



Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

Forbes, Owen, Schwenn, Paul E., Wu, Paul Pao-Yen, Santos-Fernandez, Edgar, Xie, Hong-Bo, Lagopoulos, Jim, McLoughlin, Larisa T., Sacks, Dashiell D., Mengersen, Kerrie, & Hermens, Daniel F.
(2022)

EEG-based clusters differentiate psychological distress, sleep quality and cognitive function in adolescents.

Biological Psychology, 173, Article number: 108403.

This file was downloaded from: <https://eprints.qut.edu.au/235139/>

© 2022 Elsevier B.V.

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution-Noncommercial-No Derivative Works 2.5

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

<https://doi.org/10.1016/j.biopsycho.2022.108403>

EEG-based clusters differentiate psychological distress, sleep quality and cognitive function in adolescents

Owen Forbes^{1,2*}, Paul E. Schwenn³, Paul Pao-Yen Wu^{1,2}, Edgar Santos-Fernandez^{1,2}, Hong-Bo Xie^{1,2,4}, Jim Lagopoulos³, Larisa T. McLoughlin³, Dashiell D. Sacks³, Kerrie Mengersen^{1,2}, Daniel F. Hermens³

1 School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

2 ARC Centre of Excellence in Mathematical and Statistical Frontiers (ACEMS), Brisbane, QLD, Australia

3 Thompson Institute, University of the Sunshine Coast, Birtinya, QLD, Australia

4 School of Information Science and Engineering, Yunnan University, Kunming, 650500, China

* owen.forbes@hdr.qut.edu.au

Level 8, Y Block

QUT Gardens Point Campus

2 George St

Brisbane, QLD, Australia, 4000

1 Abstract

Introduction

To better understand the relationships between neurophysiology, cognitive function and psychopathology risk in adolescence there is value in identifying data-driven subgroups based on measurements of brain activity and function, and then comparing cognition and mental health between such subgroups.

Methods

We developed a flexible and scaleable multi-stage analysis pipeline to identify data-driven clusters of 12-year-olds ($M = 12.64$, $SD = 0.32$) based on frequency characteristics calculated from resting state, eyes-closed electroencephalography (EEG) recordings. For this preliminary cross-sectional study, EEG data was collected from 59 individuals in the Longitudinal Adolescent Brain Study (LABS) being undertaken in Queensland, Australia. Applying multiple unsupervised clustering algorithms to these EEG features, we identified well-separated subgroups of individuals. To study patterns of difference in cognitive function and mental health symptoms between clusters, we applied Bayesian regression models to probabilistically identify differences in these measures between clusters.

Results

We identified 5 core clusters associated with distinct subtypes of resting state EEG frequency content. Bayesian models demonstrated substantial differences in psychological distress, sleep quality and cognitive function between clusters. By examining associations between neurophysiology and health measures across clusters, we have identified preliminary risk and protective profiles linked to EEG characteristics.

Conclusion

This method provides the potential to identify neurophysiological subgroups of adolescents in the general population based on resting state EEG, and associated patterns of health and cognition that are not observed at the whole group level. This approach offers potential utility in clinical risk prediction for mental and cognitive health outcomes throughout adolescent development.

Keywords: electroencephalography (EEG), psychopathology, adolescence, clustering, Bayesian statistics

2 Introduction

Mental health issues are a significant and growing problem in youth populations, and confer significant burden on public health, communities and individual wellbeing (Sawyer et al., 2018; McGorry and Mei, 2018). Adolescence is a period in which many mental health problems are thought to emerge, as the brain is rapidly developing and changing (Paus et al., 2008; McGorry et al., 2011). It is also a crucial period in which to identify emerging psychopathology, to anticipate risk trajectories and enable early intervention (McGorry and Mei, 2018). Given the high burden of mental health problems in adolescents along with the overlapping risk across disorder diagnoses in this age range, there is urgent need for novel data-driven methods to support improved diagnoses, risk prediction and early intervention for psychopathology and cognitive development in adolescence.

