melle I, Andreassen OA. The validity of d prime as a working memory index: Results from the “Bergen n-back task”. Journal of Clinical and Experimental Neuropsychology. 2010;32(8):871-80.
328. Ni L, Ma WJ. A computational approach to the N-back task. Scientific Reports. 2024;14(1):30211.
329. McMillan KM, Laird AR, Witt ST, Meyerand ME. Self-paced working memory: validation of verbal variations of the n-back paradigm. Brain Res. 2007;1139:133-42.
330. Byrne EM, Gilbert RA, Kievit RA, Holmes J. Evidence for separate backward recall and n-back working memory factors: a large-scale latent variable analysis. Memory. 2024;32(9):1182-98.
331. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. Journal of Clinical Epidemiology. 2010;63(11):1179-94.

332. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment. 1995;7(4):524-32.
333. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock PA, Meshkati N, editors. Advances in Psychology. 52: North-Holland; 1988. p. 139-83.
334. Paas F. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. Journal of Educational Psychology. 1992;84:429-34.
335. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Applied Ergonomics. 2018;69:66-73.
336. Amtmann D, Cook KF, Jensen MP, Chen W-H, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. Pain. 2010;150(1):173-82.
337. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proc Hum Factors Ergon Soc Annu Meet. 2006;50(9):904-8.
338. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educ Psychol. 2003;38(1):63-71.
339. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.
340. Python Software Foundation. Python Language Reference, version 3.12. Wilmington, DE, USA: Python Software Foundation; 2023.
341. Granholm E, Asarnow RF, Sarkin AJ, Dykes KL. Pupillary responses index cognitive resource limitations. Psychophysiology. 1996;33(4):457-61.
342. Klingner J, Tversky B, Hanrahan P. Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. Psychophysiology. 2011;48(3):323-32.
343. Klingner J. The pupillometric precision of a remote video eye tracker. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications; Austin, Texas: Association for Computing Machinery; 2010. p. 259–62.
344. Salari M, Niehorster DC, Nyström M, Bednarik R. The effect of pupil size on data quality in head-mounted eye trackers. Behav Res Methods. 2025;58(1):17.

345. Lee JY, Jong Nd, Donkers J, Jarodzka H, Merriënboer JJGv. Measuring Cognitive Load in Virtual Reality Training via Pupillometry. IEEE Transactions on Learning Technologies. 2024;17:704-10.
346. Thomay C, Fermitsch A, Fessler J, Garatva P, Gollan B, Lietz AK, et al., editors. Towards Cognitive Load-Based Decision Making in VR Training. 2023 IEEE 2nd International Conference on Cognitive Aspects of Virtual Reality (CVR); 2023 26-27 Oct. 2023.
347. Kim J, Sunil Kumar Y, Yoo J, Kwon S. Change of Blink Rate in Viewing Virtual Reality with HMD. Symmetry. 2018;10(9):400.
348. Bologna M, Paparella G, Valls-Solé J, Hallett M, Berardelli A. Neural control of blinking. Clinical Neurophysiology. 2024;161:59-68.
349. Jongkees BJ, Colzato LS. Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. Neuroscience & Biobehavioral Reviews. 2016;71:58-82.
350. Dang LC, Samanez-Larkin GR, Castrellon JJ, Perkins SF, Cowan RL, Newhouse PA, et al. Spontaneous Eye Blink Rate (EBR) Is Uncorrelated with Dopamine D2 Receptor Availability and Unmodulated by Dopamine Agonism in Healthy Adults. eNeuro. 2017;4(5):ENEURO.0211-17.2017.
351. Siegle GJ, Ichikawa N, Steinhauer S. Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. Psychophysiology. 2008;45(5):679-87.
352. Hoppe D, Helfmann S, Rothkopf CA. Humans quickly learn to blink strategically in response to environmental task demands. Proceedings of the National Academy of Sciences. 2018;115(9):2246-51.
353. Nakano T, Kato M, Morito Y, Itoi S, Kitazawa S. Blink-related momentary activation of the default mode network while viewing videos. Proceedings of the National Academy of Sciences. 2013;110(2):702-6.
354. Volkmann FC, Riggs LA, Moore RK. Eyeblinks and Visual Suppression. Science. 1980;207(4433):900-2.
355. Rac-Lubashevsky R, Slagter HA, Kessler Y. Tracking Real-Time Changes in Working Memory Updating and Gating with the Event-Based Eye-Blink Rate. Scientific Reports. 2017;7(1):2547.
356. Attridge N, Noonan D, Eccleston C, Keogh E. The disruptive effects of pain on n-back task performance in a large general population sample. Pain. 2015;156(10):1885-91.

357. Eckstein MK, Guerra-Carrillo B, Miller Singley AT, Bunge SA. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? Dev Cogn Neurosci. 2017;25:69-91.
358. Sarkar U, Lee JE, Nguyen KH, Lisker S, Lyles CR. Barriers and Facilitators to the Implementation of Virtual Reality as a Pain Management Modality in Academic, Community, and Safety-Net Settings: Qualitative Analysis. J Med Internet Res. 2021;23(9):e26623.
359. Webb CA, Hilt LM, Swords CM, Bolt DM, Fisher H, Goldberg SB. Ecological Momentary Assessment as a Measure of Intervention Change: Evaluation in 4 Digital Mental Health Trials. J Med Internet Res. 2025;27:e69297.
360. Elser A, Lange M, Kopkow C, Schäfer AG. Barriers and Facilitators to the Implementation of Virtual Reality Interventions for People With Chronic Pain: Scoping Review. JMIR XR Spatial Comput. 2024;1:e53129.
361. Eckert M-L, Habets E, Rummukainen OS. Cognitive Load Estimation Based on Pupillometry in Virtual Reality with Uncontrolled Scene Lighting. 2021 13th International Conference on Quality of Multimedia Experience (QoMEX). 2021:73-6.
362. Souchet AD, Philippe S, Lourdeaux D, Leroy L. Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review. International Journal of Human–Computer Interaction. 2022;38(9):801 - 24.
363. Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. PAIN Reports. 2021;6(1).
364. Bronte-Stewart HM, Beudel M, Ostrem JL, Little S, Almeida L, Ramirez-Zamora A, et al. Long-Term Personalized Adaptive Deep Brain Stimulation in Parkinson Disease: A Nonrandomized Clinical Trial. JAMA Neurology. 2025;82(11):1171-80.
365. Stamatis CA, Farlow DN, Mercaldi C, Suh M, Maple A, Savarese A, et al. Two single arm trials of AKL-T01, a digital therapeutic for adolescents and adults with ADHD. npj Mental Health Research. 2024;3(1):30.
366. Apple Inc. Apple Vision Pro: technical specifications. Cupertino, CA: Apple Inc.; 2024.
367. Meta. Meta Quest 3: next-generation mixed reality headset. 2023.
368. Pico Interactive. PICO 4 Ultra: VR/MR headset specifications. Pico Interactive; 2024.
369. Snap Inc. SPS 2024: introducing new Spectacles and Snap OS. Snap Inc.; 2024.
370. Xreal. XREAL One Pro. XREAL; 2025.

371. OpenAI. Introducing ChatGPT. San Francisco, CA: OpenAI; 2022.
372. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc.; 2017. p. 6000–10.
373. Venerito V, Lopalco G, Iannone F, Gupta L. Agentic AI in rheumatology. The Lancet Rheumatology. 2026;8(6):e486-e91.
374. Jiang W-B, Zhao L-M, Lu B-L. Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI. ArXiv. 2024;abs/2405.18765.
375. Cui W, Jeong W, Tholke P, Medani T, Jerbi K, Joshi AA, et al. Neuro-GPT: Towards A Foundation Model For EEG. 2024 IEEE International Symposium on Biomedical Imaging (ISBI). 2024:1-5.
376. Ansari AF, Stella L, Turkmen C, Zhang X, Mercado P, Shen H, et al. Chronos: Learning the Language of Time Series. ArXiv. 2024;abs/2403.07815.
377. Dempster A, Schmidt DF, Webb GI. Hydra: competing convolutional kernels for fast and accurate time series classification. Data Mining and Knowledge Discovery. 2023;37(5):1779-805.
378. British Pain Society. Outcome Measures. London: The British Pain Society; 2019 2019/01.
379. HealthMeasures. PROMIS Pain Intensity Scoring Manual. Evanston, IL: HealthMeasures, Northwestern University; 2021 2021/06/18.
380. HealthMeasures. PROMIS Pain Interference Scoring Manual. Evanston, IL: HealthMeasures, Northwestern University; 2023 2023/12/05.
381. Sullivan MJL. The Pain Catastrophizing Scale: User Manual. Montreal: McGill University; 2009.

11. Appendices

Appendix 1: Published parent feasibility study conducted alongside the VIPA-EEG study

This appendix includes the published article reporting the parent feasibility study conducted alongside the VIPA-EEG study described in Chapter 6. The publication provides the clinical and intervention context for the VIPA-EEG analysis, including the feasibility study design, VR system used, VR activities, acceptability outcomes, adverse-effect monitoring and preliminary clinical outcomes.

Tsigarides J, Grove V, Chipping J, Dainty J, Miles S, Shenker N, Sami S, MacGregor A. Feasibility of attention-based virtual reality interventions in fibromyalgia syndrome: comparing systems, virtual environments and activities. *British Journal of Pain*. 2025;19(2):71-85. doi:10.1177/20494637241310696.

Reproduced under the Creative Commons Attribution 4.0 Licence.

Article

THE BRITISH PAIN SOCIETY
EXPERTISE WHERE IT MATTERS

British Journal of Pain
2025, Vol. 19(2) 71–85
© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20494637241310696
journals.sagepub.com/home/bjp
Sage

Feasibility of attention-based virtual reality interventions in fibromyalgia syndrome: comparing systems, virtual environments and activities

Jordan Tsigarides[1], Vanessa Grove[1], Jacqueline Chipping[1], Jack Dainty[1], Susan Miles[1], Nicholas Shenker[2], Saber Sami[1] and Alexander Macgregor[1]

Abstract

Background: Fibromyalgia Syndrome (FMS) is highly prevalent with a significant associated morbidity and socioeconomic burden. Effective treatments for FMS remain elusive with pharmacological management (including use of opioids) often proving ineffective. There is a need to develop accessible, innovative management approaches to improve patient care.

Virtual reality (VR) interventions have shown evidence of efficacy in the management of acute pain and chronic low back pain, but their feasibility in FMS has not hitherto been explored.

Methods: This feasibility study investigates the use of four different VR systems, four interactive VR activities and two virtual environments in patients with FMS. Acceptability (including adverse effects) and study engagement were the main outcomes investigated. Clinical outcome data on pain and mood were also collected to gather preliminary information for future studies.

Results: The results demonstrated good feasibility across VR systems, activities and virtual environments with high levels of acceptability, low frequency of adverse effects, and positive perceptions of VR in patients with FMS. Reporting of adverse effects (including fatigue) varied across different VR components, with system comfort and virtual environmental design being particularly important.

Conclusions: The findings increase our confidence with respect to the feasibility of using VR in people with FMS, help to inform future randomised controlled trials and emphasise the importance of tailored interventional design for future VR therapeutics.

Keywords
Chronic pain, fibromyalgia, pain management, virtual reality intervention, feasibility

Introduction

Fibromyalgia Syndrome (FMS) affects between 4 and 6% of the adult population[1,2] and is characterised by chronic widespread pain, fatigue, sleep disturbance, cognitive 'fog', and significant psychological distress.[3] Despite the socioeconomic burden and associated morbidity, effective treatments for FMS remain elusive with pharmacological management (including use of opioids) often proving ineffective.[4] The lack of pharmacological efficacy has led to the European League Against Rheumatism (EULAR) and National Institute of for Health and Care Excellence (NICE) guidance recommending non-pharmacological approaches as first-line.[5,6] However, the accessibility and acceptability of

[1]Faculty of Medicine and Health Sciences, Norwich Medical School, University of East Anglia, Norwich, UK
[2]Department of Rheumatology, Addenbrookes University Hospital, Cambridge, UK

Corresponding author:
Jordan Tsigarides, Faculty of Medicine and Health Sciences, Norwich Medical School, Bob Champion Research and Education Building, Norwich Research Park, Rosalind Franklin Road, Norwich NR4 7UQ, UK.
Email: J.tsigarides@uea.ac.uk

recommended therapeutic options including supervised aerobic exercise can be challenging due to factors including resource constraints and psychological factors.[7] There is a need for novel, accessible, and well-tolerated therapeutic strategies for this patient population.

Virtual Reality (VR) has attracted increasing interest in pain management given evidence of analgesic efficacy in acute pain.[8,9] This novel technology immerses users within a fully interactive, computer-generated virtual environment through use of a head-mounted display (HMD). The immersion in VR is achieved by fully engaging the user's visual and auditory senses and effectively removing distractions from the outside world. Pain demands attention[10] and the immersive nature of VR has great potential in re-allocating attentional resources away from pain (also known as attentional modulation) and supporting personalised management approaches.[11,12]

However, VR interventions are not uniform; they consist of multiple modifiable parts including hardware, virtual environments and activities. These components can significantly influence the overall experience and efficacy of VR as a pain management tool.[13–15] Small feasibility studies investigating attention-based VR (interventions primarily aimed at engaging attention through VR activities) in heterogenous chronic pain cohorts have shown promising results. These interventions have only included single activities completed within one virtual environment (often using a 'rail-shooter' game mechanic within a cold virtual environment as investigated within the acute pain literature).[16,17]

When evaluating VR in chronic pain, the challenges of specific patient groups need consideration to avoid a potentially inappropriate 'one-size-fits-all' approach. This is particularly important in FMS, given those living with the condition face specific challenges including severe fatigue, sensory hypersensitivity (including sound, light and temperature), and cognitive symptoms[18] that will inevitably influence their experience of VR interventions. Despite highly promising results from recent RCTs investigating VR in chronic low back pain,[19,20] there are no studies that have specifically investigated the feasibility of different VR head-mounted systems (with different technical capabilities), interactive activities (with different mechanics), and virtual environments in FMS.

To address the lack of VR research conducted in FMS cohorts, this feasibility study's primary goal was to determine the acceptability of different VR system configurations, interactive activities, and virtual environments. This work is essential to ensure that subsequent larger studies, designed to rigorously test the efficacy of VR pain interventions in FMS, are built on a solid foundation of feasibility and patient acceptability data.

Methods

Study design and ethics

This study was designed as a feasibility study aiming to investigate the acceptability of different (1) VR systems, (2) Virtual environments, and (3) interactive VR activities in patients with FMS. This conforms to a 'VR2' design as outlined by the VR-CORE international working group.[21] Data collection sessions (i.e. sessions in which participants completed the interventions) were conducted within an allocated room at Norwich Medical School, with adjustments made to the data collection protocol to account for safety considerations during the COVID-19 pandemic. This study was approved by the Cornwall and Plymouth Research Ethics Committee (REC Reference 20/SW/0050) and registered with the International Standard Randomised Controlled Trial Number registry (ISRCTN46681140).

Participants and screening

Recruitment took place from October 2020 to April 2021. Participants were directly identified by clinicians within general rheumatology outpatient clinics at the Norfolk and Norwich University Hospital and Addenbrookes Hospital (Cambridge). Recruitment sites were chosen based on locality to the study site for the convenience of researchers and participants. The inclusion criteria were (1) adults ≥18 years of age; (2) ability to understand and speak conversational English; (3) capacity to consent; (4) diagnosis of FMS fulfilling the 2016 diagnostic criteria.[18] Exclusion criteria were (1) presence of co-morbid condition(s) exacerbated by exposure to flashing lights or screens; (2) a diagnosis of cognitive impairment; (3) current visual or hearing impairment that would prevent the participant from using the technology included the research.

Procedures

Overall study flow. The overall study flow is outlined in Figure 4, with the stages in each data collection session outlined in Figure 1. Following identification, individuals were given an invitation letter and participant information sheet clearly explaining the study and stating the inclusion and exclusion criteria. Those wishing to participate returned a reply slip contained within the invitation letter or contacted the research team via email or telephone to express their interest. Participants were then screened via telephone by a member of the research team and invited to data collection sessions. Each of the three data collection

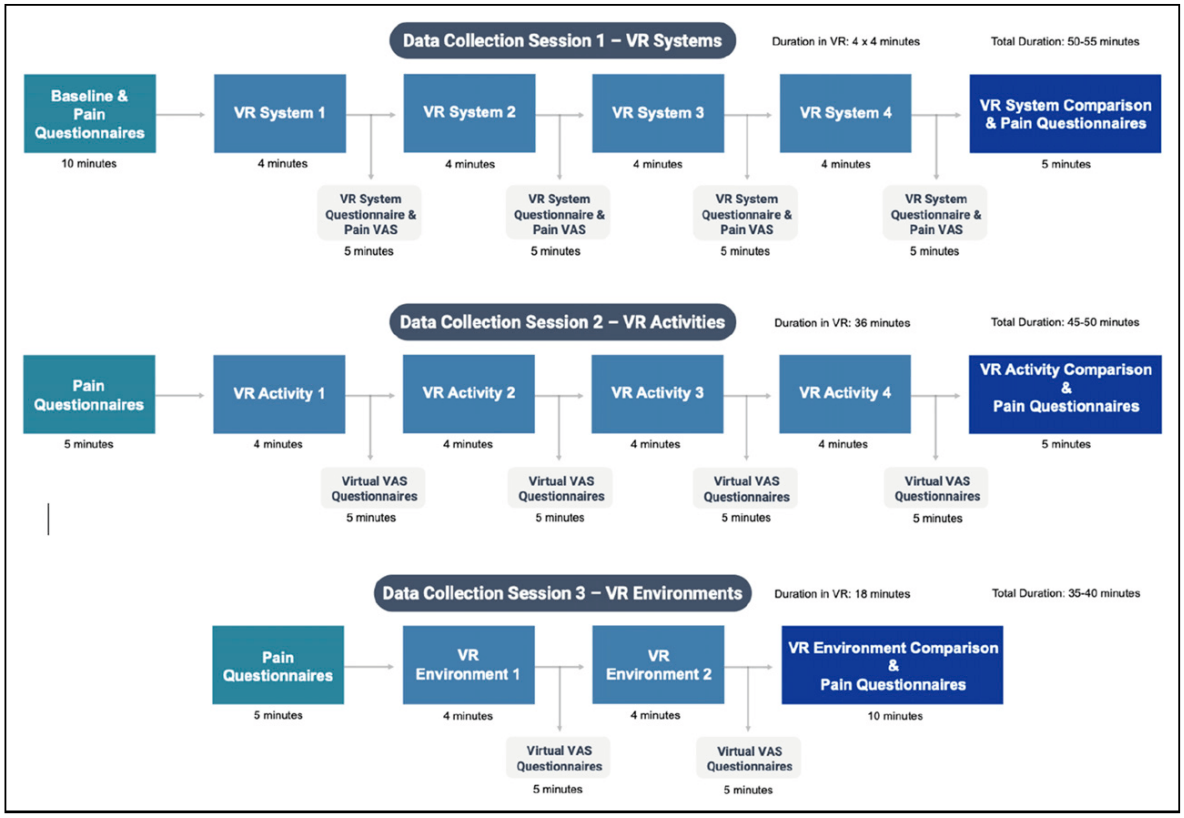


Figure 1. Flow diagram showing the steps of each data collection session, durations of each step, and total durations.

sessions investigated a different component of the VR experience:

- Data Collection Session 1: Four different VR systems
- Data Collection Session 2: Four different VR activities
- Data Collection Session 3: Two different VR environments

To minimise carryover effects, there was a mandatory 'washout' period of at least 1 week between sessions. One week has been used as a washout period in other VR studies that investigated similar interventions.[22] Within each session, the sequence of interventions was determined using simple randomisation, employing random number tables to ensure an unbiased distribution of the order in which interventions were experienced.

Upon attending the first data collection session, participants completed a written consent form and baseline questionnaires including demographics, disease severity (including pain severity), co-morbidities and previous exposure to VR. Capacity for participants to provide informed consent for the procedures included in the study was assessed by the research team at the time of consent form completion and reassessed at subsequent data collection sessions. Ability to read and understand written English questionnaires was checked with each participant by a member of the study team. No adaptations to procedures or technologies included within data collection sessions were required to allow participants to take part.

Facets of feasibility tested. Acceptability (including adverse effects) and study engagement were the main outcomes investigated in this study. Acceptability pertains to the degree in which participants found VR interventions engaging, immersive, usable, comfortable and satisfactory overall. Adverse effects are defined as any unfavourable or harmful events that occur during or after the VR interventions and are reasonably associated with their use (such as physical discomfort). Study engagement relates to participants' willingness to participate (including rates of study recruitment and retention), and their perceptions of VR after use. Feasibility encapsulates all of the above, and provides

an overall indicator of the practicality and potential for successful implementation of the VR interventions in larger-scale studies.[23] Clinical outcome data on pain and mood were also collected to gather preliminary information for future larger-scale studies.

The VR application development and content. The VR application was developed specifically for the purposes of this research alongside industry partners (Orbital Global Ltd) with input from the research team comprising of academic clinicians and researchers with expertise in chronic pain, neuroscience, virtual reality, game design and psychology. The application was designed specifically for patients with chronic pain with careful consideration of the audio-visual experience, activity-related movement requirements, and orientation of the experiences. VR environments were set within a tranquil forest, accompanied by relaxing sounds to create a serene atmosphere for participants during the activities. Similar forest settings have been used to promote relaxation in previous VR studies such as the 'Virtual Meditative Walk'.[24] Over-ear headphones were used to optimise audio quality and promote immersion, with participants always remaining seated during VR use (although able to stand up and move around when not using VR). Hand-held controllers were used to enable interaction (Figure 2). A starting area with a selection menu allowed the user to choose from four activities and two virtual environments. This starting area also included a VR questionnaire that enabled collection of virtual visual analogue scale (VAS) data for evaluation of activities and virtual environments in Data Collection Sessions 2 and 3. Additional details are provided in Appendix 1.

Data collection session 1: The VR systems. Four head-mounted VR systems were investigated, each with different characteristics representing the spectrum of commercially available technology at the time. Systems included the Samsung Gear VR (Samsung Electronics, Ridgefield Park, New Jersey) with a Samsung Galaxy S8 smartphone, Oculus Go, Oculus Quest and Oculus Rift S (Reality Labs, Menlo Park, California). One hand-held controller specific to each VR system was used to enable interaction (e.g. Gear VR controller, Oculus Go controller, Oculus Quest and Rift touch controllers). See the Appendix for a comparison of technical specifications. Custom head mounting straps were used with the Oculus Go and disposable fascia covers used with the Samsung Gear VR to ensure adequate infection control procedures between users.

Participants used each system in a random order, experiencing the same interactive VR activity (rail-shooter activity) within the same virtual environment



Figure 2. Participant seated, using the Oculus Rift S during Data Collection Session 1.

(cold) with each system. See Figure 1 for the durations of each step in this session, and sections describing data collection sessions 2 and 3 below for descriptions of the rail-shooter activity and virtual environments.

Data collection session 2: The VR activities. Four time-restricted VR activities were created, each requiring the user to utilise skills in attention but with each challenging further skills through different game 'mechanics' (Table 1).

Elements of 'gamification' were added to each of the activities to encourage engagement, motivation and entry into a psychological state of 'flow' which is thought to promote task focus with increased attentional modulation.[25] This included a scoring system dependent on performance, with scores displayed to the user during the activity. The types of activities included were carefully selected for the following reasons:

(1) Each activity type challenges different cognitive skills (as outlined in Table 1) – Many of which can be impaired in FMS (e.g. short-term memory).
(2) Activity types included (rail-shooter, memory, multitasking, match-3) are commonly used in other contexts as cognitively engaging, attention-demanding tasks.
(3) Previous VR pain research have demonstrated immersion, engagement and pain relief in similar tasks.[9,16,17,19]

(4) Each activity used the same mode of interaction (controller-based), with a similar intensity of interaction, and a similar low intensity of required movement (both the arms and head). This was to ensure an even level of physical activity intensity across the activities, and to reduce risk of pain exacerbation in FMS participants that commonly suffer with pain affecting the head, neck and arms.

Participants used the Rift S VR system (with rift touch controller) to complete each of the four VR activities within the same virtual environment (cold). The order in which they experienced the activities was random. See Figure 1 for the durations of each step in this session.

Data collection session 3: The virtual environments. Two contrasting virtual environments were investigated, one simulating a cold (snowy) world and another a warm (sunny) world within VR (Figure 3). These contrasting virtual environments were designed based on previous acute pain VR research that utilised virtual environments depicting winter scenes[26] and evidence of thermodysregulation states in FMS.[27] Each virtual environment provided a different audio-visual experience but did not alter the functionality or interactivity of the VR program. Participants used the Rift S VR system (with rift touch controller) to complete the same VR activity (rail-shooter) within each of the two virtual environments. The order in which they experienced the virtual environments was random. See Figure 1 for the durations of each step in this session.

Measures

All measures were delivered in English and completed during the lab-based data collection sessions with no longitudinal follow-up given the feasibility nature of the study. Table 2 provides a summary of measures used. The Appendix provides more detailed information on each measure used including the calculation method for composite scores for each outcome.

Acceptability – User experience, usability, overall comfort and adverse effects. Post-intervention self-report questionnaires were used to gain information on the overall comfort, usability and immersive nature of VR. Written questionnaires posed a combination of rating style (1-7 scale), multiple choice, and white space questions. Rating question data were dichotomised; participants reporting a rating of 5-7 were categorised as agreeing with the statement and participants reporting 1-3 as disagreeing. Virtual questionnaires (completed in VR) gained VAS ratings specific to the VR activity or virtual environment experienced.

The Virtual Reality Sickness Questionnaire (VRSQ)[32] was used post-intervention to collect data on adverse effects. The VRSQ is a validated tool designed to assess the physical discomfort experienced with VR interventions and has been shown to have better psychometric qualities for assessing VR applications than the traditionally used Simulator Sickness Questionnaire (SSQ).[33] The VRSQ includes a list of nine symptoms observed with use of VR systems, each scored based on severity (0-3, none to severe), with a score range of 0-100.

Study engagement and perceptions. Rating questions within the self-report post-intervention questionnaires were used to understand participants' perception of VR following each intervention. This included questions enquiring about their openness to using VR at home for pain management. In addition, data on participant recruitment and retention rates were recorded.

Clinical outcomes. The McGill Pain Questionnaire Short Form (MPQ-SF),[31] pain VAS, and specific questions within subjective experience questionnaires were used to gain data on pain intensity and quality, as well as changes in pain perception over the course of the study and across different interventions. The MPQ-SF and written pain VAS were completed at the start of each data collection session prior to VR use. The MPQ-SF was completed again at the end of sessions after completing interventions. A written pain VAS was completed immediately after use of each VR system within Data Collection Session 1. A virtual pain VAS was completed immediately after experiencing each VR activity or environment in Data Collection Sessions 2 and 3 respectively, enabling participants to report pain intensity without needing to remove VR systems. Both VAS measures included a continuous horizontal scale in which participants marked their pain intensity at the time of measurement from 'No Pain' to 'Worst Possible Pain'. The horizontal line for the written VAS was 10 cm in length and an equivalent non-numerical scale was used for the virtual VAS. See the Appendix for more detailed information on clinical outcome instruments used.

At baseline, two validated scales were used to measure symptoms of depression and anxiety: the Patient Health Questionnaire-9 (PHQ-9)[29] and the Generalised Anxiety Disorder 7-item scale (GAD-7).[30]

Measures of mood and subjective stressfulness of each of the two virtual environments were used, with participants completing a blinded virtual VAS for 'feelings of stress' (0 least stressed, 10 most stressed)

Table 1. Comparison of the VR activities investigated in Data Collection Session 2.

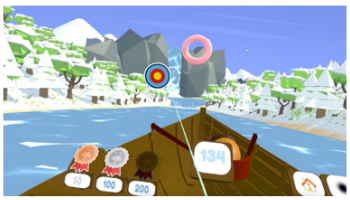
Activity type	Main skills required	Short activity description	Image
Rail-shooter	Hand-eye co-ordination and agility	To aim at and shoot targets appearing in the sky whilst travelling steadily along a calm river	
Memory	Short-term and working memory	To recall sequences of objects emerging from plots within a 3x3 grid. Upon correct recall, sequences increased in length	
Multitasking	Strategic planning and task prioritisation	To follow a series of steps to plant, maintain and harvest trees across five plots simultaneously	
Match-3	Pattern recognition, visuospatial awareness and strategic planning	To match three of the same objects in a row across three vines by moving vines up or down	



Figure 3. Images of the VR application used. Top left: Starting menu area; Top right: VR data collection questionnaire with virtual VAS; Bottom left and right: Cold and warm environments respectively.

Table 2. A summary table describing the evaluation instruments used and their frequency of administration.

Questionnaire	Frequency of administration	Session used	Description
Widespread Pain Index (WPI) & Symptom Severity Scale (SSS) Score[18]	Baseline only	Session 1	To measure Fibromyalgia Severity (FS) and ensure participants fulfil 2016 FMS diagnostic criteria
Revised Fibromyalgia Impact Questionnaire (FIQR)[28]	Baseline only	Session 1	To determine the impact of FMS on activities of daily living, overall wellbeing and severity of symptoms
Patient Health Questionnaire-9 (PHQ-9)[29]	Baseline only	Session 1	To determine presence and severity of depressive symptoms
Generalised Anxiety Disorder Assessment 7 (GAD-7)[30]	Baseline only	Session 1	To determine presence and severity of anxiety symptoms
McGill Pain Questionnaire – Short Form[31]	Pre- and post-intervention	All	To understand baseline and the changes in self-reported pain severity and character
Visual Analogue Scale – Pain	Pre- and post-intervention	All	To understand baseline and the changes in self-reported pain severity
Subjective experience questionnaires	Post-intervention	All	Specific questionnaires for each session used to gain information on acceptability, feasibility, tolerability and perceptions of VR following interventions
Virtual Reality Sickness Questionnaire (VRSQ)[32]	Post-intervention	All	To quantify adverse effects experienced with interventions
Virtual Visual Analogue Scale – Enjoyment, engagement, difficulty, length of time, motivation to perform well related to VR activity (Virtual VAS)	Post-intervention	Session 2	To quantify participant experience of different VR activities
Virtual Visual Analogue Scale – Enjoyment, immersion, feelings of stress, changes in mood related to virtual environment	Post-intervention	Session 3	To quantify participant experience of different virtual environments

and 'changes in mood' (0 worsened mood, 10 improved mood) within session 3. Multiple choice and white space questions were included within subjective experience questionnaires in sessions 2 and 3 to understand post-intervention changes in mood.

Data analysis

Statistical analyses were conducted in RStudio 2022.02.1 Build 461. Given the feasibility nature of this study, sample size was not calculated based on efficacy outcomes. A target of between 20 and 30 participants for recruitment was set as this sample range is large enough to enable meaningful qualitative interpretation and to detect obvious flaws in the study design, yet resource-efficient considering the nature of the investigation.

Descriptive statistics were used to summarise demographic data. Normality of continuous data was assessed using the Shapiro-Wilk test. Categorical data were reported as absolute numbers, while continuous data were reported as means, standard deviations, and ranges where appropriate. Changes in means with 95% confidence intervals (to provide an estimate of the precision of the findings) were reported for pain outcome data within the Appendix. Data from 7-point rating scale questions were dichotomised, with scores of 5-7 categorised as agreement with the statement, and scores of 1-3 as disagreement.

For normally distributed data, t-tests were used, such as in comparing acceptability rating total scores across the two virtual environments (warm vs cold) in Data Collection Session 3. For non-normally distributed data, a Wilcoxon Signed Rank test was used for pairwise comparisons, including the VRSQ total scores and total acceptability rating scores between VR systems. To compare acceptability ratings across multiple VR systems, a Friedman test was used. A repeated measures correlation was conducted to examine the relationship between baseline and post-VR fatigue reporting.

Results

See Figure 4 for a flow diagram showing participant numbers at each phase of the study.

26 participants entered the data collection phase. Patient demographic characteristics were similar between data collection sessions (Table 3) with the majority being female, white British and >45 years of age

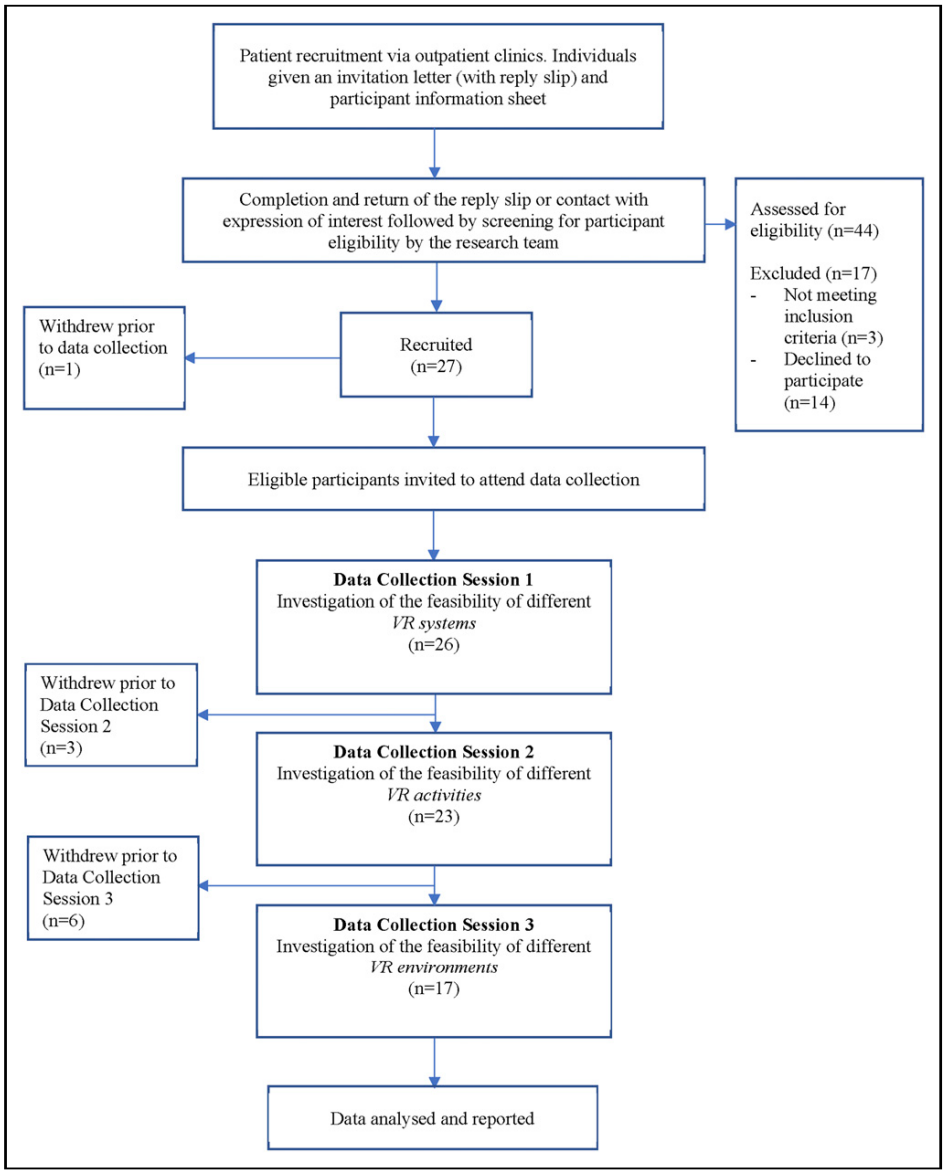


Figure 4. Flow diagram showing overall study flow, processes and participant numbers.

(mean age at recruitment 48.2 ± 14.3, range 20-72yrs). Most had no previous exposure to VR prior to participation (n = 21 out of 26). FMS disease severity ranged from mild to high[34] (mean FIQR at baseline 65.0 ± 16.2 (moderate severity), total FIQR score range 35.5-92.0).

Acceptability outcomes

See Figure 5 for a radar diagram outlining the key acceptability findings across VR systems, activities and virtual environments.

Virtual reality systems. Participants indicated high levels of acceptability across the VR systems when asked to rate their agreement with 15 statements related to system comfort, audio-visual quality, usability (including controllers), and immersion. The mean total scores for these ratings ranged from 77 to 97 out of 105. A comparative analysis between the four VR systems indicated overall differences in reported acceptability ($p < 0.0001$, Friedman test), with the Oculus Rift S being the most acceptable VR system ($p < 0.0001$, Wilcoxon Signed Rank test).

Table 3. Participant characteristics at baseline across the three data collection sessions.

Variable	Number of participants [unless stated otherwise]		
	Session 1: VR systems	Session 2: VR activities	Session 3: VR environments
n	26	23	17
BMI (kg/m^2) [mean (SD)]	29.5 (6.6)	30.0 (6.9)	31.0 (7.3)
Duration of FMS diagnosis (years) [mean (SD)]	5.2 (7.5)	4.3 (5.9)	3.3 (5.0)
FS [mean (SD)]	21.5 (4.9)	21.7 (4.5)	21.7 (4.2)
FIQR [mean (SD)]	65.0 (16.2)	65.4 (16.6)	67.3 (13.9)
PHQ-9 [mean (SD)]	13.7 (5.3)	14.1 (5.5)	13.7 (4.8)
GAD-7 [mean (SD)]	9.5 (5.6)	9.9 (5.6)	9.5 (4.8)
Age			
Age (years) [mean (SD)]	48.2 (14.3)	46.4 (13.8)	46.7 (14.9)
Sex Assigned at birth			
Male:Female ratio	3:23	3:20	2:15
Ethnicity			
White British	24	21	16
Black/African/Caribbean	1	1	1
Other (unspecified)	1	1	0
Previous VR exposure			
No	21	19	14

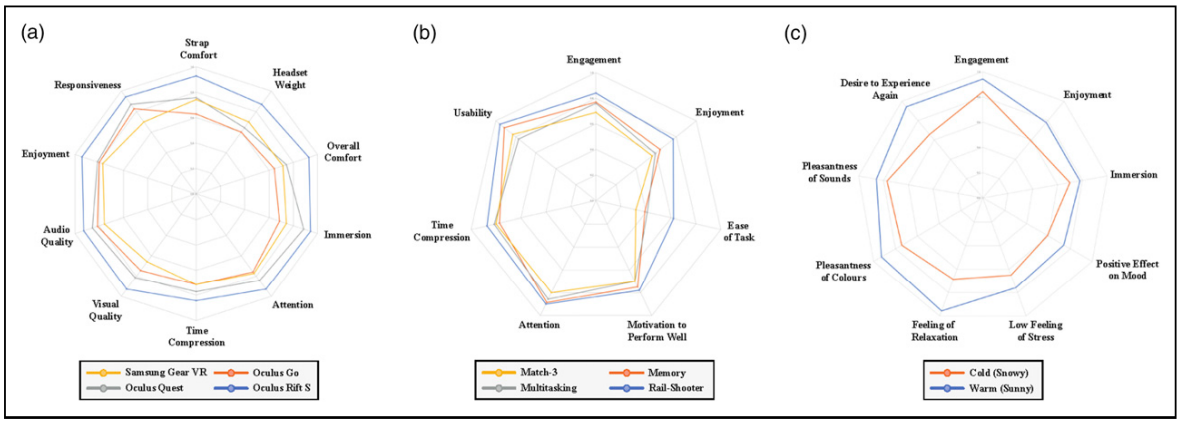


Figure 5. Radar diagram to show acceptability questionnaire scores across the VR systems, activities and environments[a]. [a] This diagram includes acceptability data collected through both rating questions (0-7 scale) and virtual VAS. Data has been presented using a standardised scale to aid interpretation (0 = lowest, 1 = highest scores).

Directly comparing the least acceptable headset (Oculus Go) and most acceptable headset (Oculus Rift S) revealed differences in comfort related to the strap and weight distribution in favour of the Oculus Rift S.

Virtual reality activities. Participants reported high levels of acceptability with VR when reporting on their experience with VR generally, across all activities. They rated their agreement with 7 statements enquiring about audio-visual comfort, usability, and engagement within the post-intervention subjective experience questionnaire. The mean total scores for these ratings ranged from 40 to 49 out of 49. Most participants agreed that being presented with a score based on their performance within the VR activities motivated them to perform better (n = 19 out of 23).

Participants reported high levels of engagement and immersion across activities, with most agreeing that they had 'more attention for the activities than for their own thoughts' and that they 'lost track of time whilst completing the activities' (time compression[35]). Upon comparison, the rail-shooter

activity was consistently rated the most engaging, user-friendly, and least difficult. In contrast, the match-3 task was rated most difficult and was less favoured in terms of engagement and enjoyment. When accounting for order effect within a mixed effects model, the match-3 task engendered significantly less attention than the rail-shooter task (p = 0.028), with no significant differences in time compression.

Virtual environments. Participants indicated high levels of acceptability across both virtual environments when asked to rate their agreement with five statements enquiring about audio-visual comfort, engagement and relaxation within the post-intervention subjective experience questionnaire (mean total scores 26-33 out of 35). Greater acceptability across the five statements was seen with the warm (sunny) virtual environment compared to the cold (snowy) environment (p = 0.003, t test), with superiority also seen in virtual VAS reporting. Most of the participants reported they would prefer to experience the warm virtual environment again in the future (n = 14 out of 17).

Adverse effects. A low rate of adverse effects were reported across VR systems, activities and environments through the VRSQ (Figure 6). The most commonly experienced adverse effects were 'general discomfort', 'fatigue' and 'eye strain', with most being 'mild' or 'moderate' and settling upon removal of the VR system. 'Fatigue' was the most likely adverse effect to persist after removing the VR system. This showed a general increasing trend when accounting for headset order in Data Collection Session 1, suggesting increased reporting with greater durations of VR use. Similarly, adverse effects were more common in Data Collection Session 2 (VR activity comparison), most likely as a consequence of longer durations of continuous VR use (approximately 30 minutes). There was also a positive correlation between baseline fatigue reporting (on the symptom severity scale) and reporting of moderate-severe fatigue post-VR (n = 11, r = 0.39).

When comparing VR systems, the Oculus Rift S produced significantly lower adverse effects than the Gear VR and Oculus Go (p = 0.013 & 0.018, Wilcoxon Signed Rank test). When asked to specify which VR activities led to adverse effects, similar numbers of participants reported adverse effects with each activity. Fewer adverse effects were reported with the warm compared with the cold virtual environment (VRSQ mean total scores 9.75 ± 10.55 vs 17.94 ± 13.62).

Other feasibility outcomes

Recruitment numbers are described in Figure 4. 14 individuals declined to participate upon contact, with the majority either feeling too unwell to participate, or conversely too well at the time of recruitment with minimal pain. Overall, 10 out of 26 participants dropped out of the study, with one participant dropping out before data collection commenced, three participants after Data Collection Session 1, and six participants after Data Collection Session 2. Reasons for drop out included ill health (n = 2), lack of time or practical difficulties with attending sessions (n = 2), inability to contact (n = 4), and other (n = 2). No participants stated their reasons for dropping out were due to adverse effects experienced with VR. All questionnaires were fully completed by participants, resulting in no missing data for the study.

Following use of all VR systems, participants reported their openness to using VR in the future at home and as part of their pain treatments. Assuming that VR was as effective as the participants' current method for managing pain and was made freely available, all of the participants agreed that they would be 'open to using VR technology at home regularly for the management of [their] pain' and most agreed that they could 'see a future for the use of VR as a pain management option' following completion of Data Collection Session 1 (n = 23 out of 26).

Clinical outcomes

Clinical outcome data relating to pain and mood are shown in the Appendix.

Reductions in both mean pain VAS and MPQ-SF total scores were reported across all VR systems, activities and virtual environments. In Data Collection Session 1, the Oculus Rift S produced the greatest pain reduction, and the Oculus Go the least. In Data Collection Session 3, the warm (sunny) virtual environment produced the greatest pain reduction, and the cold (snowy) environment the least. There was considerable variability in VAS reporting between participants after interventions. For example, one participant reported a reduction in VAS from 7.3 (baseline) to 1.4 (post-intervention) after using all VR systems in Data Collection Session 1, whereas another reported a small increase in VAS.

After experiencing all VR activities in Data Collection Session 2, the majority of participants reported either improvements or no change in their mood, with only small differences between activities.