In this work we report on a flexible analysis pipeline that enables identification of subgroups of individuals based on frequency features of electroencephalography (EEG) data, and allows comparison of differences in neurophysiology and health measures between these subgroups. Our aim was to identify empirical, data-driven patterns relating brain activity to cognitive function and mental health outcomes that are more robust and generalisable than the findings available through a case-control approach. Traditional approaches to characterising individuals based on brain measurements typically rely on a case-control approach, which separates individuals into groups on the basis of a clinical diagnosis, contrasted with ‘healthy controls’ without any diagnosis (Latzman and DeYoung, 2020). There are several shortcomings to the case-control approach, including that it is bound to find patterns of difference between groups categorised by traditional diagnostic structures based on clinician judgement. As a result case-control studies have limited ability to generalise and capture information relating to psychopathology and cognitive function issues occurring on a continuum, and are limited to either contradicting or strengthening existing assumptions about these categorical differences. In this framework, differences in brain characteristics of interest between cases and controls are also likely to be confounded by many extraneous factors (Latzman and DeYoung, 2020). Recent studies have suggested that EEG studies of mental health should strive to be more independent of clinical diagnostic categories and examine mental health symptoms from a data-driven neurophysiology perspective (Loo et al., 2016; Keizer, 2019).

Previous research has demonstrated that features calculated from EEG data are associated with mental health and cognitive function outcomes across multiple life stages (McVoy et al., 2019; Loo et al., 2016). Particularly in the frequency domain, a large variety of studies have demonstrated these sorts of associations. For example, the ratio of theta to beta activity has been used to support diagnosis and prognosis in individuals with attention deficit hyperactivity disorder (ADHD) (Arns et al., 2013). Resting state theta power has been associated with cognitive performance in children with ADHD (Hermens et al., 2005). Various EEG-based biomarkers have been associated with treatment response for depression in adults (Baskaran et al., 2012). EEG signal complexity or entropy has been shown to be associated with neurological conditions including epilepsy and autism spectrum disorders (Acharya et al., 2012; Catarino et al., 2011). These are just a few of the many studies which have identified differences between controls and individuals with a particular diagnosis, based on specific features calculated from EEG data. By studying novel combinations of summary features in a data-driven framework, we can develop unique EEG-based characterisations of individuals and investigate associations between distinct EEG profiles and health outcomes of interest.

Mental disorder risk in childhood and adolescence is understood to be pluripotential, with non-specific risk overlapping across a number of categorical diagnostic outcomes (Raballo and Poletti, 2020; Hartmann et al., 2019). The prevailing case-control approach to understanding the relationships between neurophysiology and psychopathology is hindered by flaws in traditional categorical approaches to mental health diagnoses, including arbitrary boundaries between pathology and normality, overlap and co-occurrence within diagnostic categories (Ko-

to et al., 2017). Examples of alternative approaches to the case-control paradigm emphasise data-driven patterns of co-occurrence between dimensional mental health symptoms (Ousley and Cermak, 2014; Craddock et al., 2018; Forbes et al., 2021), or neurocognitive test scores (Crouse et al., 2020).

Departing from the case-control approach, we can identify data-driven subgroups of individuals and examine differences between subgroups. By starting with analyses of brain activity to characterise individuals before looking at health outcomes, we have greater opportunity to build robust models and generate novel hypotheses from an empirical perspective. The methodological pipeline developed in this work offers a novel, flexible and scaleable approach for identifying subgroups of individuals on the basis of resting state EEG measurements. This builds on a nascent but growing body of work that uses data-driven analysis of EEG as a starting point for investigating the relationships between neurophysiology, health and functional outcomes (Barry et al., 2019; Bosl et al., 2018; Pillay et al., 2020). To enable development of these data-driven insights and improve our understanding of the relationships between brain characteristics and mental health during adolescent development, there is a need for population-based studies to obtain large datasets on these factors together in this understudied age range.