In Data Collection Session 3, participants reported low subjective feelings of stress and strong feelings of

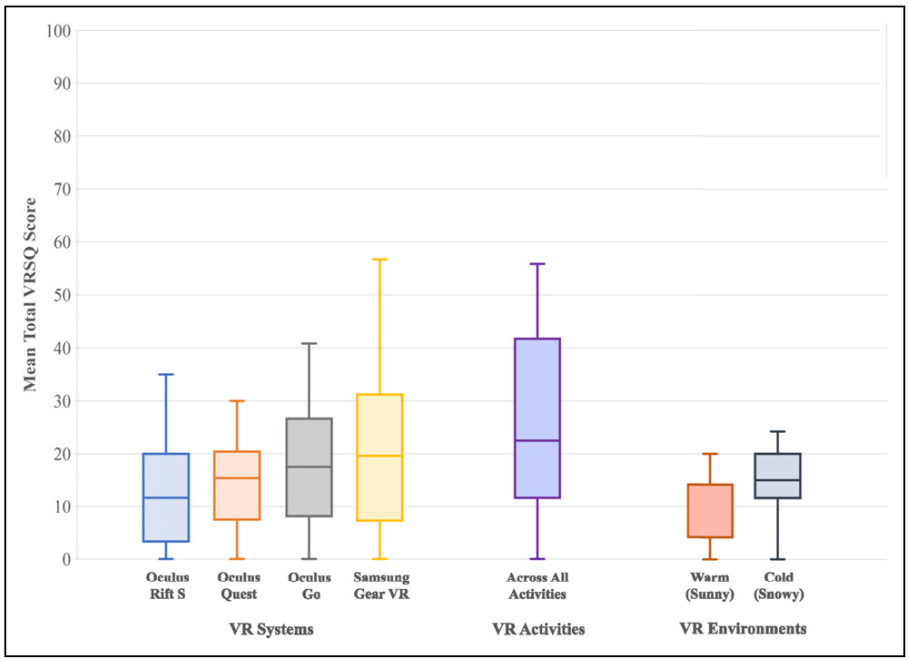


Figure 6. Adverse effects reported following use of each VR system through the Virtual Reality Sickness Questionnaire (VRSQ, maximum total score = 100).

relaxation with both virtual environments on the virtual VAS, with preference for the warm (sunny) environment (Figure 5). A considerably greater number of participants reported improvements in mood after experiencing the warm (sunny) virtual environment (n = 16 out of 17) compared with the cold (snowy) environment (n = 4 out of 17).

Discussion

Study overview and context

This study is the first to specifically investigate the feasibility of different head-mounted VR systems, environments and activities in patients with FMS. Across this multi-faceted study, the findings indicated that interactive VR designed to engage attention had good feasibility in patients with FMS, as shown by high levels of acceptability, low incidence of adverse effects, and positive perceptions. In common with previous VR pain studies, most patients with FMS felt interactive VR activities were highly immersive and effective at engaging attention. These findings provide additional evidence regarding the feasibility of using VR as a future digital therapeutic in FMS and provide information to inform future larger-scale studies.

This study builds on previous VR research within non-specific chronic pain[17,36] and chronic low back pain.[19,20] The recent longitudinal randomised controlled trials investigating VR interventions in low back pain demonstrate the efficacy and utility of VR as a home-based non-pharmacological treatment, leading to one application (RelieVRx) gaining approval from the Food and Drug Administration (FDA). However, given people living with FMS have specific challenges which would likely influence their experience of VR, studies in this cohort are essential. The few studies investigating 'virtual reality' specifically in FMS have mostly used non-immersive (not head-mounted) systems such as 2D projection screens.[37] These studies were not designed to investigate modern, head-mounted, interactive VR interventions aiming to engage attention. Similarly, most therapeutic VR studies evaluate one VR system, activity and virtual environment without considering the impact of each component.

Challenges and considerations with VR system components

VR interventions are 'complex' given they have multiple interacting components that each can influence patient outcomes.[38,39] The strength of this study lies in its multi-faceted approach to investigate each of the important components that make up an attention-

based VR intervention. Despite positive acceptability reported across VR systems, it is clear from this study that suboptimal audio-visual quality, latency, strap comfort and weight distribution can adversely affect feelings of immersion, engagement of attention and overall enjoyment in FMS. Consequently, these system-specific features can lead to greater adverse effects, and the possibility that this could impact clinical outcomes should not be ignored. Although the VR systems used in this study are now outdated by newer, more technologically advanced versions, there remains considerable variation in available devices with more customisation options than even before. This study's findings are particularly important in this context, as they highlight the importance of optimising comfort through carefully considered VR system setup to enhance user experience and minimise adverse effects. Clear and systematic reporting of such adverse effects is essential yet remains inconsistent across VR studies. To address this, future research should prioritise the development of standardised protocols for assessing and reporting adverse effects to ensure a comprehensive understanding of safety, especially for populations with specific needs like those with FMS.

Significance of the virtual environment

Study findings also demonstrated the importance of virtual environment design when creating VR content for patients with FMS. Perhaps as a consequence of the early work conducted by Hoffman et al utilising cold virtual environments for patients with burns injuries,[16] other studies have also used cold environments for chronic pain.[17] In contrast, our findings indicate improved acceptability related to virtual environments simulating warm conditions in FMS. This may be disease-specific, given the unique sensory hypersensitivity and thermodysregulatory states seen in FMS. It may also be an indication of a more complex relationship informed by a person's pain characteristics or previous experiences, indicating the need to tailor the virtual environment to the user. Qualitative responses perhaps explain differences in reported acceptability, with participants indicating an association with relaxation in warm settings and a perceived worsening of symptoms in cold settings. This is also shown in our mood data, with the majority indicating an improvement in their mood with the warm virtual environment. Although no studies have specifically explored the role of patients' phobias or past experiences in this context within relevant chronic pain populations, this is an important consideration that could influence the acceptability and safety of VR interventions. Regardless of the mechanism, it is clear that the virtual environment is a crucial element in VR chronic pain therapeutics. Further research should investigate a range of environments in well-characterised pain cohorts and consider evaluating factors like phobias or past experiences before starting VR therapy to better tailor environments to individual needs.

The importance of fatigue with VR

This study's findings provide reassurance of the safety of using VR in FMS, with low adverse effects reported across VR systems, activities and environments. However, a general increase in fatigue reporting with longer VR sessions was noted, particularly apparent during Data Collection Session 2 where VR use duration was at its peak. Upon closer examination, there was a relationship between baseline fatigue levels and subsequent fatigue reporting with VR use. Considering that fatigue is a defining characteristic of FMS and other chronic pain conditions, these indirect findings underscore the importance of customising VR 'dosing' for high-risk groups. This includes integrating safety measures such as 'pacing checkpoints' and conducting more targeted research to understand the 'dose-response' relationship of VR in specific disease subgroups.

Perceptions and tailored VR interventions

Prior to this study, the perceptions of individuals with FMS regarding VR use for pain management were unclear. Although only within a small, select sample of patients, study findings demonstrated positive attitudes towards the use of VR for future, home-based therapy.

The rapid evolution of VR in recent years, particularly with the advent of affordable and portable systems, has opened new avenues for the home-based management of chronic pain. With the increasing availability of VR, it is inevitable that the use of VR therapeutics will become more widespread. Our data in FMS highlights the risk of adopting a generalised approach with use of VR applications across disease subtypes that are unsuitable and may lead to adverse effects.

To ensure the acceptability and safety of VR interventions, future research and application development should promote: (1) co-design with individuals who have lived experience of chronic pain, (2) consideration of each VR component's impact on acceptability, tolerability, and usability, and (3) evaluation in specific chronic pain subgroups, taking into account unique disease-specific challenges. This approach will help tailor VR applications to meet the diverse needs of chronic pain patients.

Limitations and recommendations for future research

Limitations of the present study need to be acknowledged. The study was primarily designed to determine feasibility, with a small sample size appropriate for the study design. While the study sample was representative of the UK FMS population,[40] the inclusion of mostly white females based in the East of England limits the generalisability of findings. Measurement of outcomes at multiple time points provided a good understanding of intervention feasibility and helped reduce recall bias by collecting relevant data for each intervention as close to the intervention completion as possible. Despite this, use of self-report questionnaires risks recall bias, particularly with questionnaires comparing interventions at the end of the session. The use of non-validated subjective experience questionnaires and virtual scales may limit the generalisability and robustness of the findings. These tools were specifically developed for each data collection session to gather targeted feedback on comfort, usability, and overall user experience. Available validated scales were either not specific for VR interventions, or evaluated single constructs like immersion and lacked the specificity needed for assessing factors such as the comfort of individual VR system components.[41] Data on the ease of completion of the questionnaires for participants or of data analysis for researchers was not collected. The sequential nature of the data collection sessions in this study leads to the possibility of intervention novelty influencing outcomes. This should be taken into account for future studies. The one-week washout period aimed to minimise potential carryover effects between sessions. However, due to limited knowledge about the duration of effects from attention-based VR, some risk of residual influence remains. Future trials should account for this uncertainty in their design to ensure accurate interpretation of outcomes.

While the absence of a control intervention is not a specific limitation for our feasibility study, it is an important consideration for future VR trials evaluating effect. The VR-CORE international working group emphasise the lack of a 'perfect' or 'standardised' control condition for VR therapeutics research, with the choice of control depending on the research question.[21] Our feasibility study focused on evaluating short-term outcomes, but future studies should consider longer follow-up to gain insights into the sustained feasibility of VR use, including the persistence of benefits or potential delayed adverse effects with extended use.

The overall attrition rate of 37.0% resulted in complete datasets for 16 participants, falling short of our target sample size of 20-30. Multiple data collection sessions organised over time may have increased the risk of participant dropout, especially given the fluctuating nature of symptoms in individuals with FMS. While participants did not cite COVID-19 as a direct reason for dropout, many communicated difficulties with transport and increased anxieties related to social contact during the pandemic. Additionally, national lockdowns required us to pause data collection, extending the data collection period to 10 months, which may have further influenced retention. These factors should be considered when interpreting our attrition rate, as they represent barriers that may not be as prevalent in future studies. Despite these challenges, the high rate of questionnaire completion by those who attended suggests that the outcome measures were not overly burdensome and could be suitable for future research. To account for attrition in similar studies, a larger initial sample size could help ensure that sufficient data are collected. Future trials should carefully consider using multiple lab-based data collection sessions over time, and may also consider remote or hybrid data collection methods to reduce barriers to participation.

Conclusion

In conclusion, this study increases confidence with respect to the feasibility of using attention-based, head-mounted VR interventions in patients with FMS. Strong acceptability, minimal adverse effects and positive perceptions of using VR highlight the potential of this technology to support future home-based, non-pharmacological treatments in FMS. However, it is clear that VR interventions are complex with the need to consider the acceptability of each component of the VR system, environment and activity when attempting to improve clinical outcomes. It is therefore essential that we avoid a 'one-size-fits-all' approach to future VR therapeutics, and instead promote co-development and tailoring of interventions to the patients that they are designed for.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the British Society for Rheumatology, Action Arthritis, Enabling Innovation: Research to Application.

ORCID iD

Jordan Tsigarides https://orcid.org/0000-0001-9893-8002

Supplemental Material

Supplemental material for this article is available online.

References

1. Heidari F, Afshari M and Moosazadeh M. Prevalence of fibromyalgia in general population and patients, a systematic review and meta-analysis. *Rheumatol Int* 2017; 37: 1527–1539. DOI: 10.1007/s00296-017-3725-2.
2. Vincent A, Lahr BD, Wolfe F, et al. Prevalence of fibromyalgia: a population-based study in Olmsted county, Minnesota, utilizing the rochester epidemiology project. *Arthritis Care Res* 2013; 65: 786–792. DOI: 10.1002/acr.21896.
3. Berwick R, Barker C and Goebel A. The diagnosis of fibromyalgia syndrome. *Clin Med* 2022; 22: 570–574. DOI: 10.7861/clinmed.2022-0402.
4. Rivera J, Molina-Collada J, Martínez-Barrio J, et al. Opioids and fibromyalgia: frequency of use and factors associated with increased consumption in patients remitted to a tertiary care center. *BMC Muscoskel Disord* 2024; 25: 121. DOI: 10.1186/s12891-024-07263-x.
5. NIfHaC Excellence. *Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain*. National Institute for Health and Care Excellence (NICE), 2023.
6. Macfarlane GJ, Kronisch C, Dean LE, et al. EULAR revised recommendations for the management of fibromyalgia. *Ann Rheum Dis* 2017; 76: 318–328. DOI: 10.1136/annrheumdis-2016-209724.
7. Becker WC, Dorflinger L, Edmond SN, et al. Barriers and facilitators to use of non-pharmacological treatments in chronic pain. *BMC Fam Pract* 2017; 18: 41. DOI: 10.1186/s12875-017-0608-2.
8. Dreesmann NJ, Su H and Thompson HJ. A systematic review of virtual reality therapeutics for acute pain management. *Pain Manag Nurs* 2022; 23: 672–681. DOI: 10.1016/j.pmn.2022.05.004.
9. Spiegel B, Fuller G, Lopez M, et al. Virtual reality for management of pain in hospitalized patients: a randomized comparative effectiveness trial. *PLoS One* 2019; 14: e0219115. DOI: 10.1371/journal.pone.0219115.
10. Eccleston C and Crombez G. Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychol Bull* 1999; 125: 356–366. DOI: 10.1037/0033-2909.125.3.356.
11. Davydov DM, Galvez-Sánchez CM, Montoro CI, et al. Personalized behavior management as a replacement for medications for pain control and mood regulation. *Sci Rep* 2021; 11: 20297. DOI: 10.1038/s41598-021-99803-x.
12. Tsigarides J, Shenker NG and MacGregor A. Ready Patient One: the role of therapeutic virtual reality in the future management of chronic pain. *Rheumatology* 2022; 61: 482–483. DOI: 10.1093/rheumatology/keab594.
13. Lier EJ, de Vries M, Steggink EM, et al. Effect modifiers of virtual reality in pain management: a systematic review and meta-regression analysis. *Pain* 2023; 164: 1658–1665. DOI: 10.1097/j.pain.0000000000002883.
14. Malloy KM and Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. *Clin Psychol Rev* 2010; 30: 1011–1018. DOI: 10.1016/j.cpr.2010.07.001.
15. Tesarz J, Herpel C, Meischner M, et al. Effects of virtual reality on psychophysical measures of pain: superiority to imagination and nonimmersive conditions. *Pain* 2024; 165: 796–810. DOI: 10.1097/j.pain.0000000000003083.
16. Hoffman HG, Richards TL, Coda B, et al. Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI. *Neuroreport* 2004; 15: 1245–1248. DOI: 10.1097/01.wnr.0000127826.73576.91.
17. Jones T, Moore T and Choo J. The impact of virtual reality on chronic pain. *PLoS One* 2016; 11: e0167523. DOI: 10.1371/journal.pone.0167523.
18. Wolfe F, Clauw DJ, Fitzcharles MA, et al. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin Arthritis Rheum* 2016; 46: 319-329. DOI: 10.1016/j.semarthrit.2016.08.012.
19. Eccleston C, Fisher E, Liikkanen S, et al. A prospective, double-blind, pilot, randomized, controlled trial of an "embodied" virtual reality intervention for adults with low back pain. *Pain* 2022; 163: 1700–1715. DOI: 10.1097/j.pain.0000000000002617.
20. Garcia LM, Birckhead BJ, Krishnamurthy P, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. *J Med Internet Res* 2021; 23: e26292. DOI: 10.2196/26292.
21. Birckhead B, Khalil C, Liu X, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. *JMIR Ment Health* 2019; 6: e11973. DOI: 10.2196/11973.
22. Reynolds LM, Cavadino A, Chin S, et al. The benefits and acceptability of virtual reality interventions for women with metastatic breast cancer in their homes; a pilot randomised trial. *BMC Cancer* 2022; 22: 360. DOI: 10.1186/s12885-021-09081-z.
23. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. *Am J Prev Med* 2009; 36: 452–457. DOI: 10.1016/j.amepre.2009.02.002.
24. Gromala D, Tong X, Choo A, et al. The virtual meditative Walk: virtual reality therapy for chronic pain

management *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 521–524.
25. Csikszentmihalyi M. *Flow: the psychology of optimal experience*. New York: Harper-Collins, 1990.
26. Hoffman HG, Richards TL, Bills AR, et al. Using FMRI to study the neural correlates of virtual reality analgesia. *CNS Spectr* 2006; 11: 45–51. DOI: 10.1017/s1092852900024202.
27. Berwick RJ, Siew S, Andersson DA, et al. A systematic review into the influence of temperature on fibromyalgia pain: meteorological studies and quantitative sensory testing. *J Pain* 2021; 22: 473–486. DOI: 10.1016/j.jpain.2020.12.005.
28. Bennett RM, Friend R, Jones KD, et al. The revised fibromyalgia impact questionnaire (FIQR): validation and psychometric properties. *Arthritis Res Ther* 2009; 11: R120. DOI: 10.1186/ar2783.
29. Kroenke K, Spitzer RL and Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16: 606–613. DOI: 10.1046/j.1525-1497.2001.016009606.x.
30. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; 166: 1092–1097. DOI: 10.1001/archinte.166.10.1092.
31. Melzack R. The short-form McGill pain questionnaire. *Pain* 1987; 30: 191–197. DOI: 10.1016/0304-3959(87)91074-8.
32. Kim HK, Park J, Choi Y, et al. Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl Ergon* 2018; 69: 66–73. DOI: 10.1016/j.apergo.2017.12.016.
33. Sevinc V and Berkman MI. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Appl Ergon* 2020; 82: 102958. DOI: 10.1016/j.apergo.2019.102958.
34. Salaffi F, Di Carlo M, Arcà S, et al. Categorisation of disease severity states in fibromyalgia: a first step to support decision-making in health care policy. *Clin Exp Rheumatol* 2018; 36: 1074–1081.
35. Mullen G and Davidenko N. Time compression in virtual reality. *Timing & Time Percept* 2021; 9: 377–392. DOI: 10.1163/22134468-bja10034.
36. Darnall BD, Krishnamurthy P, Tsuei J, et al. Self-administered skills-based virtual reality intervention for chronic pain: randomized controlled pilot study. *JMIR Form Res* 2020; 4: e17293. DOI: 10.2196/17293.
37. Cortés-Pérez I, Zagalaz-Anula N, Ibancos-Losada MDR, et al. Virtual reality-based therapy reduces the disabling impact of fibromyalgia syndrome in women: systematic review with meta-analysis of randomized controlled trials. *J Personalized Med* 2021; 11: 20211109. DOI: 10.3390/jpm11111167.
38. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new medical research council guidance. *Br Med J* 2008; 337: a1655. DOI: 10.1016/j.ijnurstu.2012.09.010.
39. Rodriguez A, Smith J and Barrett D. Research made simple: developing complex interventions. *Evid Base Nurs* 2020; 23: 35–37. DOI: 10.1136/ebnurs-2020-103261.
40. Cooksey R and Choy E. Exploring gender differences, medical history, and treatments used in patients with fibromyalgia in the UK using primary-care data: a retrospective, population-based, cohort study. *The Lancet Rheumatology* 2022; 4: S20. DOI: 10.1016/S2665-9913(22)00296-X.
41. Jennett C, Cox AL, Cairns P, et al. Measuring and defining the experience of immersion in games. *Int J Hum Comput Stud* 2008; 66: 641–661. DOI: 10.1016/j.ijhcs.2008.04.004.

Title & Description: Appendix 1 – Specifications of the VR Application and Systems Used

	VR System			
	Samsung Gear VR	**Oculus Go**	**Oculus Quest**	**Oculus Rift S**
Image				
Internal Chipset	Exynos 8895	Qualcomm Snapdragon 821	Qualcomm Snapdragon 835	N/A[a]
Random-access Memory (RAM)	4gb	3gb	4gb	N/A[a]
Screen Resolution (per eye)	1280x1440	1280x1440	1440x1600	1280x1440
Screen refresh rate	60hz	72hz	72hz	80hz
Weight	500g	468g	571g	563g

Table 4: Comparison of the VR system specifications.

[a] A Dell Mobile Precision 7740 laptop (Windows 10 Pro, Intel Core i9 processor, 64gb DDR4 ram, Nvidia Quadro RTX 5000 16gb GDDR6) streamed VR content to the Rift S, tethered through a USB-Display Port cable

The VR Application Technical Specifications

The application was built and tested using Unity 2019.4. It was designed to be stereoscopic with the ability for the user to perceive depth and three-dimensional structures. The untethered systems (Gear VR, Oculus Go and Oculus Quest) used an Android build, whereas the tethered Oculus Rift S used a Windows™ build. Data collected from 'virtual' VR questionnaires and on program use (including timestamped data on environment/activity selection and activity performance) were stored within a unique JSON file written directly to

the study computer. This data was only recorded during Data Collection Sessions 2 and 3 when using the Oculus Rift S (tethered to the study laptop).

Custom head mounting straps were used with the Oculus Go and disposable fascia covers used with the Gear VR to enable the hygiene protocol to be completed.

The VR Activities Description

Rail-Shooter Activity – A 'rail shooter' activity requiring hand-eye co-ordination and agility. Travelling slowly aboard a boat down a calm river, the user is presented a variety of targets that they are required to aim at and hit. There are two different target types, with some moving and some remaining static. A score was earned when hitting targets, with higher scores provided with hits closer to the centre of the target.

Memory Activity – A sequence recall activity requiring short-term memory and spatial awareness. A 3x3 grid of soil plots presents a sequence of emerging plants to the user who has to replicate the sequence. This sequence escalates in length (starting at 3), adding one extra to the sequence each time the user correctly replicates the preceding sequence until failure. Upon the first instance of an incorrect selection, the user is given a warning prior to the sequence resetting back to 3 if they select the incorrect plot on two occasions sequentially. The user gained a score equal to the length of the sequence upon correctly recalling and replicating the sequence.

Multitasking Activity – A multitasking activity requiring working memory and reasoning skills. A patch of soil with five plots, a watering can and seed container are presented to the user. The user is challenged to follow a sequence of four tasks (digging a plot, planting a seed, watering and harvesting) for each of the five plots in order to grow trees. Each harvested tree provided a score for the user, with the requirement to manage multiple plots simultaneously.

Match-3 Activity – A 'match three' activity requiring visual searching, visuospatial awareness and problem-solving skills. Three vines with multiple vegetables attached are presented to the user. Pulling each vine downwards or upwards changed the position of the vegetable and revealed a new vegetable from above or below (depending on the direction the vine is pulled). When the user matched three identical vegetables in a horizontal row they would gain a positive score. An escalating 'multiplier' was gained for successful matching of three items in quick succession, improving the overall score gained.

Title & Description: Appendix 2 – Detailed Description of the Outcomes Measures Used

Disease Severity Measures (Baseline)

Widespread Pain Index (WPI) and Symptom Severity Scale (SSS) Scores:

The Widespread Pain Index (WPI) is a questionnaire enquiring about pain experienced across 19 body areas, to evaluate the presence and distribution of pain experienced by the patient over the past week.

The Symptom Severity Scale (SSS) is a questionnaire with two parts: the first part includes 13 questions assessing the severity of fatigue, waking unrefreshed, cognitive symptoms, and the extent of somatic symptoms in general, while the second part evaluates the number of somatic symptoms experienced.

The sum of the Widespread Pain Index (WPI) and Symptom Severity Scale (SSS) scores are used to calculate the Fibromyalgia Severity (FS) Score to categorise FMS symptom severity (FS, total score 0-31).

Revised Fibromyalgia Impact Questionnaire (Revised FIQR):

Self-report questionnaire to determine impact of FMS on activities of daily living, overall wellbeing and severity of symptoms. 21 questions rated on a 0-10 numeric rating scale (NRS) assessing the impact of FMS on three domains (1) function, (2) overall impact, and (3) symptoms. The total FIQR score is the sum of the three domain scores (0-100).

Depression and Anxiety Measures (Baseline)

Patient Health Questionnaire 9 (PHQ-9):

Self-report questionnaire to determine the presence and severity of depressive symptoms. 9 items each scored from 0 (not at all) to 3 (nearly every day). Total scores range between 0-27.

Generalised Anxiety Disorder Assessment 7 (GAD-7):

Self-report questionnaire to determine the presence and severity of anxiety symptoms. 7 items scored from 0 (not at all) to 3 (nearly every day). Total scores range between 0-21.

Subjective Experience Questionnaires (Post-Intervention)

These questionnaires included a combination of rating questions, asking participants to rate their level of agreement with statements related to acceptability on a scale of 1-7 (1 strongly disagree, 7 strongly agree), multiple choice questions, white space questions (to contextualise quantitative self-report), and virtual VAS ratings (10-point blinded scale). Questionnaires varied between data collection sessions and were tailored to the intervention:

1. VR Systems:
    - After use of a VR system: 15 rating questions and 3 white space questions.
    - After use of all VR systems: 6 multiple choice questions (asking the participant to choose the best performing headset for a variety of acceptability questions) and 3 rating questions enquiring about perceptions of VR.
2. VR Activities:
    - After experiencing a VR activity: 5 virtual VAS rating questions enquiring about enjoyment, engagement, difficulty, length of time in activity and motivation to perform the activity.

- After experiencing all VR activities: 8 initial rating questions (generic acceptability across the activities), followed by 3 further rating questions enquiring about perceptions of VR. This was followed by 20 rating questions (5 per activity, activity-specific acceptability), 8 multiple choice questions (asking about specific effects on pain and mood for each activity) and 12 white space questions (3 per activity).

3. VR Environments:
    - After experiencing a VR environment: 4 virtual VAS rating questions enquiring about enjoyment, immersion, feelings of stress and changes in mood related to the virtual environment.
    - After experiencing both VR environments: 2 multiple choice questions asking participants about the effect of each environment on pain and mood with 2 associated white space questions. 10 initial rating questions (5 per environment) enquiring about acceptability of the colours and sounds, in addition to engagement, feelings of relaxation and desire to experience the environment again in the future. 1 further multiple choice question asking the participant to choose which one environment they would rather experience again with white space questions to describe likes and dislikes of each. 3 further rating questions enquiring about perceptions of VR.

The Virtual Reality Sickness Questionnaire (VRSQ, Post-intervention)

Participants completed the VRSQ after use of each VR headset in Data Collection Session 1 and after experiencing all environments/activities in sessions 2 and 3. Participants were asked to rate the severity of each symptom on a four-point scale ranging from 'None' (0) to 'Severe'

(3). VRSQ scores were calculated by first summing the scores for each symptom in the two categories. These raw scores were then weighted according to the procedure outlined by Kim et al. (2018): oculomotor score was the sum of raw scores divided by 12 and then multiplied by 100, and disorientation was divided by 15 and multiplied by 100. The total VRSQ score was computed as the sum of the oculomotor and disorientation scores divided by 2. The VRSQ allowed assessment of both the incidence (i.e., the number of participants reporting symptoms) and severity (i.e., the intensity of symptoms reported) of adverse effects post-VR.

Pain Intensity (Pre/Post Intervention)

To quantify participants' perception of pain, we implemented a multimodal approach using the McGill Pain Questionnaire Short Form (MPQ-SF), a Visual Analog Scale (VAS), and a custom post-intervention questionnaire.

The MPQ-SF, a brief yet robust tool for pain assessment, was administered to all participants both before (pre-exposure baseline) and after (post-exposure) intervention. This instrument includes 15 pain descriptors (11 sensory; 4 affective), which participants rate based on their intensity from 0 (none) to 3 (severe). The MPQ-SF yields three scores: the sensory, affective, and total pain scores, providing a multidimensional representation of the pain experience.

The VAS, a simple and reliable tool for subjective pain measurement, was also employed. Two forms of the VAS were used, a physical paper-based version and a ‘virtual’ version. For the written VAS, participants were asked to rate their pain intensity on a 10cm line, ranging from "no pain" on one end (0cm) to "worst possible pain" on the other end (10cm). The ‘virtual’ VAS replicated the written VAS but used a 10-point blinded scale (participants were only able to appreciate the position of a marker on a line, without labelling of numbers) with data recorded by the application. The ‘virtual’ pain VAS was completed in VR after

experiencing each environment and activity during Studies 2 and 3. The written pain VAS was completed at baseline and post-intervention in all studies, as well as after use of each VR system in Study 1. Recording of VAS pre- and post-intervention, and between multiple interventions, allowed for a direct comparison of pain levels at different time points.

Repeated measures correlation analysis demonstrates a correlation coefficient of 0.403 ($p<001$) when assessing the correlation of virtual and written VAS reporting in Data Collection Session 2. In line with work by Boonstra et al (2014), interpretation of VAS scores are reported as follows: mild pain ($\leq$3.4cm), moderate pain (3.5-7.4cm), and severe pain ($\geq$7.5cm).

In addition to the MPQ and VAS, we included questions within the subjective experience questionnaires to gain information about participants' perception of pain in relation to specific interventions. Participants indicated whether each intervention increased, decreased, or did not change their perception of pain using a multiple-choice format. Additionally, participants were asked to complete white space questions to explain why they felt their pain levels were impacted.

Title & Description: Appendix 3 – Clinical Outcome Data Tables

Pain reporting

Reporting Time Point	Pain VAS [mean (SD)]	Difference in VAS from Session Baseline [mean (95% CIs)]	Number achieving MCID in VAS	MPQ-SF total score [mean (SD)]	Difference in MPQ-SF from Session Baseline [mean (95% CIs)]
Data Collection Session 1 (VR Systems, n=26)					
Session 1 Baseline	5.5 (1.9)	-	-	24.3 (9.0)	-
Post-VR System: Samsung Gear VR	4.7 (1.8)	-0.8 (-1.3, -0.4)	12	-	-
Post-VR System: Oculus Go	4.9 (1.8)	-0.6 (-1.2, -0.1)	8	-	-
Post-VR System: Oculus Quest	4.3 (1.6)	-1.2 (-1.9, -0.5)	13	-	-
Post-VR System: Oculus Rift S	3.9 (1.8)	-1.6 (-2.3, -0.9)	19	-	-
Post-Intervention (following use of all VR systems)	4.0 (1.9)	-1.5 (-2.2, -0.8)	16	14.6 (9.1)	-9.7 (-13.0, -6.4)
Data Collection Session 2 (VR Activities, n=23)					
Session 2 Baseline	5.5 (2.3)	-	-	21.7 (8.8)	-
Post-Activity: Rail-shooter	4.1 (2.4)	-1.4 (-2.1, -0.6)	14	-	-
Post-Activity: Memory	4.0 (2.4)	-1.5 (-2.2, -0.7)	13	-	-
Post-Activity: Multitasking	4.3 (2.4)	-1.2 (-2.0, -0.4)	13	-	-
Post-Activity: Match-3	4.0 (2.2)	-1.5 (-2.1, -0.8)	14	-	-

Post-Intervention: After Experiencing All Activities	4.5 (2.5)	-1.0 (-1.9, -0.1)	14	16.9 (10.5)	-4.8 (-7.4, -2.2)
Data Collection Session 3 (VR Environments, n=17)					
Session 3 Baseline	5.4 (2.1)	-	-	21.6 (10.1)	-
Post-environment: Warm (sunny)	4.2 (2.0)	-1.2 (-2.1, -0.4)	11	-	-
Post-environment: Cold (snowy)	4.6 (2.2)	-0.8 (-1.5, -0.1)	6	-	-
Post-Intervention: After Experiencing Both Environments	4.8 (2.5)	-0.6 (-1.4, 0.2)	8	15.2 (9.1)	-6.4 (-9.0, -3.8)

Table 5: Comparison of pain outcomes across the study. Minimal Clinically Important Difference (MCID) is defined as a reduction in VAS by ≥15% (as described by Salaffi et al in musculoskeletal disease[1])

Mood Reporting

Type of Activity or Environment	**Post-intervention: Number Reporting Improvement in Mood**	**Post-intervention: Number Reporting Worsening in Mood**
Data Collection Session 2 (VR Activities, n=23)		
Rail-shooter	10	1
Memory	9	4
Multitasking	9	4
Match-3	5	4
Data Collection Session 3		

(VR Environments, n=17)		
Warm (Sunny)	16	0
Cold (Snowy)	4	1

Table 6: Comparison of mood-related outcome measures across the VR activities and environments.

REFERENCES

1. Salaffi F, Stancati A, Silvestri CA, et al. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain* 2004; 8: 283-291. DOI: 10.1016/j.ejpain.2003.09.004.

Appendix 2: Description of study-specific questionnaires and outcome measures

This appendix summarises the questionnaire and self-reported outcome measures used in the PainWaves and VR-FOCUS studies. Its purpose is to provide additional detail on the content, scoring and study-specific role of these measures without duplicating the methods sections in Chapters 5 and 7. Measures used in the VIPA-EEG study are not reproduced here, as these were collected within the parent feasibility study and are described in the published article included as Appendix 1. Chapter 6 specifies which of those measures were used in the VIPA-EEG secondary analysis.

PainWaves study measures

London Fibromyalgia Epidemiology Study Screening Questionnaire (LFESSQ)

Questionnaire-derived screening measure used to identify widespread musculoskeletal pain. The LFESSQ pain component includes four items asking whether participants had pain in muscles, bones or joints lasting at least 1 week; pain in the shoulders, arms or hands; pain in the legs or feet; and pain in the neck, chest or back. For the upper- and lower-limb pain items, participants also indicated whether pain was present on the right side, left side or both sides. The LFESSQ pain criteria were met when all four pain items were reported and pain was distributed bilaterally, defined as either both right- and left-sided positive responses or a “both sides” positive response. In PainWaves, existing TwinsUK LFESSQ data were used to derive questionnaire-defined CWP status for twin-pair selection. The measure was used as a screening-derived phenotype rather than as a clinical diagnosis of fibromyalgia. Delivered historically as a paper version by the TwinsUK research team.

Patient Health Questionnaire-2 (PHQ-2)

Brief self-report screening measure of depressive symptoms over the previous 2 weeks. Two items assess low mood and anhedonia. Each item is scored from 0 (not at all) to 3 (nearly every day). Total scores range from 0-6, with higher scores indicating greater depressive symptom burden. A score of 3 or above is commonly used as the threshold for a positive depression screen, warranting further assessment rather than constituting a diagnosis[291]. Delivered as a paper version.

Generalised Anxiety Disorder-7 (GAD-7)

Self-report measure of anxiety symptoms over the previous 2 weeks. Seven items assess nervousness, uncontrollable worry, excessive worry, difficulty relaxing, restlessness, irritability and fear that something awful might happen. Each item is scored from 0 (not at all) to 3 (nearly every day). Total scores range from 0-21. Conventional severity categories are 0-4 minimal anxiety, 5-9 mild anxiety, 10-14 moderate anxiety and 15-21 severe anxiety. A score of 10 or above is commonly used as a clinically relevant screening threshold[292]. Delivered as a paper version.

Widespread Pain Index (WPI)

Self-report measure of pain distribution over the previous 7 days. Participants indicated whether they had experienced pain in 19 predefined body regions: right and left jaw, neck, right and left shoulder, right and left upper arm, right and left forearm, chest, abdomen, upper back, lower back, right and left buttock, right and left upper leg, and right and left lower leg. Each painful region contributes 1 point. Total scores range from 0-19, with higher scores indicating more widespread pain. In the revised fibromyalgia criteria, the WPI contributes to classification when considered alongside symptom severity and pain distribution[282, 283]. Delivered as a paper version.

Symptom Severity Scale (SSS)

Self-report measure of fibromyalgia-related symptom burden. Participants rated the severity of fatigue, waking unrefreshed and cognitive symptoms over the previous 7 days. Each item was scored from 0 to 3, where 0 indicates no problem, 1 indicates slight or mild problems, 2 indicates moderate or considerable problems, and 3 indicates severe, pervasive, continuous or life-disturbing problems. Participants also reported whether they had experienced headaches, pain or cramps in the lower abdomen, and depression over the previous 6 months, with each scored as absent or present. The total SSS score ranges from 0-12, with higher scores indicating greater symptom burden[282, 283]. Delivered as a paper version.

Fibromyalgia Severity (FS) score

Composite measure calculated by summing the WPI and SSS. Total scores range from 0-31, with higher scores indicating greater combined burden of widespread pain and fibromyalgia-related symptoms[282, 283]. In the 2016 fibromyalgia criteria[282], fibromyalgia may be classified when symptoms have been present at a similar level

for at least 3 months, generalised pain is present in at least 4 of 5 regions, and either WPI is 7 or above with SSS 5 or above, or WPI is 4-6 with SSS 9 or above. In PainWaves, these scores provided clinical characterisation of symptom burden rather than acting as the primary grouping variable. Delivered as a paper version.

Pain and fatigue visual analogue scales

State measures of current pain and fatigue intensity were collected immediately before EEG recording. Participants marked a vertical line on separate 10cm horizontal scales anchored by “no pain” and “worst possible pain” for pain, and “no fatigue” and “worst possible fatigue” for fatigue. The research team recorded the corresponding numeric VAS value by measuring the distance in centimetres from the left-hand anchor to the participant’s mark, with higher scores indicating greater current pain or fatigue intensity. Published pain severity cut-offs vary across settings; when 0-10 numeric categories are used, a common convention is 0 no pain, 1-3 mild pain, 4-6 moderate pain and 7-10 severe pain[378]. These categories should be interpreted cautiously and were not used as primary inferential thresholds. No equivalent categorical severity thresholds are recognised for fatigue VAS scores. Delivered as paper versions.

EEG-relevant contextual variables

Additional contextual variables were recorded at the time of EEG acquisition because of their potential relevance to EEG data quality or interpretation. These included head circumference, nasion-inion distance, whether the participant had washed their hair in the preceding 24 hours, caffeine consumption in the preceding 24 hours, sleep duration on the previous night, and whether the participant had woken during the night. These variables were not primary clinical outcomes, but provided useful contextual information for interpreting acquisition quality and potential sources of physiological variability during analysis.

VR-FOCUS study measures

Baseline questionnaire

Self-report questionnaire completed before VR exposure. Items captured demographic information, previous technology exposure, medical history, comorbidities and regular medications. All participants completed the baseline questionnaire. Participants with chronic pain completed additional pain-specific

measures. Delivered through a Microsoft Forms electronic format using a tablet or laptop.

Current pain and fatigue numeric rating scales

Participants with chronic pain provided current pain and fatigue ratings at baseline. These were used to characterise symptom state shortly before VR exposure to support interpretation of task performance, subjective workload and eye tracking signals. For ratings on a 1-10 scale, 1 indicated minimal pain or fatigue and 10 indicated worst pain or fatigue imaginable. Delivered through a Microsoft Forms electronic format using a tablet or laptop.

PROMIS Pain Intensity and Pain Interference

PROMIS Pain Intensity, Short Form 3a:

Completed by participants with chronic pain. The measure contains three items assessing pain intensity: pain at its worst over the previous 7 days, average pain intensity over the previous 7 days, and pain intensity at the time of completion. Each item is scored from 1 to 5, with 1 indicating no pain and 5 indicating very severe pain. Raw summed scores therefore range from 3-15[331].

PROMIS Pain Interference, Short Form 8a:

Completed by participants with chronic pain. The measure contains eight items assessing how much pain interfered with functioning over the previous 7 days. Items assess interference with day-to-day activities, work around the home, participation in social activities, household chores, usual activities undertaken for fun, enjoyment of social activities, enjoyment of life, and family life. Each item is scored from 1 to 5, with response options ranging from not at all to very much. Raw summed scores therefore range from 8-40[331].

PROMIS measures are commonly interpreted using T-scores, where the reference population mean is 50 and the standard deviation is 10. Raw PROMIS scores were calculated by summing item responses. Raw summed scores were converted to PROMIS T-scores using the instrument-specific scoring tables. Higher T-scores indicate greater pain intensity or burden. HealthMeasures interpretive guidance for negatively valenced domains describes scores up to 55 as within normal limits, scores above 55 to 60 as mild, scores above 60 to 70 as moderate, and scores above 70 as severe. These thresholds were used descriptively to aid interpretation of

chronic pain burden, not as primary inferential thresholds[379, 380]. Delivered through a Microsoft Forms electronic format using a tablet or laptop.

Pain Catastrophising Scale (PCS)

Participants with chronic pain completed a paper version of the Pain Catastrophising Scale using the licensed United Kingdom/English version supplied through Mapi Research Trust. The PCS is a 13-item self-report measure of pain-related catastrophic thoughts and feelings, covering rumination, magnification and helplessness. Each item is scored from 0 (not at all) to 4 (all the time), giving a total score range of 0-52, with higher scores indicating greater pain catastrophising[332]. A total score of 30 or above is commonly used to indicate clinically relevant catastrophising[381]. The PCS was used under Mapi Research Trust special terms for the VR-FOCUS doctoral research project; as the scale is copyrighted, the full item wording is not reproduced in this appendix.

In-VR pain and fatigue ratings

Current pain and fatigue was assessed within VR using separate interactive 10-step sliders, completed using the VR system controllers. The scales were anchored from minimal to worst possible (pain or fatigue). Responses were saved directly to headset storage alongside other in-VR questionnaire outputs.

In-VR NASA Task Load Index (NASA-TLX):

Subjective cognitive workload was assessed in VR after each N-back task using the six NASA-TLX subscale items: mental demand, physical demand, temporal demand, perceived performance, effort and frustration. Each item was answered using an interactive slider comprising 20 unlabelled response intervals, corresponding to the standard NASA-TLX 0-100 scoring metric in 5-point increments. For mental demand, physical demand, temporal demand, effort and frustration, higher scores indicated greater perceived demand or burden, with 0 indicating very low and 100 indicating very high. For perceived performance, the scale ranged from perfect performance to failure, so higher scores indicated poorer perceived performance, with 0 indicating perfect performance and 100 indicating failure. No established mild, moderate or severe thresholds were applied; scores were interpreted continuously.

In-VR PAAS mental effort scale:

A single-item PAAS mental effort rating was collected in VR after each N-back task

as an additional subjective measure of cognitive load. Participants rated the mental effort invested in the task using an in-VR interactive slider. The PAAS scale is conventionally scored from 1 to 9, with higher scores indicating greater perceived mental effort.

A Unity slider implementation issue meant that scores could only be recorded using a scale of 0-9. Values of 0 were corrected by recoding 0 to 1 before analysis. This affected 48 of 420 PAAS ratings (11.4%), occurring predominantly in ratings related to the baseline task (44 ratings), with 4 ratings during 1-back and none during 2-back, 3-back or 4-back. Corrected values were similarly distributed across groups, with 26 from the chronic pain cohort and 22 from healthy controls. No values above the minimum score were altered. PAAS scores were analysed as continuous subjective cognitive-load ratings without categorical severity thresholds.

In-VR subjective task-experience items

Additional in-VR ratings assessed immersion, comfort, distraction, anxiety, attention consistency, difficulty maintaining focus, motivation and engagement. Each item was answered using an interactive slider scored from 0 to 100 (0 indicating the lowest rating and 100 indicating the highest rating for all questions). These items were developed to contextualise task-defined cognitive load and to identify experiential factors that could influence task performance or eye tracking signals. No published severity thresholds were applied.

N-back performance metrics

Task performance was logged automatically during each N-back level and scored using a signal detection framework. Trials were classified as hits when participants correctly responded to target stimuli, misses when they failed to respond to target stimuli, false alarms when they responded to non-target stimuli, including lure trials, and correct rejections when they correctly withheld a response to non-target stimuli. Hit rate and false alarm rate were calculated for each participant and N-back level. To avoid infinite values when hit rate or false alarm rate equalled 0 or 1, loglinear corrections were applied, such that corrected hit rate was calculated as (hits+0.5)/(hits+misses+1), and corrected false alarm rate as (false alarms+0.5)/(false alarms+correct rejections+1). These corrected rates were then used to calculate discrimination sensitivity and response bias. Discrimination sensitivity was calculated as $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$, where Z is the

inverse cumulative distribution function of the standard normal distribution. Higher d' values indicated better discrimination between target and non-target stimuli. Response bias was calculated as $c = -0.5 \times [Z(\text{hit rate}) + Z(\text{false alarm rate})]$. Positive c values indicated a more conservative response tendency, reflecting a greater tendency to withhold responses or classify stimuli as non-targets, whereas negative values indicated a more liberal tendency to respond as though stimuli were targets. Raw hit rate and false alarm rate were retained for interpretability, while $d'$and c were used as the principal signal detection indices of N-back task performance, as described in Chapter 7[327].

Eye tracking measures:

Continuous eye tracking data were recorded during the VR N-back task. The primary ocular measures were pupil diameter and blink-related behaviour, with gaze direction also recorded. These signals were used to investigate whether headset-derived measures acquired during VR could support inference of task-defined cognitive load. Eye tracking variables were not interpreted using clinical severity thresholds.