The Longitudinal Adolescent Brain Study (LABS), being conducted at the Thompson Institute in Queensland, Australia, is a longitudinal cohort study examining the interactions between environmental and psychosocial risk factors, and outcomes including cognition, self-report mental health symptoms, neuroimaging measures, and psychiatric diagnoses (Beaudequin et al., 2020). The aims for LABS include: identifying early manifestations of psychopathology in young people aged 12 - 17 years; monitoring neurophysiological changes and their relationships with external factors and mental health outcomes over this developmental period; and enabling better identification of risk and early intervention to improve cognitive development and mental health trajectories for young people. Further information on data collection and study protocols for LABS are provided in the Methods section. In this paper we aim to identify data-driven subgroups of LABS participants using EEG data. This is a preliminary study using cross-sectional EEG data to develop and validate this method for finding data-driven groups of young people with shared resting state EEG characteristics. Compared to other neuroimaging modalities, EEG has a number of strengths which motivated us to focus on EEG data for this study. Advantages of EEG include that it is low cost, widely available, and EEG acquisition from scalp electrodes is safe, non-invasive, and places a low burden on individuals being measured (Liedorp et al., 2009; Alizadeh and Dehzangi, 2016; Debener et al., 2015). Resting state functional magnetic resonance imaging (MRI) data was not available for all participants in this stage of LABS data, because it was the last modality in the MRI protocol and some participants ended the scanning session prior to acquisition of resting state data. Future research using LABS data will investigate longitudinal patterns of change in EEG and mental health characteristics, and integrate analyses of EEG with data from other neuroimaging modalities including structural and functional MRI.

To identify subgroups of adolescents based on EEG characteristics, we implement a selection of popular unsupervised clustering algorithms which together demonstrate good coverage of key common features across existing clustering algorithms. Unsupervised clustering methods are a useful set of statistical tools for identifying subgroups within data, without requiring prior information about labels of classifications of groups to which individual data points are expected to belong (Aggarwal and Reddy, 2014). Different clustering algorithms will emphasise or be sensitive to different aspects of input data in their clustering solutions, so it is valuable to consider results from multiple clustering algorithms especially when there is limited prior evidence about any subgroups that are expected to exist within a dataset. Comparing results across multiple algorithms allows identification of well-separated subgroups with more confidence and enables inference that is less sensitive to choice of method, as opposed to choosing one ‘best’ model.

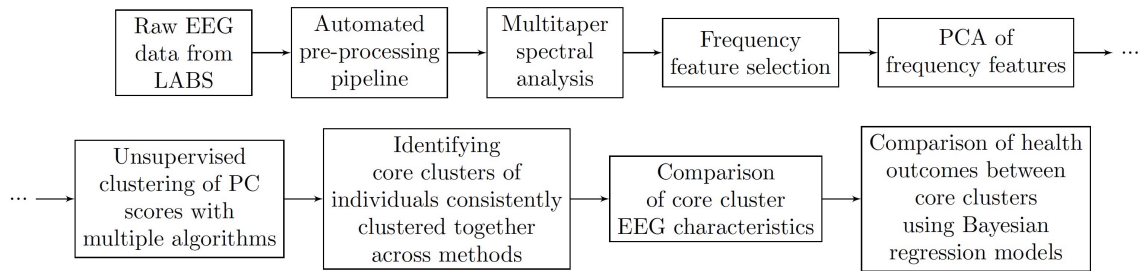


Figure 1: Methods overview block diagram

In the context of the literature introduced here it is evident that to better understand the relationships between brain activity, cognitive function and mental health risk in adolescence, there is value in identifying data-driven subgroups based on brain measurements and comparing cognitive function and mental health outcomes between these subgroups. The present study had several goals: (i) using frequency information from resting state EEG data collected at the baseline timepoint in LABS, to cluster 12-year-olds into subgroups; (ii) to identify stable, well-separated subgroups of individuals across different clustering methods; (iii) to interpret the EEG characteristics differentiating these subgroups; (iv) and to test for differences in cognitive function, mental health symptoms and other health indicators between these EEG-based clusters.

With these goals in mind, we aimed to address the following research questions:

1. Using popular unsupervised clustering algorithms, how can we develop a flexible analysis pipeline to identify subgroups of individuals based on frequency information from resting state EEG?
2. Across clusters of individuals, what characteristic profiles of resting state EEG content differentiate these clusters?
3. How do individuals across clusters vary in terms of psychological distress, wellbeing, sleep quality, and cognitive function?
4. Are these data-driven neurophysiological groups associated with different risk or protective profiles for mental health and cognitive function?