Table A2.1. Example eye tracking CSV output from a single VR N-back task

timestamp	time_ms	pupil_l_mm	pupil_r_mm	gaze_l_X	gaze_l_Y	gaze_l_Z	gaze_r_X	gaze_r_Y	gaze_r_Z
1970-01-01T00:11:59.025956Z	0	4.970632076263430	4.545835494995120	-0.05285393446683880	-0.06474073976278310	0.9965015649795530	-0.10379531979560900	-0.05042309686541560	0.9933197498321530
1970-01-01T00:11:59.042618Z	17	4.972482681274410	4.507942199707030	-0.05293635278940200	-0.06479065120220180	0.9964938759803770	-0.10204433649778400	-0.05563380196690560	0.9932230114936830
1970-01-01T00:11:59.059279Z	33	4.972607135772710	4.538609981536870	-0.05351888388395310	-0.06478196382522580	0.9964632987976070	-0.10235671699047100	-0.05665585398674010	0.9931331276893620

Example of the task-level eye tracking CSV structure generated from the raw headset output. The table shows the first three recorded samples from one task file. timestamp denotes the absolute sample timestamp, time_ms denotes elapsed time from task onset in milliseconds, pupil_l_mm and pupil_r_mm represent left and right pupil diameter in millimetres, and gaze_l_X/Y/Z and gaze_r_X/Y/Z represent left and right eye gaze-direction vectors.

Virtual Reality Sickness Questionnaire (VRSQ):

Post-VR measure of adverse effects and tolerability. The VRSQ was completed after headset removal and assessed VR-related symptoms across oculomotor and disorientation-related symptom domains. The VRSQ uses nine items scored from 0 to 3, corresponding to none, slight, moderate and severe symptom ratings, and can

be summarised as oculomotor, disorientation and total scores scaled to 0-100. No universally accepted mild/moderate/severe cut-offs for the total VRSQ score were applied; higher scores were interpreted as greater VR sickness burden. Delivered through a Microsoft Forms electronic format using a tablet or laptop.

Appendix 3: Additional rsEEG data and figures from PainWaves and VIPA-EEG

This appendix provides supplementary rsEEG data supporting Chapters 5 and 6. Tables and figures are included where they provide additional transparency for quality control, sensitivity analyses, secondary EEG analyses or exploratory outcomes, without duplicating the primary results presented in the main chapters.

Table A3.1. PainWaves quality-control tier composition

Tier	Inclusion criteria	Pairs retained (MZ, DZ)	Median retained EC duration in seconds (IQR; range)	Median channels interpolated (IQR; range)
0	No criteria	36 (17, 19)	192 (153-226; 60-290)	3 (2-5; 0-13)
1	<10 channels removed AND retained duration ≥75s	30 (15,15)	195 (153-235; 88-290)	3 (2-5; 0-8)
2	<8 channels removed AND retained duration ≥120s	22 (12,10)	196 (163-236; 122-290)	3 (2-4; 0-7)
3	<8 channels removed AND retained duration ≥180s	12 (6,6)	227 (206-247; 184-290)	3 (2-4; 0-7)

Retained eyes-closed (EC) duration and number of interpolated channels are summarised at participant level within each retained paired dataset. Values are presented as median (IQR; range). MZ, monozygotic; DZ, dizygotic.

Table A3.2. PainWaves zygosity-stratified spectral results

Zygosity group	n	Frequency band	Cluster direction and topography	Cluster-level p-value	Cohen's dz (95% CI)
MZ	17	Delta	N/A	N/A	-0.19 (-0.80, 0.28)
		Theta	N/A	N/A	-0.16 (-0.71, 0.33)
		Alpha	N/A	N/A	-0.11 (-0.69, 0.36)
		Broad beta	N/A	N/A	-0.61 (-1.3,-0.15)
		Low beta	N/A	N/A	-0.40 (-0.97, 0.06)
		High beta	Negative, frontocentral	<0.05	-0.82 (-1.63, -0.34)
		Gamma	N/A	N/A	-0.51 (-0.97, -0.11)
DZ	19	Delta	N/A	0.09	-0.35 (-0.96, 0.09)
		Theta	N/A	0.05	-0.50 (-1.21, -0.10)
		Alpha	Negative, bilateral frontocentral-central	0.02	-0.56 (-1.21, -0.13)
		Broad beta	Negative, frontocentral-centroparietal	0.01	-0.58 (-1.25, -0.13)
		Low beta	Negative, bilateral frontocentral-central	0.01	-0.68 (-1.36, -0.25)
		High beta	Negative, left frontotemporal-centroparietal	0.04	-0.44 (-1.09, <0.05)
		Gamma	N/A	0.08	-0.42 (-0.99, 0.02)

Sensor-level cluster-based permutation results are shown separately for monozygotic (MZ) and dizygotic (DZ) twin pairs. Cluster direction and topography describe significant sensor-level clusters only; negative clusters indicate lower spectral power in the twin with CWP relative to their co-twin control. ROI-based effect sizes are reported as Cohen's dz with 95% confidence intervals (95% CI). CWP, chronic widespread pain; DZ, dizygotic; MZ, monozygotic.

Table A3.3. PainWaves exploratory connectivity summary

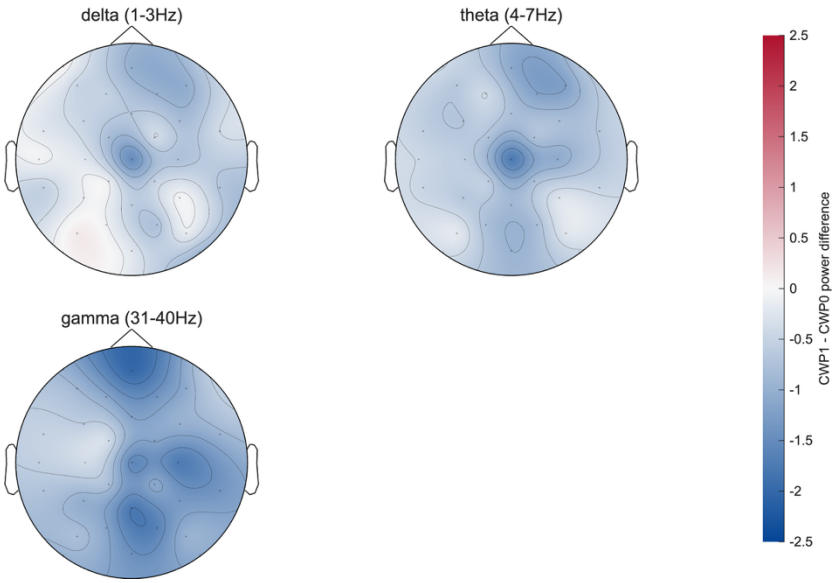
Band	Contrast	Direction of effect	Cohen's dz (95% CI)	Uncorrected p-value
Alpha (8-12Hz)	Central-central connectivity	Lower in CWP	-0.61 (-1.63, -0.05)	0.06
Alpha (8-12Hz)	Frontal-central connectivity	Lower in CWP	-0.58 (-1.28, -0.06)	0.07
Alpha (8-12Hz)	Central-parietal connectivity	Lower in CWP	-0.57 (-1.14, -0.08)	0.08
Broad beta (13-30Hz)	Central-central connectivity	Lower in CWP	-0.79 (-2.17, -0.22)	0.03
Broad beta (13-30Hz)	Frontal-central connectivity	Lower in CWP	-0.51 (-1.58, 0.01)	0.10
Broad beta (13-30Hz)	Central-parietal connectivity	Lower in CWP	-0.52 (-2.27, 0.06)	0.107
Low beta (13-20Hz)	Central-central connectivity	Lower in CWP	-0.66 (-1.67, -0.10)	0.05
Low beta (13-20Hz)	Frontal-central connectivity	Lower in CWP	-0.47 (-1.45, 0.08)	0.13
Low beta (13-20Hz)	Central-parietal connectivity	Lower in CWP	-0.57 (-1.46, -0.04)	0.08

High beta (21-30Hz)	Central-central connectivity	Lower in CWP	-0.77 (-2.01, -0.17)	0.03
High beta (21-30Hz)	Frontal-central connectivity	Lower in CWP	-0.44 (-1.08, 0.06)	0.14
High beta (21-30Hz)	Central-parietal connectivity	Lower in CWP	-0.54 (-1.42, 0.01)	0.09

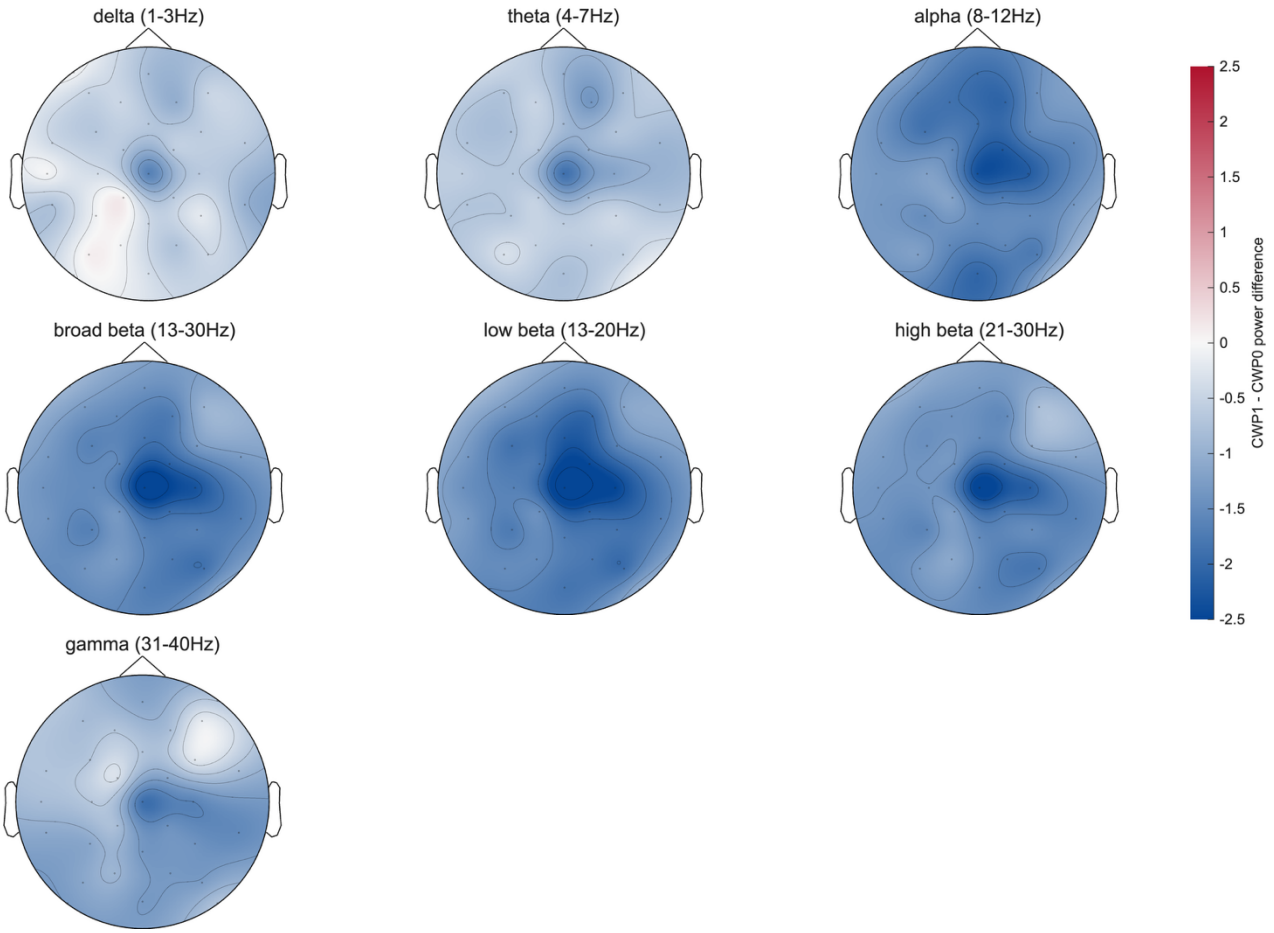
The table summarises selected ROI-level amplitude envelope correlation (AEC) findings from the Tier 3 connectivity dataset. Contrasts are shown for alpha, broad beta, low beta and high beta bands, focusing on central, frontal-central and central-parietal connections because these were the most recurrent nominal patterns in the exploratory analyses. Direction of effect refers to the within-pair contrast between the twin with CWP and the co-twin control; lower in CWP indicates reduced connectivity in the twin with CWP. Effect sizes are reported as Cohen's dz with 95% confidence intervals. P-values are uncorrected and should be interpreted descriptively, as no connectivity result survived correction for multiple comparisons. AEC, amplitude envelope correlation; CWP, chronic widespread pain; ROI, region of interest.

Figure A3.1. PainWaves supplementary spectral topographies across all frequency bands

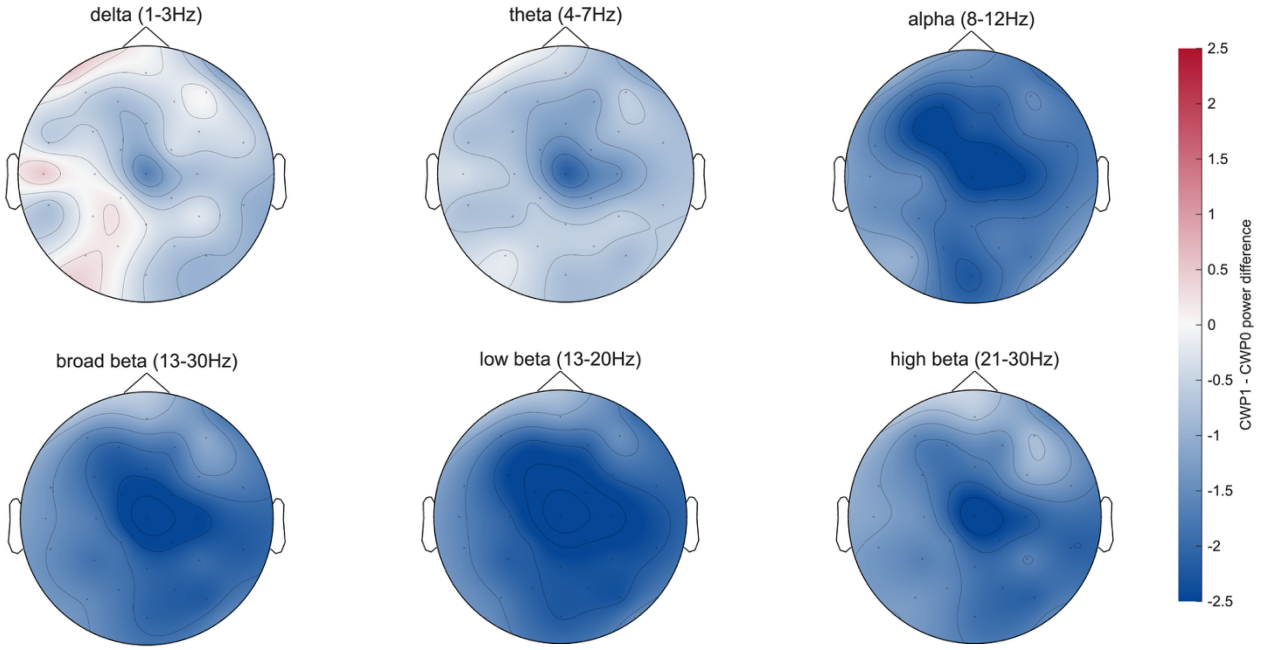
Tier 0



Tier 1


delta (1-3Hz)
theta (4-7Hz)
alpha (8-12Hz)
broad beta (13-30Hz)
low beta (13-20Hz)
high beta (21-30Hz)
gamma (31-40Hz)
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
CWP1 - CWP0 power difference

Tier 2


delta (1-3Hz)
theta (4-7Hz)
alpha (8-12Hz)
broad beta (13-30Hz)
low beta (13-20Hz)
high beta (21-30Hz)
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
CWP1 - CWP0 power difference

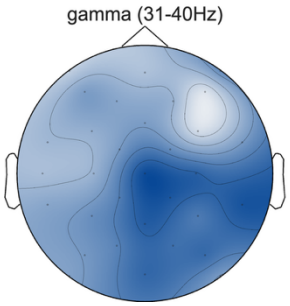


Table A3.6. VIPA-EEG quality-control characteristics by recording state

Tier	Inclusion criteria	Participants retained	Recording state	Median retained EC duration in seconds (IQR; range)	Median channels interpolated (IQR; range)
0	No criteria	23	Pre-intervention	62 (47-74, 30-78)	2 (1-3, 0-10)
			Post-activity 1	58 (50-68, 28-78)	2 (1-4, 0-11)
			Post-activity 2	58 (49-63, 14-84)	2 (1-6, 0-11)
			Post-activity 3	58 (49-69, 12-94)	3 (1-5, 0-9)
			Post-activity 4	54 (46-67, 26-88)	3 (2-4, 0-13)
			Post-intervention	54 (41-64, 22-84)	2 (1-3, 0-7)
1	<8 channels removed AND retained duration ≥30s	21	Pre-intervention	64 (48-74, 40-78)	2 (1-3, 0-7)
			Post-activity 1	58 (50-68, 38-78)	2 (1-3, 0-5)
			Post-activity 2	58 (52-64, 36-84)	2 (1-5, 0-7)
			Post-activity 3	60 (56-70, 42-94)	2 (1-3, 0-7)
			Post-activity 4	54 (46-68, 34-88)	2 (2-4, 0-7)
			Post-intervention	54 (44-68, 36-84)	1 (1-3, 0-5)

Retained eyes-closed (EC) duration and number of interpolated channels are summarised at participant level within each retained paired dataset. Values are presented as median (IQR; range). MZ, monozygotic; DZ, dizygotic.

Appendix 4: Additional performance and self-reported data from VR-FOCUS

This appendix provides additional descriptive and exploratory data supporting the VR-FOCUS results reported in Chapter 7. Tables summarise task-level performance, participant-reported outcomes, tolerability and exploratory clinical-context analyses by group and N-back level where relevant. Values are presented as median (IQR) unless otherwise stated. Between-group comparisons are exploratory and unadjusted unless specified.

Table A4.1. Exploratory between-group comparisons for task performance indices

Measure	N-back level	HC, median (IQR)	CP, median (IQR)	p-value	Rank-biserial r
Hits (out of 10)	1-Back	10.0 (9.0-10.0)	9.0 (7.3-9.8)	<0.05	-0.46
	2-Back	8.0 (7.0-9.0)	6.0 (4.0-7.0	<0.05	-0.42
	3-Back	4.0 (3.0-6.0)	3.0 (2.0-4.0)	0.08	-0.22
	4-Back	4.0 (2.0-4.3)	3.0 (2.0-4.0)	0.18	-0.17
d’	1-Back	3.8 (3.2-3.8)	3.2 (2.6-3.8)	<0.05	-0.46
	2-Back	2.4 (1.9-2.9)	1.6 (1.2-2.2)	<0.05	-0.40
	3-Back	1.2 (1.0-1.7)	1.2 (0.7-1.4)	0.46	-0.09
	4-Back	1.3 (0.9-1.7)	1.0 (0.6-1.4)	0.04	-0.26
c	1-Back	0.2 (0.2-0.5)	0.5 (0.2-0.8)	<0.05	0.46
	2-Back	0.5 (0.3-0.7)	0.7 (0.5-0.9)	<0.05	0.37
	3-Back	0.8 (0.7-1.1)	0.9 (0.8-1.1)	0.06	0.24
	4-Back	1.0 (0.8-1.2)	1.0 (0.8-1.2)	0.74	0.04
Mean hit reaction time (ms)	1-Back	501.1 (476.4-529.5)	537.8 (508.4-579.2)	<0.05	-0.45
	2-Back	539.6 (497.5-580.3)	558.9 (515.8-604.5)	0.12	-0.19

	3-Back	547.9 (499.6-599.6)	588.0 (537.0-640.0)	0.03	-0.28
	4-Back	545.3 (467.4-616.1)	589.4 (529.4-637.7)	0.05	-0.25

Values are presented as median (IQR) for healthy controls (HC) and participants with chronic pain (CP). Between-group p-values were calculated using two-sided Mann-Whitney U tests. Rank-biserial r is reported as the effect size; positive values indicate higher values in CP and negative values indicate higher values in HC. d′ represents discrimination sensitivity and c represents response criterion.

Table A4.2. Participant-reported questionnaire data by group and task

Measure	Group	Baseline	1-Back	2-Back	3-Back	4-Back
Post-task pain VAS, median (IQR)	HC	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
	CP	18.0 (9.3-25.3)	15.0 (1.5-22.3)	13.5 (0.0-29.0)	10.5 (1.0-19.0)	11.0 (2.5-32.5)
Post-task fatigue VAS, median (IQR)	HC	0.0 (0.0-2.3)	0.0 (0.0-9.3)	0.5 (0.0-8.0)	1.5 (0.0-7.8)	1.5 (0.0-10.0)
	CP	26.0 (9.8-63.0)	23.0 (7.8-50.0)	26.0 (10.3-68.0)	29.0 (8.3-67.5)	27.5 (7.0-68.3)
NASA-TLX total score, median (IQR)	HC	17.1 (16.7-18.5)	25.8 (20.0-31.9)	44.2 (38.3-50.8)	52.9 (47.5-57.5)	56.7 (49.0-60.2)
	CP	17.5 (16.7-23.3)	30.8 (23.3-39.6)	50.0 (40.6-59.4)	57.5 (48.3-65.2)	62.5 (52.1-67.9)
PAAS score, median (IQR)	HC	1.0 (1.0-1.0)	2.0 (1.0-5.0)	7.0 (6.0-8.0)	8.0 (7.0-9.0)	8.0 (7.0-9.0)
	CP	1.0 (1.0-1.0)	2.5 (1.0-6.0)	7.0 (6.0-8.0)	8.0 (7.0-9.0)	8.0 (8.0-9.0)
	HC	81.5 (50.0-100.0)	96.0 (76.0-100.0)	98.0 (84.5-100.0)	99.5 (89.8-100.0)	99.5 (86.8-100.0)

Immersion, median (IQR)	CP	66.5 (38.0-89.0)	73.5 (55.8-93.8)	88.5 (76.5-100.0)	93.5 (77.8-100.0)	96.5 (82.8-100.0)
Comfort, median (IQR)	HC	100.0 (89.8-100.0)	100.0 (92.0-100.0)	100.0 (92.8-100.0)	100.0 (99.3-100.0	100.0 (90.0-100.0)
	CP	93.5 (74.3-100.0)	86.5 (72.3-97.3)	83.0 (56.3-94.3)	84.5 (62.5-95.8)	80.0 (35.0-95.0)
Anxiety, median (IQR)	HC	0.0 (0.0-0.3)	5.0 (0.0-13.3)	20.0 (0.0-62.3)	51.0 (0.8-74.5)	50.0 (12.5-67.5)
	CP	2.0 (0.0-21.3)	14.5 (3.0-50.0)	42.5 (18.0-72.3)	66.0 (23.5-80.5)	62.5 (25.0-85.0)
Attention, median (IQR)	HC	82.5 (47.8-92.3)	87.5 (80.3-100.0)	87.0 (67.8-100.0)	83.5 (60.3-100.0)	84.5 (49.3-100.0)
	CP	84.0 (66.0-92.8)	84.0 (64.8-97.3)	82.0 (61.8-91.0)	78.0 (32.5-93.3)	83.5 (61.0-96.0)
Difficulty focusing, median (IQR)	HC	20.0 (4.8-39.5)	11.0 (4.8-24.0)	25.5 (3.5-60.5)	35.5 (2.3-82.0)	31.5 (1.8-73.3)
	CP	13.5 (3.5-31.3)	20.5 (7.5-38.0)	37.0 (18.0-66.3)	67.0 (30.8-85.5)	79.0 (25.0-89.3)
Motivation, median (IQR)	HC	99.5 (74.3-100.0)	100.0 (92.0-100.0)	100.0 (90.5-100.0)	100.0 (91.5-100.0)	85.5 (56.3-100.0)
	CP	97.5 (78.3-100.0)	97.0 (80.0-100.0)	97.5 (88.0-100.0)	98.0 (87.8-100.0)	99.0 (91.5-100.0)
Engagement, median (IQR)	HC	85.5 (56.3-100.0)	100.0 (91.5-100.0)	100.0 (88.3-100.0)	100.0 (86.0-100.0)	100.0 (87.5-100.0)
	CP	84.0 (65.8-100.0)	91.0 (80.8-100.0)	94.5 (84.3-100.0)	91.5 (78.8-100.0)	92.5 (83.3-100.0)

Values are presented as median (IQR) for healthy controls (HC) and participants with chronic pain (CP). Post-task pain and fatigue VAS scores, immersion, comfort, anxiety, attention, difficulty focusing, motivation and engagement were recorded to characterise subjective task experience, tolerability and perceived workload across increasing N-back levels. NASA-TLX, NASA Task Load Index; PAAS, PAAS Cognitive Load Rating Scale; VAS, visual analogue scale. Baseline values are shown where available.