3 Methods

3.1 The Longitudinal Adolescent Brain Study

3.1.1 Participants

The present study uses cross-sectional data collected from a community sample of young people at baseline entry to the Longitudinal Adolescent Brain Study conducted at the Thompson Institute, University of the Sunshine Coast. Initiated in July 2018, LABS collects participant data every 4 months over 5 years, following participants from ages 12 to 17 — totalling 15 timepoints for each participant. At each timepoint, data collected includes: a self-report questionnaire on demographic information, mental health and wellbeing measures; a neuropsychiatric interview; a battery of cognitive assessments; and neuroimaging scans including resting state and task-based EEG, and structural and functional MRI. Inclusion criteria are being 12 years of age and in Grade 7 (the first year of secondary school) at the time of enrolment to participate. Exclusion criteria include neurological disorder diagnosis, intellectual disability,

major illness, or head injury with unconsciousness exceeding 30 minutes.

Further details on data collection and study protocols for LABS are available in previous publications (Beaudequin et al., 2020; Jamieson et al., 2020).

3.1.2 Ethical Approval

LABS received ethical approval from the University of the Sunshine Coast Human Research Ethics Committee (Approval Number: A181064). Written informed assent and consent was obtained from all participants and their guardian/s. For data analysis conducted at the Queensland University of Technology (QUT), the QUT Human Research Ethics Committee assessed this research as meeting the conditions for exemption from HREC review and approval in accordance with section 5.1.22 of the National Statement on Ethical Conduct in Human Research (2007). Exemption Number: 2021000159.

3.1.3 Data collection

EEG Data

For EEG recordings, participants were fitted with a Biosemi headcap mounted with 32 Ag/AgCl active electrodes and arrayed in a 10-10 configuration at a sampling rate of 1024 Hz using a Biosemi ActiveTwo system. To monitor eye movements, horizontal electrooculogram (EOG) was recorded using electrodes placed one-centimetre lateral to the outer canthus of each eye, while vertical EOG was recorded using supra- and infraorbital electrodes on the left eye. Each electrode was filled with a conductive gel and DC offsets were checked to assess electrode contact quality before recording. Eyes-closed, resting state data was recorded for 4 minutes, and participants were instructed to rest and minimise eye movements during the recording. For the present study we analysed data recorded from the vertex electrode (Cz) only, as we were not aiming to investigate patterns across different brain regions.

Mental Health & Cognitive Function Measures

In the present study we examine a variety of outcomes related to mental health, collected by self-report questionnaire in LABS. These include: psychological distress as measured by the K10 (Kessler et al., 2003), and SPHERE-12 (Hickie et al., 2001); Suicidal ideation measured by the Suicidal Ideation Attributes Scale (SIDAS) (Van Spijker et al., 2014); Wellbeing measured by the COMPAS-W (Gatt et al., 2014); and sleep quality, measured by the Pittsburgh Sleep Quality Index (PSQI) (Buysse et al., 1989).

We also examined a number of cognitive function measures from the CogState cognitive test battery including reaction time measures for the Detection (DET), Identification (IDN) and One Back (OBK) tasks, with log10 transformed reaction times for correct responses where lower scores indicate better performance. These measures also included accuracy for the Two Back (TWOB) task using an arcsine transformation of the square root of the proportion of correct responses, for which higher scores indicate better performance (Maruff et al., 2009). This is the standard scoring for these tests. Full details on LABS data collection protocols for these measures are available in previous publications (Beaudequin et al., 2020).

3.2 Statistical analyses

To address the research questions outlined above, this study implemented a multi-stage analysis pipeline. The first stage for clustering based on EEG frequency characteristics included: EEG pre-processing; frequency decomposition with multitaper analysis; and selection and calculation of frequency features. The second stage included: dimensionality reduction using principal component analysis; and applying a number of popular unsupervised clustering algorithms to

this dimension-reduced data.

In the third stage to identify well-separated subgroups of participants who are differentiated by characteristic EEG profiles, we selected core clusters of individuals who were consistently grouped together across the different clustering methods. This allowed us to identify well-separated subgroups with distinct profiles of EEG characteristics. After implementing this pipeline and identifying core clusters, we described the principal components and EEG frequency characteristics that differentiate these core clusters. Finally, we applied Bayesian regression models with cluster allocation as a categorical predictive variable to examine between-cluster differences in these health and cognitive function measures. Expected marginal means for pairwise contrasts between clusters were calculated from these regression models to identify the magnitude and associated uncertainty for pairwise differences in these measures between the clusters.

3.2.1 Pre-processing

EEG data tend to be characterised by non-stationarity, a low signal-to-noise ratio, and artifacts from a number of sources including endogenous biological signals, background electrical interference, and variable electrode contact quality (Reddy and Narava, 2013). In this study we implemented an automated data cleaning and pre-processing pipeline, using a variety of tools available through the EEGLAB toolbox in MATLAB (Delorme and Makeig, 2004), including: ICLabel using a trained algorithm to probabilistically label noise components from Independent Component Analysis (ICA) (Pion-Tonachini et al., 2019); Adaptive Mixture ICA (AMICA), which is demonstrated to be an effective variant of ICA for EEG artifact removal (Delorme et al., 2012); and Artifact Subspace Reconstruction (ASR) (Chang et al., 2019). An automated approach was implemented to enable efficient and consistent pre-processing for EEG data across a large number of participants, and to be consistent with the flexible and scalable design approach of this analysis pipeline. Further details on the automated EEG cleaning and pre-processing pipeline are presented in the Supplementary Materials.

3.2.2 Multitaper Analysis & selection of frequency features

To study frequency content of EEG data we used multitaper analysis, a frequency decomposition method that is suitable for non-stationary signals and offers good frequency specificity (Walden et al., 1998). This is a popular technique for frequency analysis that has been used widely in recent EEG research literature (Babadi and Brown, 2014; Prerau et al., 2017), and offers an improved signal-to-noise ratio for detecting rhythmic activity in a signal relative to other standard frequency decomposition methods (Delorme et al., 2007; Thomson, 1982).

Multitaper analysis was conducted using the Chronux toolbox in MATLAB (Bokil et al., 2010). This method enabled us to calculate power spectral densities with good frequency resolution and minimal ‘bleeding’ of power across adjacent frequency bands (Cohen, 2014). This analysis used a 2 second time segment (W) and a 2Hz frequency bandwidth (W), resulting in a total of 3 tapers ($2TW - 1$). These parameters were chosen to provide adequate frequency resolution, giving a detailed characterisation of the frequency distribution in each individual’s power spectral density (Prerau et al., 2017). Multitaper time-frequency spectrograms were calculated from resting state, eyes closed EEG recordings, and were then averaged over time to yield power spectral densities (PSDs) representing the average frequency content of EEG activity for each individual LABS participant during the recording session. In this context since we were using resting state EEG data we were not specifically interested in any time-varying features, and we aimed to characterise the time-robust frequency content across each 5 minute recording session of resting state EEG. Averaging frequency content over windowed time segments is a standard practice in EEG analysis (Lehembre et al., 2012; Jaramillo-Jimenez et al., 2021), implemented in common signal processing analyses such as Welch’s method (Zhao and He, 2013). This time averaging enables smoother estimates of the frequency distribution to

enable comparison of key features across individuals.

From these PSDs we calculated a selection of 8 commonly used summary metrics to capture important characteristics of individual EEG frequency content. Individual Alpha Frequency (IAF) was calculated as the frequency at which maximum power is observed in the range 7-14 Hz (Smit et al., 2006; Klimesch, 1999), which is known to typically increase through adolescence before reaching a peak frequency that has a typical mean of around 10 Hz in adulthood (Cragg et al., 2011). 95% Spectral Edge Frequency (SEF95) was calculated to identify the frequency below which 95% of the power for a PSD is contained, and indicative of the degree of dispersion in the PSD (Thomas et al., 2010). Spectral entropy was calculated as the Shannon entropy of the self-normalised PSD (Tian et al., 2017), implemented with the ‘ForeCA’ package in R (Goerg, 2013, 2020). Finally we included fractional power in bands of width 4 Hz between 0 and 16 Hz, and total power for the PSD. In adults, typical frequency bands and their approximate boundaries are delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-30 Hz), and gamma (over 30 Hz) (Campisi and La Rocca, 2014). For ease of interpretation we include these traditional frequency band labels as an approximate heuristic guide for labeling the fractional power features we used in 4 Hz intervals up to 16 Hz (low beta), though these bands are not uniformly representative of discrete functional categories of EEG activity in children and adolescents (Saby and Marshall, 2012).

Prior to principal component analysis, all of these frequency features were mean centred and scaled by the standard deviation to have zero-mean and unit-variance.