Table A4.3. VRSQ tolerability scores by group

VRSQ measure	Healthy Controls	Chronic Pain	p-value	Rank-biserial r
VRSQ total score, median (IQR)	4.2 (4.2-12.5)	12.5 (6.7-17.3)	0.006	0.33
Oculomotor domain score, median (IQR)	8.3 (8.3-25.0)	25.0 (8.3-25.0)	0.008	0.33
Disorientation domain score, median (IQR)	0.0 (0.0-6.7)	0.0 (0.0-6.7)	0.142	0.16

Values are presented as median (IQR). Between-group p-values were calculated using two-sided Mann-Whitney U tests. Rank-biserial r is reported as the effect size, with positive values indicating higher scores in the chronic pain group. VRSQ, Virtual Reality Sickness Questionnaire.

Because the chronic pain group had substantial multimorbidity and medication exposure, exploratory models examined whether selected clinical factors were associated with pupil diameter or blink rate. These analyses did not identify a consistent pattern suggesting that pain severity, pain interference, pain catastrophising, diagnosis, medication burden, opioid exposure, anticholinergic medication or SSRI/SNRI use dominated the ocular outcomes.

Table A4.4. Exploratory linear regression associations between clinical factors and ocular outcomes in the chronic pain group

Clinical factor	Pupil diameter, p-value	Blink rate, p-value
PROMIS Pain Intensity T-score	0.423	0.403
PROMIS Pain Interference T-score	0.356	0.675
Pain Catastrophizing Scale total score	0.520	0.448
Inflammatory arthritis	0.153	0.689
Fibromyalgia	0.894	0.428
Autoimmune connective tissue disease	0.279	0.479

Specific eye disease mentioned	0.409	0.321
Number of regular medications	0.617	0.187
Current opioid use	0.440	0.815
Opioid use in previous 12 hours	0.244	0.379
Eye-drop use	0.245	0.712
Drugs with anticholinergic effects	0.150	0.396
Drugs that can increase blink rate	0.410	0.940
SSRI/SNRI use	0.375	0.844

P-values are from exploratory linear regression models restricted to participants with chronic pain. Pupil diameter p-values refer to the task period. Analyses were uncorrected and are included as contextual checks rather than primary inferential outcomes. PROMIS, Patient-Reported Outcomes Measurement Information System; SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin-norepinephrine reuptake inhibitor.

Appendix 5: Ethics Approval Forms

Health Research Authority

North West - Liverpool Central Research Ethics Committee
2 Redman Place
Stratford
London
E20 1JQ

09 May 2024

Professor Claire Steves
Clinical Director
King's College London
Department of Twin Research & Genetic Epidemiology
3rd Floor, South Wing, Block D
St. Thomas' Hospital Campus
SE1 7EH

Dear Professor Steves

Study title:	**TwinsUK Programme of Research**
REC reference:	**24/NW/0107**
Protocol number:	**N/A**
IRAS project ID:	**342169**

Thank you for your letter of 24 April 2024, responding to the Research Ethics Committee's (REC) request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation [as revised], subject to the conditions specified below.

Good practice principles and responsibilities

The UK Policy Framework for Health and Social Care Research sets out principles of good practice in the management and conduct of health and social care research. It also outlines the responsibilities of individuals and organisations, including those related to the four elements of

research transparency:

1. registering research studies
2. reporting results
3. informing participants
4. sharing study data and tissue

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

All research should be registered in a publicly accessible database and we expect all researchers, research sponsors and others to meet this fundamental best practice standard.

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a public registry before the first participant is recruited and no later than six weeks after. For this purpose, 'clinical trials' are defined as:

- clinical trial of an investigational medicinal product
- clinical investigation or other study of a medical device
- combined trial of an investigational medicinal product and an investigational medical device
- other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice.

A 'public registry' means any registry on the WHO list of primary registries or the ICMJE list of registries provided the registry facilitates public access to information about the UK trial.

Failure to register a clinical trial is a breach of these approval conditions, unless a deferral has been agreed by the HRA (for more information on registration and requesting a deferral see: Research registration and research project identifiers).

Where a deferral is agreed we expect the sponsor to publish a minimal record on a publicly accessible registry. When the deferral period ends, the sponsor should publish the full record on the same registry, to fulfil the condition of the REC favourable opinion.

If you have not already included registration details in your IRAS application form you should notify the REC of the registration details as soon as possible.

Where the study is registered on ClinicalTrials.gov, please inform deferrals@hra.nhs.uk and the Research Ethics Committee (REC) which issued the final ethical opinion so that our records can be updated.

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter. Where a deferral is agreed, a minimum research summary will still be published in the research summaries database. At the end of the deferral period, we will publish the full research summary.

Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit: Research summaries - Health Research Authority (hra.nhs.uk)

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report
- Reporting results

The latest guidance on these topics can be found at Managing your approval - Health Research Authority (hra.nhs.uk)

Ethical review of research sites

NHS/HSC sites

The favourable opinion applies to all NHS/HSC sites taking part in the study, subject to confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or management permission (in Scotland) being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS/HSC sites

I am pleased to confirm that the favourable opinion applies to any non-NHS/HSC sites listed in the application, subject to site management permission being obtained prior to the start of the study at the site.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Covering letter on headed paper [Cover Letter]		15 March 2024
Covering letter on headed paper [Cover Letter]	2	24 April 2024
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Sponsor Insurance]		26 July 2023
Other [Data Linkage Information]	v0.1	13 March 2024
Other [MPE Report for DXA]	N/A	13 March 2024
Other [UKLLC Fair Processing Expansion]	N/A	13 March 2024
Other [Substudy1 Protocol]	v1.2	13 March 2024
Other [Substudy1 VIS]	v1.1	13 March 2024
Other [Substudy2 Protocol]	v1.1	13 March 2024
Other [Substudy2 VIS]	v1.1	13 March 2024
Other [Substudy3 Protocol]	v0.1	13 March 2024
Other [Substudy3 VIS]	v0.1	13 March 2024
Other [Substudy4 Protocol]	v0.1	13 March 2024
Other [Substudy4 VIS]	v0.1	13 March 2024
Other [Substudy5 Protocol]	v0.1	13 March 2024
Other [Substudy6 Protocol]	v0.1	13 March 2024
Other [Substudy7 Protocol]	v0.1	13 March 2024
Other [Substudy7 VIS]	v0.1	13 March 2024
Other [Substudy8 Protocol]	v0.1	13 March 2024
Other [Substudy5 VIS]	1.0	18 April 2024
Other [Substudy6 VIS]	1.0	18 April 2024
Other [Substudy8 VIS]	1.0	18 April 2024
Other [Substudy 6 Twinclusion Microbiome Questionnaire]	4.0	18 April 2024
Other [Substudy 6 Twinclusion Oral Inflammatory Load Questionnaire]	3.0	18 April 2024
Other [Substudy 7 Ageing Immune ChrX Questionnaire]	0.1	23 April 2024
Other [TwinsUK Study Recruitment Email]	0.1	18 April 2024
Participant consent form [Young People Assent 6-15 Yrs]	v0.1	13 March 2024
Participant consent form [Parents Consent]	1.0	18 April 2024
Participant consent form [Adult Consent]	1.0	18 April 2024
Participant information sheet (PIS) [10-15 Yrs VIS]	v0.1	13 March 2024
Participant information sheet (PIS) [6-9 Yrs VIS]	v0.1	13 March 2024
Participant information sheet (PIS) [Under 6 Yrs VIS]	v0.1	13 March 2024
Participant information sheet (PIS) [Adult VIS]	1.0	18 April 2024
Participant information sheet (PIS) [Parents VIS]	1.0	18 April 2024
REC Application Form [REC_Form_21032024]		21 March 2024

Research protocol or project proposal	v0.1	13 March 2024
Research protocol or project proposal	1.0	18 April 2024
Summary CV for Chief Investigator (CI) [CI's CV]		13 March 2024

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: Quality assurance - Health Research Authority (hra.nhs.uk)

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: Learning - Health Research Authority (hra.nhs.uk)

IRAS project ID: 342169 Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely

Tina Cavaliere

On behalf of

Mr Paul Mooney
Chair

Email: liverpoolcentral.rec@hra.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

After ethical review guidance for sponsors and investigators – Standard Conditions for Programmes of Research

Copy to: Professor Bashir Al-Hashimi

University of East Anglia
Norwich Research Park
Norwich. NR4 7TJ

Email: ethicsmonitor@uea.ac.uk
Web: www.uea.ac.uk

Study title: VR-FOCUS: Leveraging sensor-based technologies and machine learning for real-time optimisation of cognitive load within virtual reality.

Application ID: ETH2324-2498

Dear Jordan,

Your application was considered on 23rd August 2024 by the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee).

The decision is: **approved**.

You are therefore able to start your project subject to any other necessary approvals being given.

If your study involves NHS staff and facilities, you will require Health Research Authority (HRA) governance approval before you can start this project (even though you did not require NHS-REC ethics approval). Please consult the HRA webpage about the application required, which is submitted through the IRAS system.

This approval will expire on **1st July 2025**.

Please note that your project is granted ethics approval only for the length of time identified above. Any extension to a project must obtain ethics approval by the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) before continuing.

It is a requirement of this ethics approval that you should report any adverse events which occur during your project to the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) as soon as possible. An adverse event is one which was not anticipated in the research design, and which could potentially cause risk or harm to the participants or the researcher, or which reveals potential risks in the treatment under evaluation. For research involving animals, it may be the unintended death of an animal after trapping or carrying out a procedure.

Any amendments to your submitted project in terms of design, sample, data collection, focus etc. should be notified to the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) in advance to ensure ethical compliance. If the amendments are substantial a new application may be required.

Approval by the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) should not be taken as evidence that your study is compliant with the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018. If you need guidance on how to make your study UK GDPR compliant, please contact the UEA Data Protection Officer (dataprotection@uea.ac.uk).

Please can you send your report once your project is completed to the FMH S-REC (fmh.ethics@uea.ac.uk).

I would like to wish you every success with your project.

On behalf of the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee)

Yours sincerely,

Dr Paul Linsley

University of East Anglia
Norwich Research Park
Norwich. NR4 7TJ

Email: ethicsmonitor@uea.ac.uk
Web: www.uea.ac.uk

Study title: PainWaves: An Investigation of EEG Biomarkers in Chronic Pain to Inform Future Targeted Treatments

Application ID: ETH2425-2568

Dear Jordan,

Your application was considered on 12th September 2025 by the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee).

The decision is: **approved**.

You are therefore able to start your project subject to any other necessary approvals being given.

This approval will expire on **30th September 2030**.

Please note that your project is granted ethics approval only for the length of time identified above. Any extension to a project must obtain ethics approval by the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) before continuing.

It is a requirement of this ethics approval that you should report any adverse events which occur during your project to the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) as soon as possible. An adverse event is one which was not anticipated in the research design, and which could potentially cause risk or harm to the participants or the researcher, or which reveals potential risks in the treatment under evaluation. For research involving animals, it may be the unintended death of an animal after trapping or carrying out a procedure.

Any amendments to your submitted project in terms of design, sample, data collection, focus etc. should be notified to the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) in advance to ensure ethical compliance. If the amendments are substantial a new application may be required.

Approval by the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee) should not be taken as evidence that your study is compliant with the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018. If you need guidance on how to make your study UK GDPR compliant, please contact the UEA Data Protection Officer (dataprotection@uea.ac.uk).

Please can you send your report once your project is completed to the FMH S-REC (fmh.ethics@uea.ac.uk).

I would like to wish you every success with your project.

On behalf of the FMH S-REC (Faculty of Medicine and Health Sciences Research Ethics Subcommittee)

Yours sincerely,

Dr Paul Linsley

NHS
Health Research Authority
South West - Cornwall & Plymouth Research Ethics Committee
Level 3
Block B
Whitefriars
Lewins Mead
Bristol
BS1 2NT

Telephone: 02071048033/53

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

01 June 2020

Dr Jordan Tsigarides
NIHR Academic Clinical Fellow (Rheumatology)
Norfolk and Norwich University Hospital
Norfolk and Norwich University Hospital
Colney Lane, Norwich
Norfolk
NR47UY

Dear Dr Tsigarides

Study title:	**Investigating Virtual Immersive Experiences in the Management of Chronic Widespread Pain – The VIPA Study**
REC reference:	**20/SW/0050**
Protocol number:	**2.0**
IRAS project ID:	**271209**

Thank you for your responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a publicly accessible database. For this purpose, 'clinical trials' are defined as the first four project categories in IRAS project filter question 2. Registration is a legal requirement for clinical trials of investigational medicinal products (CTIMPs), except for phase I trials in healthy volunteers (these must still register as a condition of the REC favourable opinion).

Registration should take place as early as possible and within six weeks of recruiting the first research participant at the latest. Failure to register is a breach of these approval conditions, unless a deferral has been agreed by or on behalf of the Research Ethics Committee (see here for more information on requesting a deferral:
https://www.hra.nhs.uk/planning-and-improving-research/research-planning/research-registration-research-project-identifiers/

As set out in the UK Policy Framework, research sponsors are responsible for making information about research publicly available before it starts e.g. by registering the research project on a publicly accessible register. Further guidance on registration is available at:
https://www.hra.nhs.uk/planning-and-improving-research/research-planning/transparency-responsibilities/

You should notify the REC of the registration details. We will audit these as part of the annual progress reporting process.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report

The latest guidance on these topics can be found at https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

NHS/HSC sites

The favourable opinion applies to all NHS/HSC sites listed in the application subject to confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or management permission (in Scotland) being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS/HSC sites

I am pleased to confirm that the favourable opinion applies to any non-NHS/HSC sites listed in the application, subject to site management permission being obtained prior to the start of the study at the site.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Professional Indemnity Letter]	1	11 February 2020
GP/consultant information sheets or letters [The VIPA Study - Appendix 19 - Letter to GP (Ver 1.0)]	2	15 April 2020
IRAS Application Form [IRAS_Form_07052020]		07 May 2020
Letter from funder [Letters from Funders]	1	29 November 2019
Letter from sponsor [Letter from Sponsor]	1	11 February 2020
Letters of invitation to participant [The VIPA Study - Appendix 3 - Invitation Letter (Ver 2.0) - Updated for REC]	2.0	15 April 2020
Non-validated questionnaire [The VIPA Study - Appendix 6 - Baseline Questionnaire (Ver 2.0) - Updated for REC]	2.0	15 April 2020
Non-validated questionnaire [The VIPA Study - Appendix 13 - Experiment 1 Survey (Ver 1.0)]	1	15 April 2020
Non-validated questionnaire [The VIPA Study - Appendix 15 - Experiment 2 Survey (Ver 1.0)]	1	15 April 2020
Non-validated questionnaire [The VIPA Study - Appendix 16 - Experiment 3 Survey (Ver 1.0)]	1	15 April 2020
Non-validated questionnaire [The VIPA Study - Appendix 17 - Experiment 4 Survey (Ver 1.0)]	1	15 April 2020
Other [GANTT Chart]	2.0	15 April 2020
Other [Information Document for Participants - EEG]	1	15 April 2020

Other [The VIPA Study - Amendments following REC meeting (Ver 1.0)]	1.0	15 April 2020
Participant consent form [The VIPA Study - Appendix 4 - Consent Form (Ver 2.0) - Updated for REC]	2.0	15 April 2020
Participant information sheet (PIS) [The VIPA Study - Appendix 5 - Participant Information Sheet (Ver 2.0) - Updated for REC]	2.0	15 April 2020
Research protocol or project proposal [The VIPA Study - Research Protocol]	2.0	15 April 2020
Summary CV for Chief Investigator (CI) [J Tsigarides CV]	1	31 January 2020
Summary, synopsis or diagram (flowchart) of protocol in non technical language [The VIPA Study - Appendix 1 - Flow Chart (Ver 2.0) - Updated for REC]	2.0	15 April 2020
Validated questionnaire [The VIPA Study - Appendix 8 - Widespread Pain Index (WPI) (Ver 1.0)]	1	15 April 2020
Validated questionnaire [The VIPA Study - Appendix 9 - Revised Fibromyalgia Impact Questionnaire (Ver 1.0)]	1	15 April 2020
Validated questionnaire [The VIPA Study - Appendix 10 - McGill Pain Questionnaire (Ver 1.0)]	1	15 April 2020
Validated questionnaire [The VIPA Study - Appendix 11 - Patient Health Questionnaire 9 (PHQ-9) (Ver 1.0)]	1	15 April 2020
Validated questionnaire [The VIPA Study - Appendix 12 - Generalised Anxiety Disorder 7 (GAD-7) (Ver 1.0)]	1	15 April 2020
Validated questionnaire [The VIPA Study - Appendix 14 - Pain Visual Analogue Scales (Ver 1.0)]	1	15 April 2020

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at:
https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS project ID: 271209	Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely



pp Dr Stephen Coles
Chair

Email:cornwallandplymouth.rec@hra.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

Copy to: Mr Graham Horne

Appendix 6: Consent Forms

University of East Anglia

Norfolk and Norwich University Hospitals NHS Foundation Trust

Addenbrooke's NHS Trust

The VIPA Study

STUDY CONSENT FORM

Study Title: Investigating Virtual Immersive Experiences in the Management of Chronic Pain – The VIPA Study

IRAS Project ID: 271209

Responsible Researcher: Dr Jordan Tsigarides (Principal Investigator)

Additional Researchers: Miss Vanessa Grove (Co-investigator), Dr Susan Miles (Co-investigator), Dr Toby Smith (Co-investigator), Prof Iain McNamara (Co-investigator), Dr Nicholas Shenker (Co-investigator), Dr Saber Sami (Research Supervisor), Prof Alexander Macgregor (Research Supervisor)

Please read statements 1-10 below and initial in the boxes to confirm that you agree with the corresponding statement. On the back of this page, please print your name/sign/date where indicated to confirm that you agree with statements 1-10. To be considered for participation, you must agree to all of statements 1-10.

At the end of this consent form please read the additional statements 11-12 and initial in the boxes if you agree. It is not mandatory that you agree with these statements to be considered for participation.

Statements 1-10

		Please initial boxes
1.	I confirm that I have read and understand the Participant Information Sheet dated February 2020 (version 1.0) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2.	I confirm that I have been fully informed of the nature and extent of any risk to my health or well-being from participation in this study	
3.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.	
4.	I understand that The VIPA Study research team will hold personal details collected during the study for the purposes of carrying out the research project as detailed in the Participant Information Sheet. If I withdraw from the study after my data is analysed, I understand that my rights to access, change or move my information are limited. The VIPA Study researchers need to manage my information in specific ways in order for the research to be reliable and accurate. Prior to data analysis, unless I specifically request for my data to be removed, data collected up until the point of withdrawal will remain part of the study.	
5.	I understand that the data from this research will be stored for a period of 10 years as per current University of East Anglia (UEA) data protection protocol and Medical Research Council (MRC) guidelines.	
6.	I understand that personal information collected for the study will be kept on a secure database as described in the Participant Information Sheet. Authorised members of The VIPA Study research team will access personal information only where necessary and will not share any personal details with anyone	

	outside of the study. With the exception of a reasonable request from a regulatory authority.	
7.	I agree to take part in the above study.	
8.	I agree to participate in at least one of the experiments outlined in the Participant Information Sheet.	
9.	I agree to my GP being informed of my participation in the study.	
10.	I understand that after I sign and return this consent form, it will be retained by The VIPA Study research team. If I am eligible for participation, one of the research team will check this form with me prior to participation in any of the experiments.	

Participant Name in Block Capitals	**Date**	**Signature**

Name of Research Team Member in Block Capitals	**Date**	**Signature**

Original to be retained by the research team

Copy to be given to the participant

Additional Statements 11-12

11	I agree that my anonymised data can be shared with research collaborators undertaking other ethically approved research both in the UK and internationally.	
12	I agree that the research team may use my details to contact me about other research projects that may be relevant to me.	

The VR-FOCUS Study: Participant Information, Eligibility and Consent

* Required

Study Consent Form

Study Title: VR-FOCUS: Leveraging sensor-based technologies and machine learning for real-time optimisation of cognitive load within virtual reality pain management interventions

Responsible Researcher: Dr Jordan Tsigarides (Principal Investigator)

Additional Researchers: Prof Anthony Bagnall (Co-investigator), Dr Jennifer Bowler (Secondary Research Supervisor), Prof Alexander Macgregor (Primary Research Supervisor).

Please read the following statements and select "yes" to confirm your agreement. To be considered for participation, you must agree to all of the following statements and print your name at the end.

I confirm that I have read and understand the Participant Information Sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. *

- ◯ Yes

I confirm that I have been fully informed of the nature and extent of any risk to my health or well-being from participation in this study. *

- ◯ Yes

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care, educational provision or legal rights being affected. *

- ◯ Yes

I understand that The VR-FOCUS research team will hold personal details collected during the study for the purposes of carrying out the research project as detailed in the Participant Information Sheet. If I withdraw from the study after my data is analysed, I understand that my rights to access, change or move my information are limited. The VR-FOCUS research team need to manage my information in specific ways in order for the research to be reliable and accurate. Prior to data analysis, unless I specifically request for my data to be removed, data collected up until the point of withdrawal will remain part of the study. *

- ◯ Yes

I understand that the data from this research will be stored for a period of 10 years as per current University of East Anglia (UEA) data protection protocol and Medical Research Council (MRC) guidelines. *

- ◯ Yes

I understand that personal information collected for the study will be kept on a secure database as described in the Participant Information Sheet. Authorised members of The VR-FOCUS research team will access personal information only where necessary and will not share any personal details with anyone outside of the study, with the exception of a reasonable request from a regulatory authority. *

- ◯ Yes

I agree to take part in the VR-FOCUS study. *

- ◯ Yes

I agree that my anonymised data can be shared with research collaborators undertaking other ethically approved research both in the UK and internationally. *

- ◯ Yes

I agree that the research team may use my details to contact me about other research projects that may be relevant to me. *

- ◯ Yes

I understand that upon completion of this consent form, a record of my consent will be retained by the VR-FOCUS research team and, prior to participation in data collection sessions, one of the research team will re-confirm my consent to participate with me. *

- ◯ Yes

Please print your name below to confirm your consent to participation in the VR-FOCUS study. *

Enter your answer

Back | Next

Page 3 of 4

Appendix 7: Participant Information Sheets

Participant Information Sheet







Study Title: Investigating Virtual Immersive Experiences in the Management of Chronic Widespread Pain – The VIPA Study

IRAS Project ID: 271209

Date & Version: 1st October 2020, Version 2.0

Principal Investigator: Dr Jordan Tsigarides

Dear Sir/Madam,

We are inviting you to take part in a research study. Before you decide if you want to be involved, it is important that you understand why we are conducting the research and what your participation will involve.

Please read the following information carefully and take time to consider if you would like to take part. After reading this information sheet, any questions about the study can be directed to the main researcher (contact details below).

The project will be supervised by Professor Alexander Macgregor, Consultant Rheumatologist and lead researcher at the University of East Anglia Centre for Epidemiology Versus Arthritis, Norwich.

Why is this research being done?

The Burden of Chronic Pain in the UK

Chronic pain affects up to 50% of people in the UK and is a major cause of suffering and disability worldwide. In the future, with the increasing age of the population, it is likely that this percentage will continue to rise. At the moment, chronic pain is mostly managed with medications, with morphine-like drugs being commonly prescribed.

Unfortunately, these medications can cause a range of side effects that can worsen the suffering of people with an already affected quality of life. Additionally, in chronic widespread pain conditions such as Fibromyalgia Syndrome (FMS), evidence shows that these medications are not very effective. Current guidelines from the European League Against Rheumatism (EULAR) recommend other 'non-drug' treatments such as exercise and psychological therapies. However, the availability of these treatments to the people that need them is varied throughout the UK. It is important that new, innovative and accessible therapies are developed to help to treat people suffering from chronic pain.

Virtual Reality – A Potential New Treatment

Virtual reality (VR) is an innovative form of technology in which a person becomes fully immersed within a 360-degree, computer generated, 3D, 'virtual' environment. The visual element of a modern VR experience is achieved through the use of a head-mounted display (HMD) with built in motion sensors that can detect different movements of the head (see figures 1 & 2 below). When the head is moved in a VR experience, the virtual environment adapts to this movement, allowing the user to see a different aspect of the scene. The use of headphones attached to the HMD allows the user to experience the scene through both audio and visual features. Finally, depending on the VR experience, users can interact with the virtual environment in different ways, most commonly achieved with the use of hand controllers. With these three elements (visual immersion, auditory immersion and interactivity), the user can have a feeling of presence within the virtual world.

The unique multi-sensory experience offered by VR has been shown to reduce unpleasant symptoms, including pain and anxiety. We therefore believe that VR therapy has exciting potential as a new treatment to benefit both the physical and psychological wellbeing of patients with Fibromyalgia Syndrome. Multiple studies have shown that use of VR can significantly reduce the amount of pain experienced in people with acute (short term) painful conditions such as burns. However, we know that chronic (longer term) pain is different from acute pain. Early research looking at using VR for chronic pain is encouraging, although only a small number of research studies have been reported so far. At this early stage in the study of this technology, it is important that we find the best way of using this potential future therapy when using it for people with Fibromyalgia Syndrome specifically. To determine this, we aim to

conduct multiple small experiments to understand more about which VR technologies, environments and activities show the most promising influence on pain and are most acceptable to people suffering from Fibromyalgia Syndrome. It is important that we study both how the VR therapy influences symptoms and also whether people with Fibromyalgia Syndrome find this technology 'acceptable' (can tolerate and use the technology).

Why have I been invited to take part?

You have been invited to join the study because you have a diagnosis of Fibromyalgia Syndrome made by a Rheumatologist and are currently suffering from chronic pain which is thought to be caused by this condition.

Do I have to take part?

No, choice to participate in the study is entirely voluntary and you can freely choose not to take part in the study. Even if you are included as a participant in this study, you may withdraw at any time without needing to provide a reason. You may also decide to refuse to answer any question(s) involved in the study if you feel uncomfortable with answering them.

Who are you looking for?

We are looking for volunteers who have a diagnosis of Fibromyalgia Syndrome above 18 years of age, have time to take part, and have a good understanding of English. This is to ensure that participants can fully understand the study proceedings and what is expected of them before consenting to take part. To take part, you must have been suffering from persistent or recurrent pain for the last 3 months or more.

You will <u>not</u> be able to take part in the study if you have:

- Any condition in which the symptoms can be brought on by flashing lights or screens
- A diagnosis of cognitive impairment or dementia
- A seeing or hearing impairment that would prevent use of the VR included in the experiments

If I don't take part, will it affect my treatment?

No, if you decide not to be involved in this study, your treatment will be in no way affected. Similarly, if you decide to take part at first, but decide to withdraw from the study, your treatment will be in no way affected. Withdrawing from the study will also

have no effect on your legal rights, employment or participation in other research projects.

If I do take part, will you inform my general practitioner (GP)?

If after the screening process you are selected to take part in this study, we will inform your GP. Your GP will be sent a letter informing them of your participation and summarising the research project, including a short explanation of your involvement. This is so that they are aware of your participation in the unlikely event that you need to seek their medical input because of any 'side effects' from being involved.

What will I have to do if I agree to take part?

This study is going to be split into four individual experiments, each with different elements to them. On the reply slip located at the bottom of the invitation letter you will be asked to indicate which of the following experiments you would like to take part in. You may choose to participate in any of the four experiments listed below. The number of the below experiments that you choose to participate in is also your choice (you can participate in all four of the experiments if you wish to). For each of the experiments, you will be asked to attend a face-to-face session with a member of the research team at the University of East Anglia (UEA).

Prior to starting the first experiment that you attend, you will be asked to read, complete and sign a consent form to indicate that you consent to participation in this research study. This will be witnessed and countersigned by a member of the research team. At the beginning of your first experiment, you also be asked to complete some initial questionnaires asking about your condition, wellbeing and lifestyle. Before starting to use the equipment in each of the experiments that you attend, you will be given a safety briefing, and introduction to each of the pieces of technology that you will be using. A research team member will be present throughout each experiment and will answer any questions that you have. You will also have access to this information sheet.

Here are the outlines for each experiment individually:

Experiment 1 – Different VR Technologies

What do you aim to achieve with Experiment 1?

- This experiment aims to compare four different types of commercially available VR technology and determine which of these that people with Fibromyalgia Syndrome prefer and tolerate best.

What does Experiment 1 involve me doing?

- During this experiment, you will be asked to use each piece of VR technology to complete a short activity. After completing the activity using one of the pieces of VR technology, you will be given some time to reflect on your experience and will be asked to complete a questionnaire describing your thoughts/feelings. You will then be asked to complete the same activity using a different piece of VR technology. After using all four of the pieces of VR technology, you will be asked to complete a final questionnaire in which we will ask you specific questions to compare your experience with each piece of technology. Completion of this questionnaire will signal the end of this experiment. Your visit will take about 2-3 hours from arrival to this end point.

Experiment 2 – Different Virtual Environments

What do you aim to achieve with Experiment 2?

- This experiment aims to compare which of our two different virtual environments are most preferred and tolerated in people with Fibromyalgia Syndrome.
- Also, we aim to compare how two different environments can affect the pain levels and brainwave patterns in people with Fibromyalgia Syndrome. This experiment will use electroencephalography (EEG) to record the brainwave patterns before, during and after using VR (there is more information about the EEG technology further below).

What does Experiment 2 involve me doing?

- Before using the technology, you will be asked to rate your level of pain on a visual analogue scale (VAS). The EEG equipment will then be placed on your head. Following this, we will ask you to again rate your level of pain so that this can be matched with the signals recorded by the EEG. At this stage, we will record your brainwaves at rest for a short period of time. There may be some adjustment of the EEG equipment during this time to ensure your brainwaves are being recorded clearly. You will then be asked to wear the VR equipment and use it to complete a short activity using the first virtual environment. After this, you will be asked to rate your level of pain within the VR technology and answer some short questions about your experience. You will then be asked to complete the same activity using the second virtual environment. Following this, you will again be asked to rate your level of pain within the VR technology and answer the same short questions about your experience. You will then be asked to remove the VR technology and we will record your brainwaves at rest for a short period of time. Finally, a member of the research team will assist you with removing the EEG equipment. You will then be asked rate your pain for a final time and

complete a questionnaire, asking you specific questions on your experience/feelings of each of the virtual environments. Completion of this questionnaire will signal the end of this experiment. Your visit will take about 2-3 hours from arrival to this end point.

Experiment 3 – Different Virtual Activities

What do you aim to achieve with Experiment 3?

- This experiment aims to compare which of our four different VR activities are most preferred and tolerated in people with Fibromyalgia Syndrome.
- Also, we aim to compare how the different activities can affect the pain levels and brainwave patterns in people with Fibromyalgia Syndrome. This experiment will use electroencephalography (EEG) to record the brainwave patterns before, during and after using VR.

What does Experiment 3 involve me doing?

Before using the technology, you will be asked to rate your level of pain on a visual analogue scale (VAS). The EEG equipment will then be applied to your head. Following this, we will ask you to again rate your level of pain so that this can be matched to the signals recorded by the EEG. At this stage, we will record your brainwaves at rest for a short period of time. There may be some adjustment of the EEG equipment during this time to ensure your brainwaves are being recorded clearly. You will then be asked to use the VR equipment to complete each of the four short activities. After each activity, you will be asked to rate your level of pain within the VR technology and will be asked some short questions about your experience. After completing the final VR activity you will be asked to remove the VR technology and we will record your brainwaves at rest for a short period of time. Finally, a member of the research team will assist you with removing the EEG equipment. You will then be asked rate your pain for a final time and complete a questionnaire, asking you specific questions on your experience/feelings of each of the VR activities. Completion of this questionnaire will signal the end of this experiment. Your visit will take about 2-3 hours from arrival to this end point.

Experiment 4 – The Brain-Computer Interface

What do you aim to achieve with Experiment 4?

- This experiment aims to see if using a brain-computer interface (BCI) with VR is acceptable (tolerated) in people with Fibromyalgia Syndrome.
- Also, we aim to look at how BCI affects the pain levels and brainwaves in people with Fibromyalgia Syndrome. This experiment will use electroencephalography (EEG) to record the brainwave patterns before, during and after using VR.

What does Experiment 4 involve me doing?

- Before using the technology, you will be asked to rate your level of pain on a visual analogue scale (VAS). The EEG system will then be applied to your head. Following this, we will ask you to again rate your level of pain. At this stage, we will record your brainwaves at rest for a short period of time. There may be some adjustment of the EEG equipment during this time to ensure your brainwaves are being recorded clearly. You will then be asked to use the VR/EEG equipment to perform a short 'training' session. This will involve using your brainwaves to control the VR environment. When you and the research team are satisfied that you are able to control the environment effectively, you will be asked to rate your pain level. Following this, you will be asked to complete a short activity using the VR/EEG equipment to control elements of the environment with your brainwaves. After this activity you will be asked to rate your pain levels again. You will then be asked to remove the VR technolog and we will record your brainwaves at rest for a short period of time. Finally, a member of the research team will assist you with removing the EEG equipment. You will then be asked rate your pain for a final time and complete a questionnaire, asking you specific questions on your experience/feelings of using the VR BCI. Completion of this questionnaire will signal the end of this experiment. Your visit will take about 2-3 hours from arrival to this end point.

After you have selected the experiment(s) that you would like to participate in (on the return slip at the bottom of the invitation letter) and returned this to us, the research team will contact you using the contact information that you provide and offer you dates/times of available sessions. You will then be asked to select a date/time which is most convenient for you.

Why are you using EEG in your experiments?

Using EEG allows us to look at the electrical activity across the brain, which can help us to understand more about certain medical conditions or symptoms and how our brains process information. Similarly, looking at how these brainwave patterns change with different conditions (such as when using VR) can help us to understand how our therapies are working. Gaining information on the brainwave patterns of lots of people with Fibromyalgia Syndrome allows us to look for similarities that may be due to the

condition. We can also potentially look for similarities in other data sets that have been collected by other researchers, allowing us to build up a picture of how this condition affects the brain and how VR can be used to help with symptoms.

I am interested in participating in one of the experiments with the EEG, what is the procedure of using this?

Throughout the experiment, one of our research team will be there to assist you with putting on the EEG equipment and applying the electrodes to your head to measure your brainwaves.

Firstly you will be asked to put an EEG cap on your head. The cap will have holes in it to allow for the EEG sensors (also called “electrodes”) to be applied to the head. After parting the hair over the location of each of these holes, the research team will use a small cotton swab to clean the skin around where the electrodes will be placed. The electrodes will then be applied to the scalp with use of a small amount of gel to allow for a good signal. After each electrode has been secured, the research team will confirm that the computer is receiving a clear signal from the EEG equipment. There may be some readjustment or re-application of electrodes required to ensure this.

You are using a “brain-computer interface” (BCI) in Experiment 4, could you tell me more about this?

Brain-computer interfaces (BCI) are systems that enable communication between the brain and technological devices like computers. Firstly, electrical signals are collected from the brain with EEG using the procedure described above. These signals are sent to a high-powered computer which translates them into commands to control the computer program/device. By encouraging the user to think about performing a certain task lots of times, the user can start to control elements of the computer program with just the use of their ‘brainwaves’. This may be controlling the cursor of a computer mouse, or typing on a virtual keyboard with the use of brainwaves alone. Within the experiment, the EEG will be linked to a VR program and, through use of brainwaves alone, our participants will use this to control elements of the VR environment.

What are the possible benefits of taking part?

There have only been a small number of research studies that have looked at using VR as a therapy for chronic pain. Some of the participants in these studies found that their pain levels were reduced whilst using VR and for a short time afterwards.

However, these studies were small and included people with a range of different chronic pain conditions. Although research looks promising and you may experience short-term reductions in your pain when using VR, it is important to know that we are still experimenting with the use of VR technologies as treatments for chronic pain and cannot guarantee any changes in your symptom levels.

Brain-computer interfaces (BCI) are relatively new and, although they have been studied before in rehabilitation, this will be one of the first studies to use BCI in people with chronic pain. Although we think that using BCI with VR could help with the symptoms in Fibromyalgia Syndrome, we are still experimenting with this technology and cannot guarantee any changes in your symptom levels.

Ultimately, your participation will help us to target the future development of our VR programs. This may help us to produce a more effective and acceptable intervention in the future. The results from your participation in this study will help to inform the scientific community on these types of interventions. The results will also help to inform further research trials. In the future, by adding to the current research base, we can move closer to developing a VR therapy that could be used in the management of Fibromyalgia Syndrome.

What are the possible risks and discomforts of taking part in this research?

All of the VR technologies used in this study are commercially available and used regularly by millions of people all over the world. The risks of participating in this research study are minimal, and similar studies have shown very low rates of side-effects when using similar technologies. Despite this, it is important to know any potential risks before agreeing to participate.

In a small number of people, using VR technology can cause a feeling of nausea and dizziness, similar to motion sickness. This is usually mild and settles quickly once the headset has been removed. With modern VR technology this is minimised, but if you feel any discomfort when using any of the technology, please let the research team know and you can stop using it.

Whilst using VR, people lose awareness of their surroundings. In rare circumstances, the user could sustain an injury from hazards within the room. To address this problem, you will only use the VR technology whilst sitting down on a chair that is able to rotate. The room will also be cleared of potential hazards.

The VR activities may involve some fast movements of your arms and upper body. There will also be an element of rotation of the head required to look around the VR environment. These movements could potentially cause increased pain in people with Fibromyalgia Syndrome. Additionally, people with Fibromyalgia Syndrome can experience pain when objects make contact with their skin. Use of the VR/EEG equipment could potentially cause pain in some people as they are worn on the head. If you feel discomfort at any time during the experiment, you are free to stop using the equipment.

There are some medical conditions (such as migraines and seizures) where symptoms can triggered when viewing screens or experiencing flashing lights. If you do suffer from a condition like this, you will be excluded from the study as a safety precaution.

In certain individuals it is possible that the special conductive gel used to attach the EEG electrodes may cause some minor discomfort due to skin irritation. If this causes discomfort at the time of the testing, we will move the electrode to an unaffected area. This gel can also leave a flaky residue on the skin but can be removed following a hair wash.

Will my involvement in the study be confidential and how will you manage my data?

All data will be handled in accordance with the General Data Protection Regulation (GDPR; see end of document for further information). After collection, all personal details supplied by you will be stored securely on an encrypted university server and will only be used for the purposes of carrying out the research project as detailed in this participant information sheet. Your personal details will be entered onto an encrypted spreadsheet alongside your participant number. This data will be kept on a secure server, separate from other study data and will only be accessible to the research team when absolutely necessary. All other data collected will be anonymised to ensure your confidentiality as a participant. Again, this data will be kept on a secure university server. The research team will not share your identifiable details with anyone outside of the study, with the exceptions of a reasonable request from a regulatory authority or if there are significant concerns over the safety of you or another party. With your consent, the research team may use your details to contact you about other relevant research projects.

All electronic data and written consent forms will be kept available for 10 years as per current UEA data protection protocol and Medical Research Council (MRC) guidelines. After this 10-year period expires, data will be securely deleted by the research team. Other forms of written data such as the questionnaires will be stored securely for the duration of the study period. After the study period has ended, this written data will be securely disposed of.

Who has reviewed this study?

An independent group of people called a Research Ethics Committee have reviewed the research proposal. These groups look at the ethical and safety implications of research to protect participants. This study has been reviewed and approved by an NHS research ethics committee.

Am I getting paid for participating in the study?

Although we really appreciate your interest and participation, unfortunately we do not have the budget available to pay our participants. However, there will be a small amount of money available to each participant for help towards travel costs to get to/from the study location. We will also provide water refreshments when you attend for your experiment(s).

Who is funding/sponsoring this study?

The British Society of Rheumatology (BSR) have provided a grant to fund this study. The Norwich Academic Training Office (NATO) and our industry partner (Orbital Media) are also providing smaller amounts of funding. The Principal Investigator is an Academic Clinical Fellow (ACF) and his salary is funded by the National Institute for Health Research (NIHR). The University of East Anglia (UEA) are also providing financial support and the regulatory infrastructure for the research to take place. Orbital Media are a digital health company providing the VR program for use in the experiments.

What will happen to the results of the research study?

The research team intend for the anonymised results to be presented at national/international conferences and submitted for publication in peer-reviewed scientific journals.

What will happen if I decide to withdraw from the study?

You may withdraw from the study at any time and do not have to give a reason for doing so. We may ask you for the reason you decided to withdraw to help inform the design of our future studies, but you should not feel obliged to give a reason.

Any data collected while you were taking part in the study will be retained in line with GDPR. If you were to withdraw before the data analysis has taken place, you have the right to ask for your data to be excluded from the final analysis. If you do not specifically ask for your data to be excluded, it may still be included in the final analysis.

What do I do now?

If you wish to take part in the study, we would now ask you to complete and return the reply slip located at the bottom of your invitation letter. Alternatively, you can email vipa.study@uea.ac.uk to declare your interest in taking part. Either with this reply slip, or within the email, please indicate which of the experiments you would like to be involved in (see above for more details of each experiment). Please also provide your telephone number and email address so that we can contact you regarding the next steps.

After receiving your response, a member of the research team will contact you (using the details that you provide) to confirm that you are eligible to participate in the study and offer you dates/times to attend the experiment(s). At this stage and throughout the study, the research team will answer any questions that you may have.

If you have any further questions relating to this study, please get in touch with the main researcher using the contact details below.

If you do not wish to take part, please ignore this communication.

Contact details of the main researcher:

Dr Jordan Tsigarides

University of East Anglia Centre for Epidemiology Versus Arthritis

Bob Champion Research and Education Building (BCRE)

Colney Lane

Norwich

NR4 7UQ

Tel: 01603 591793

Email: vipa.study@uea.ac.uk

This page explains how health researchers use information from patients. If you are asked to take part in research, you can ask what will happen in the study.[1]

What is patient data?

When you go to your GP or hospital, the doctors and others looking after you will record information about your health. This will include your health problems, and the tests and treatment you have had. They might want to know about family history, if you smoke or what work you do. All this information that is recorded about you is called patient data or patient information.

When information about your health care joins together with information that can show who you are (like your name or NHS number) it is called identifiable patient information. It's important to all of us that this identifiable patient information is kept confidential to the patient and the people who need to know relevant bits of that information to look after the patient. There are special rules to keep confidential patient information safe and secure.

What sort of patient data does health and care research use?

There are lots of different types of health and care research.

If you take part in a clinical trial, researchers will be testing a medicine or other treatment. Or you may take part in a research study where you have some health tests or answer some questions. When you have agreed to take part in the study, the research team may look at your medical history and ask you questions to see if you are suitable for the study. During the study you may have blood tests or other health checks, and you may complete questionnaires. The research team will record this data in special forms and combine it with the information from everyone else in the study. This recorded information is research data.

In other types of research, you won't need to do anything different, but the research team will be looking at some of your health records. This sort of research may use some data from your GP, hospital or central NHS records. Some research will combine these records with information from other places, like schools or social care. The information that the researcher collects from the health records is research data.

Why does health and care research use information from patients?

In clinical trials, the researchers are collecting data that will tell them whether one treatment is better or worse than other. The information they collect will show how safe a treatment is, or whether it is making a difference to your health. Different people can respond differently to a treatment. By collecting information from lots of people, researchers can use statistics to work out what effect a treatment is having.

Other types of research will collect data from lots of health records to look for patterns. It might be looking to see if any problems happen more in patients taking a medicine. Or to see if people who have screening tests are more likely to stay healthier.

Some research will use blood tests or samples along with information about the patient's health. Researchers may be looking at changes in cells or chemicals due to a disease.

All research should only use the patient data that it really needs to do the research. You can ask what parts of your health records will be looked at.

How does research use patient data?

If you take part in some types of research, like clinical trials, some of the research team will need to know your name and contact details so they can contact you about your research appointments, or to send you questionnaires. Researchers must always make sure that as few people as possible can see this sort of information that can show who you are.

In lots of research, most of the research team will not need to know your name. In these cases, someone will remove your name from the research data and replace it with a code number. This is called coded data, or the technical term is pseudonymised data. For example, your blood test might be labelled with your code number instead of your name. It can be matched up with the rest of the data relating to you by the code number.

In other research, only the doctor copying the data from your health records will know your name. They will replace your name with a code number. They will also make sure that any other information that could show who you are is removed. For example, instead of using your date of birth they will give the research team your age. When there is no information that could show who you are, this is called anonymous data.

Where will my data go?

Sometimes your own doctor or care team will be involved in doing a research study. Often, they will be part of a bigger research team. This may involve other hospitals, or universities or companies developing new treatments. Sometimes parts of the research team will be in other countries. You can ask about where your data will go. You can also check whether the data they get will include information that could show who you are. Research teams in other countries must stick to the rules that the UK uses.

All the computers storing patient data must meet special security arrangements.

If you want to find out more about how companies develop and sell new medicines, the Association of the British Pharmaceutical Industry has information on its website (https://www.abpi.org.uk/).

What are my choices about my patient data?

- You can stop being part of a research study at any time, without giving a reason, but the research team will keep the research data about you that they already have. You can find out what would happen with your data before you agree to take part in a study.
- In some studies, once you have finished treatment the research team will continue to collect some information from your doctor or from central NHS records over a few months or years so the research team can track your health. If you do not want this to happen, you can say you want to stop any more information being collected.
- Researchers need to manage your records in specific ways for the research to be reliable. This means that they won't be able to let you see or change the data they hold about you. Research could go wrong if data is removed or changed.

What happens to my research data after the study?

Researchers must make sure they write the reports about the study in a way that no-one can work out that you took part in the study.

Once they have finished the study, the research team will keep the research data for several years, in case they need to check it. You can ask about who will keep it, whether it includes your name, and how long they will keep it.

Usually your hospital or GP where you are taking part in the study will keep a copy of the research data along with your name. The organisation running the research will

usually only keep a coded copy of your research data, without your name included. This is kept so the results can be checked.

If you agree to take part in a research study, you may get the choice to give your research data from this study for future research. Sometimes this future research may use research data that has had your name and NHS number removed. Or it may use research data that could show who you are. You will be told what options there are. You will get details if your research data will be joined up with other information about you or your health, such as from your GP or social services.

Once your details like your name or NHS number have been removed, other researchers won't be able to contact you to ask you about future research.

Any information that could show who you are will be held safely with strict limits on who can access it.

You may also have the choice for the hospital or researchers to keep your contact details and some of your health information, so they can invite you to take part in future clinical trials or other studies. Your data will not be used to sell you anything. It will not be given to other organisations or companies except for research.

Will the use of my data meet GDPR rules?

GDPR stands for the General Data Protection Regulation. In the UK we follow the GDPR rules and have a law called the Data Protection Act. All research using patient data must follow UK laws and rules.

Universities, NHS organisations and companies may use patient data to do research to make health and care better.

When companies do research to develop new treatments, they need to be able to prove that they need to use patient data for the research, and that they need to do the research to develop new treatments. In legal terms this means that they have a 'legitimate interest' in using patient data.

Universities and the NHS are funded from taxes and they are expected to do research as part of their job. They still need to be able to prove that they need to use patient data for the research. In legal terms this means that they use patient data as part of 'a task in the public interest'.

If they could do the research without using patient data they would not be allowed to get your data.

Researchers must show that their research takes account of the views of patients and ordinary members of the public. They must also show how they protect the privacy of the people who take part. An NHS research ethics committee checks this before the research starts.

What if I don't want my patient data used for research?

You will have a choice about taking part in a clinical trial testing a treatment. If you choose not to take part, that is fine.

In most cases you will also have a choice about your patient data being used for other types of research. There are two cases where this might not happen:

1. When the research is using anonymous information. Because it's anonymous, the research team don't know whose data it is and can't ask you.
2. When it would not be possible for the research team to ask everyone. This would usually be because of the number of people who would have to be contacted. Sometimes it will be because the research could be biased if some people chose not to agree. In this case a special NHS group will check that the reasons are valid. You can opt-out of your data being used for this sort of research. You can ask your GP about opting-out, or you can find out more on the HRA's website (https://www.hra.nhs.uk/information-about-patients/).

Who can I contact if I have a complaint?

If you want to complain about how researchers have handled your information, you should contact the research team. If you are not happy after that, you can contact the Data Protection Officer. The research team can give you details of the right Data Protection Officer.

If you are not happy with their response or believe they are processing your data in a way that is not right or lawful, you can complain to the Information Commissioner's Office (ICO) (www.ico.org.uk or 0303 123 1113).

1 https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/templates/template-wording-for-generic-information-document/

University of East Anglia

Study Title: VR-FOCUS: Leveraging sensor-based technologies and machine learning for real-time optimisation of cognitive load within virtual reality pain management interventions

Date & Version: July 2024, Version 1.0

Principal Investigator: Dr Jordan Tsigarides

Research Supervisor: Prof Alexander Macgregor

Dear Sir/Madam,

We are inviting you to take part in a research study. Before you decide that you want to be involved in the study, it is important that you understand why we are conducting the research and what your participation will involve.

Please read the following information carefully and take time to consider if you would like to take part. After reading this information sheet, any questions about the study can be directed to the main researcher (contact details below).

Why is this research being done?

The Burden of Chronic Pain in the UK

Chronic pain affects up to 50% of the UK population and is a major cause of suffering and disability worldwide. Unfortunately, many of the recommended non-drug treatments are hard to access in the UK, meaning that people living with pain often have few treatment options. Virtual Reality (VR) offers a new and promising way to deliver effective non-drug treatments in the home setting. VR has been shown to help manage pain in various settings and in various ways, including engaging the user's attention away from pain.

Virtual Reality – A Potential New Treatment Needing Personalisation

VR is a novel technology in which a person becomes fully immersed within a 360-degree, computer generated, three-dimensional, 'virtual' environment through use of a head-mounted system. Previous research at the University of East Anglia (UEA) has tested VR for chronic pain management. These studies showed that VR is both acceptable to patients and can be effective in reducing pain. However, not everyone in these studies experienced the same level of pain relief. This variation is thought to be due to the lack of personalisation of the VR experience to each individual's needs.

Using Sensors To Personalise VR

Everyone's experience of pain is different, influenced by many factors such as their biology, psychology, and social situation. Therefore, a "one-size-fits-all" approach to VR pain management fails to consider the individual. To address this, we can use advanced technology embedded in VR systems to collect real-time data about how each person responds to the VR content. This data includes information like changes in pupil size, which can tell us about cognitive processes (such as their cognitive load) whilst using VR. By combining this data with information provided by the users themselves, supported by advanced artificial intelligence techniques, we have the potential to develop highly personalised VR therapies of the future. These therapies can adapt in real-time to meet the specific needs of each individual, making treatment more tailored and hopefully more effective.

The VR-FOCUS Study

To achieve this goal, our study aims to create a reliable method for measuring cognitive load using VR and eye-tracking technology. We will test this method in people both with and without chronic pain to understand how well it works and how it can be improved. This foundational research is essential for developing future VR treatments that are tailored to the individual's unique experiences.

Who are we looking to include in this study?

We will be looking for individuals to participate in this study if they are either:

1) A **current student or staff member** based at the UEA who is <u>NOT</u> suffering from acute or chronic pain.

   OR

2) An **individual who is currently living with chronic pain** (pain experienced for ≥3 months) that has participated in previous ethically approved research studies at the UEA.

Those included need to be **18 years or older**, have **conversational English** language skills and the **capacity to consent**.

Who are we excluding from participation in this study?

People will not be able to take part if they:

a) Suffer with any conditions that are triggered or made worse by exposure to flashing lights or screens.
b) Have a current visual or hearing impairment that would prevent them from using the VR technology included. This would include severe near sightedness (-6 diopters or worse).
c) Have a current facial injury or condition that may make use of a VR system uncomfortable or impractical
d) Have any other conditions or are taking any medication that is known to affect the movements of either eye, or the reactivity of the pupil.

What will I have to do if I agree to take part?

(1) Signing Up to Participate

If you do choose to participate, you will be directed to the study's website where you will be asked a few questions to ensure you are eligible for the study. There will also be access to this participant information sheet and a consent form to confirm your consent to participate. You will then be asked to provide your email address, after which you will be emailed by the research team with your participant ID number and a weblink to choose a suitable time to attend for your data collection session at the UEA.

(2) Your Data Collection Session – Practical Information and Introduction

Your session will be conducted in a room at the UEA (confirmed by the research team via email in advance). Your session should take approximately 1 hour from start to finish. Upon meeting the research team, you will be welcomed to the session and will gain an introduction to the relevant technologies, questionnaires and activities involved. Prior to using the VR systems you will be asked to complete some short questionnaires using either a laptop or tablet. You will also be offered refreshments from the start (and throughout), will have access to this participant information sheet to review at any time, and will be able to ask questions to the research team.

(3) Your Data Collection Session – Setting up the VR System for You

You will then be supported by the research team to use the VR head-mounted display (with integrated eye tracking) and start the relevant application. To ensure the eye tracking system is set up for your eyes, you will complete a short calibration exercise where you will be asked to gaze at various dots in the virtual environment. This will also adjust the VR system for your focal length. Prior to completing any further activities, the research team will make sure you are comfortable with the VR system and can see elements of the environment clearly.



Figure 1: An example VR eye tracking system (HTC VIVE XR Elite). Note: This may not be the VR system used in your data collection session.

(4) Your Data Collection Session – The First VR Activity (VR N-Back Task)

Following set-up, you will be asked to complete the first VR activity. This activity is a VR version of a validated psychometric task called the N-Back task; a simple memory game that helps us to understand how the brain works whilst you are focusing, remembering and responding to things in the environment.

The N-Back task is split into different challenge 'levels' from 0-Back to 4-Back. You will start with the 0-Back task, but after this the order in which you complete the levels will be random (e.g. you may perform the 2-Back task next, followed by the 1-Back, 4-Back and finally the 3-Back tasks). Each N-Back task will present a sequence of letters to you within a large square in the centre of your vision. Each letter will be presented for a short period of time, with a slightly longer (but still short) period of time between letters. Your task will be to decide if the current letter matches the one that appeared a few steps earlier and press a button on the controller when you think it does. This will depend on the level of the task you're completing, more specifically:

- 0-Back Task: You will be asked to respond (press a button) every time the X letter appears. This is a baseline task and does not require you to remember previous letters.
- 1-Back Task: You will be asked to respond when the same letter appears twice in a row (e.g. a B letter followed by another B).
- 2-Back Task: You will be asked to respond when the same letter appears two steps after it appeared previously (e.g. a B letter, followed by a C, followed by a another B)
- 3-Back Task: You will be asked to respond when the same letter appears three steps after it appeared previously (e.g. B, C, A, B)

- 4-Back Task: You will be asked to respond when the same letter appears four steps after it appeared previously (e.g. B, C, A, Z, B)

For each task, there will be a short 'demo' run first to give you a chance to practice and familiarise yourself with the task. Please note, although your performance on these tasks will be monitored for the purposes of the study, this is NOT a test of your ability to complete the tasks successfully. The tasks are used to provide different levels of cognitive load (difficulty) to challenge your working memory. See Figure 2 below for a visual representation of the tasks.

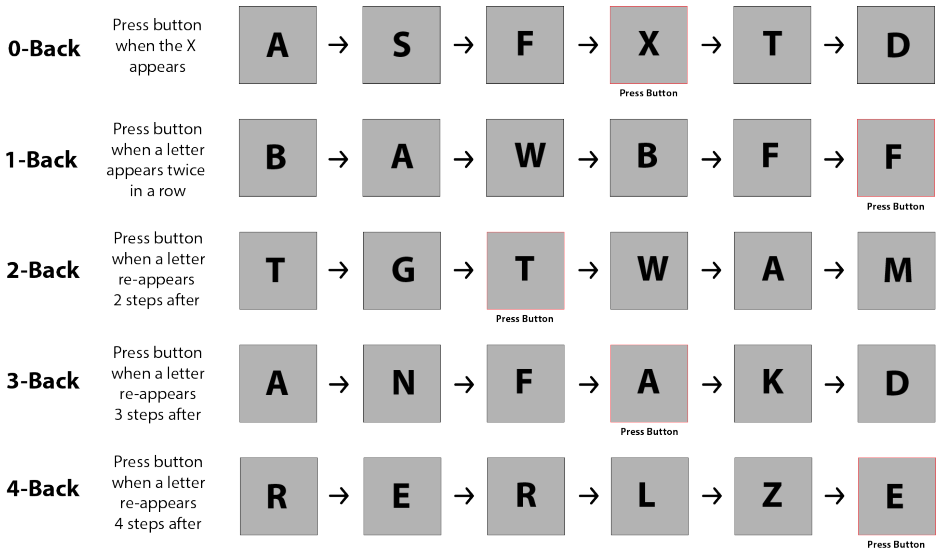


Figure 2: The First Activity: VR N-Back Task - Explanation of the different levels.

Eye tracking data will be captured throughout the tasks within the VR system, and after completing each 'level' you will be asked to complete a short questionnaire within the VR system. After completing all of the tasks, you will be asked to remove the VR system, complete a short questionnaire and given a break (with refreshments) before completing the second VR activity.

(5) Your Data Collection Session – The Second VR Activity (VR Pain Management Task)

Following your break, you will be supported to use the VR system for a second time by the research team. You will be asked to re-complete the short eye tracking calibration prior to starting the VR Pain Management activity. This activity was designed initially for people living with chronic pain as a fun, but challenging memory task and was studied in previous research at the UEA (The VIPA and RIVET studies). The activity is a simple sequence memory task set in a farm-based environment (see Figure 3 below). You will undergo a tutorial before starting to familiarise you with the task. Then, you will be given 4 minutes to complete the task in which you will be challenged to remember the sequence of trees appearing from different plots in a 3 x 3 grid. After a sequence is shown to you, you will need to remember the order and use the VR controller to point at and click the plots in the same order as the sequence. The sequence will start with three and will increase by steps of one with each successful recall. You will also receive a score based on the number of successful recalls achieved.

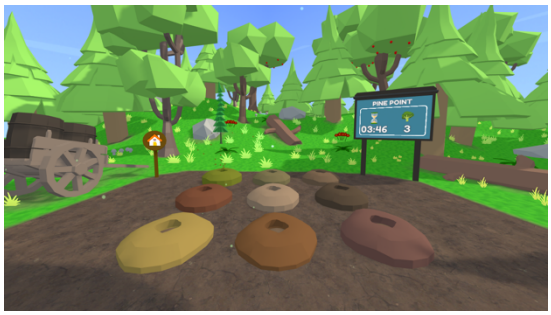


Figure 3: The Second Activity: VR Pain Management Memory-based Task

As with the first activity, your eye tracking data will be collected throughout completion of the activity. After completing the activity once you will be asked to complete a short VR questionnaire prior to removing the VR system, with a further short questionnaire after removing the VR system.

(6) Your Data Collection Session – The End of the Session

After completing the second VR activity and associated questionnaires, the research team will check you comfortable before explaining how your data will be processed and inviting any questions. Following this, the research team will close the session.

Do I have to take part?

No, choice to participate in the study is entirely voluntary and you can freely choose not to take part in the study. Even if you are included as a participant in this study, you may withdraw at any time without needing to provide a reason. You may also decide to refuse to answer any question(s) involved in the study if you feel uncomfortable with answering them.

If I don't take part, will it affect my treatment or education?

If you decide not to be involved in this study, your treatment or educational provision will be in no way affected. Similarly, if you decide to take part at first, but decide to withdraw from the study, these will be in no way affected.

Why are you using eye tracking in your experiments?

Using eye tracking allows us to look at small changes in the way your eyes adapt to various tasks as a reflection of your cognitive processes. We feel that these changes in the eyes can be used in the future as an objective marker of things like the user's 'mental effort' and 'attention' which can be used to support a personalised VR therapy. Collecting data like this in the context of VR pain management has not been done before, and we are looking to collect a rich dataset within this study to support development of advanced analysis models (powered by artificial intelligence) and inform future work.

What are the possible benefits of taking part?

Although not a direct benefit for people taking part in the study, your participation will help us to inform the future development of VR programs and will support new ways or providing personalised VR therapeutics. This may aid in the production of more effective and acceptable VR interventions in the future. The results from your participation in this study will help to inform the scientific community on these types of intervention. The results will also help to inform further research trials. In the future, by adding to the current research base, we can move closer to developing personalised VR treatments that could be used in the management of chronic pain.

For those experiencing pain at the time of the study, there is early evidence that use of VR technology can help to reduce pain in people with chronic pain. Within this study, you may experience short-term relief of pain or other symptoms when using the VR technology. It is also possible that this relief of your symptoms could last for an extended period after removing the VR technology.

What are the possible risks and discomforts of taking part in this research?

The risks of participating in this research study are minimal, and both our previous work and similar studies have shown very low rates of side-effects when using similar VR technologies. Despite this, it is important to know any potential risks before agreeing to participate.

In a small number of people, using VR technology can cause a feeling of nausea and dizziness, similar to motion sickness. This is usually mild and settles quickly once the headset has been removed. With modern VR technology this is minimised, but if you feel any discomfort when using any of the technology, please let the research team know and you can stop using it at any time.

Whilst using VR, people lose awareness of their surroundings. In rare circumstances, the user could sustain an injury from hazards within the room. To address this problem, you will only

use the VR technology whilst sitting down on a chair that is able to rotate. The room will also be cleared of potential hazards.

The movement and attention required for the VR activities could theoretically cause increased pain in people with chronic pain conditions. If you feel discomfort at any time during the experiment, you are free to stop using the equipment.

There are some medical conditions (such as migraines and seizures) where symptoms can triggered when viewing screens or experiencing flashing lights. You will be asked about this in our initial eligibility questions. If you do disclose that you suffer from a condition like this, you will be excluded from the study as a safety precaution.

Will my involvement in the study be confidential and how will you manage my data?

All data will be handled in accordance with the General Data Protection Regulation (GDPR). After collection, all personal details supplied by you will be stored securely on an encrypted university server and will only be used for the purposes of carrying out the research project as detailed in this participant information sheet. Your personal details will be entered onto an encrypted spreadsheet alongside your participant number. This data will be kept on a secure server, separate from other study data and will only be accessible to the research team when absolutely necessary. All other data collected will be anonymised to ensure your confidentiality as a participant. Again, this data will be kept on a secure university server. Eye tracking data collected can not be used to identify you. The research team will not share your identifiable details with anyone outside of the study, with the exceptions of a reasonable request from a regulatory authority or if there are significant concerns over the safety of you or another party. With your consent, the research team may use your details to contact you about other relevant research projects.

All electronic data will be kept available for 10 years as per current UEA data protection protocol and Medical Research Council (MRC) guidelines. After this 10-year period expires, data will be securely deleted by the research team.

Who has reviewed this study?

An independent group of people called a Research Ethics Committee have reviewed the research proposal. These groups look at the ethical and safety implications of research to protect participants. This study has been reviewed and approved by the UEA's Faculty of Medicine and Health research ethics committee.

Am I getting paid for participating in the study?

Although we really appreciate your interest and participation, unfortunately we do not have the budget available to pay our participants. However, we will provide a variety of refreshments when you attend for your experiment(s).

Who is funding/sponsoring this study?

This study forms part of a PhD project conducted by Dr Tsigarides, funded by the UEA and a local digital health company acting as an industry sponsor (Revolve Labs Ltd). One of the VR activities used has been provided free-of-charge by a second industry partner for use in this study (Orbital Global Ltd).

What will happen to the results of the research study?

The research team intend for the anonymous results to be presented at national/international conferences and submitted for publication in peer-reviewed scientific journals.

What will happen if I decide to withdraw from the study?

You may withdraw from the study at any time and do not have to give a reason for doing so. We may ask you for the reason you decided to withdraw to help inform the design of our future studies, but you should not feel obliged to give a reason.

Any data collected while you were taking part in the study will be retained in line with GDPR. If you were to withdraw before the data analysis has taken place, you have the right to ask for your data to be excluded from the final analysis. If you do not specifically ask for your data to be excluded, it may still be included in the final analysis.

What to do now

If you wish to take part in the study, we would now ask you to complete the eligibility questionnaire on our study webpage (see below for a QR code/link to access), prior to confirming your consent and providing your email address. Depending on how you were informed of the study, you may have already been directed to the study webpage to access this participant information sheet – in which case please progress to the eligibility questionnaire.

https://forms.office.com/e/NFS3G5YA2s

If you have any further questions relating to this study, please get in touch with the main researcher using the contact details below. We have also provided contact details for the supervising researcher if required.

If you do not wish to take part, please ignore this communication.

Contact details of the main researcher and supervising researcher:

Dr Jordan Tsigarides

University of East Anglia Centre for Epidemiology Versus Arthritis

Bob Champion Research and Education Building (BCRE)

Colney Lane

Norwich

NR4 7UQ

Email: J.Tsigarides@uea.ac.uk

Prof Alexander Macgregor

University of East Anglia Centre for Epidemiology Versus Arthritis

Bob Champion Research and Education Building (BCRE)

Colney Lane

Norwich

NR4 7UQ

Email: a.macgregor@uea.ac.uk