3.2.3 Principal Component Analysis

From the frequency features calculated from the results of multitaper analysis, we conducted dimensionality reduction with PCA using the `prcomp()` function from the ‘stats’ package in R (R Core Team, 2021). From visual inspection of pairwise scatterplots of these 8 input features, correlation plots for all pairwise combinations of features, and Bartlett’s test of sphericity to assess presence of linear correlations between variables, we identified the presence of linear correlations between these variables with no obvious nonlinear relationships, indicating that PCA was an appropriate approach for dimension reduction. We identified dimensionality reduction as an important step in this analysis pipeline, for several reasons. Due to the small number of participants in this dataset, inclusion of this dimensionality reduction step enabled flexibility and scalability of this method to accommodate a large number of input features and avoid some of the challenges around clustering high-dimensional data (Steinbach et al., 2004). It is also important to improve model parsimony and interpretability for the clustering stages of this work. The PCA step enabled flexibility of this method to accommodate higher-dimensional input data in future, for example by looking at a large number of summary features across multiple EEG channels to look at activity patterns across multiple brain regions. We chose to retain 3 principal components based on finding an elbow in the scree plot, which coincided with eigenvalues greater than 1. Looking at the principal component coefficients representing the input features in a lower dimensional space, we were able to identify the combinations and contrasts in EEG features that account for the most variance between individuals.

3.2.4 Unsupervised clustering

Based on the principal component scores from the previous stage of analysis, we implemented three popular unsupervised clustering algorithms to identify data-driven subgroups of individuals on the basis of EEG frequency characteristics. These algorithms were k -means, hierarchical clustering (HC) using Ward’s method, and Gaussian mixture model (GMM), each of which has different assumptions and sensitivities and provides different perspectives on subgroup structure within the data. These three clustering algorithms were chosen on the basis that they

are among the most widely used, and together they provide good coverage of some common attributes of existing clustering algorithms: being model-based (GMM and k -means) and not model-based (HC), probabilistic (GMM) and non-probabilistic (k -means and HC), nested (HC) and not nested (k -means and GMM) (Aggarwal and Reddy, 2014).

There are a number of available methods for clustering of nonstationary time series (Harvill et al., 2017; Gorji Sefidmazgi et al., 2014). Since we were principally interested in frequency content of resting state EEG in the context of mental health cognition, we chose to focus on clustering in the frequency domain rather than the time domain. As noted above, there is a well-established research literature connecting frequency characteristics to clinically relevant aspects of mental health and cognitive function, and we aimed to contribute to this area of EEG research.

We calculated results for k -means using the `kmeans()` function with default settings from the base package ‘stats’ in R (R Core Team, 2021). We calculated results for hierarchical clustering using the `hclust()` function with method “ward.D2” from the base package ‘stats’ in R (R Core Team, 2021). We calculated results for GMM with default settings including Euclidean distance and a diagonal covariance matrix using the package ‘ClusterR’ in R, version 1.22 (Mouselimis, 2020). For the applications in this work, we use the crisp projection of GMM results with allocation of each data point to the mixture component in which it is most likely to belong. Further details on each algorithm are presented in the Supplementary Materials.

For each clustering algorithm, the optimal number of clusters k was selected on the basis of a number of internal validation indices which could be calculated for all three methods. Internal validation indices can be used to identify the number of clusters that creates the most compact and well-separated set of subgroups in the data (Aggarwal and Reddy, 2014). Each index is calculated based on slightly different metrics, so using a selection of multiple indices can be more robust than relying on a single index. Selected indices included the Dunn index (Dunn, 1974), silhouette coefficient (Rousseeuw, 1987), Davies-Bouldin index (Davies, 1979), Calinski-Harabasz index (Caliński and Harabasz, 1974), and Xie-Beni index (Xie and Beni, 1991). Clustering internal validation indices were calculated using the R package ‘clusterCrit’, v1.2.8 (Desgraupes, 2018). We also assessed clustering stability metrics, calculated using the R package ‘fpc’, v2.2-9 (Hennig, 2020).

3.2.5 Identifying core clusters across methods

To identify subgroups of individuals that were consistent across the clustering methods implemented, we selected ‘core’ clusters that co-occurred in all three sets of results. The labels and locations of clusters were not consistent across methods — this is likely due to differing approaches to identifying cluster structure across the algorithms, as well as the modest degree of dispersion of individuals in the principal component space. To address the disagreement between clustering algorithms and identify clusters with distinct EEG profiles, we identified overlapping membership in the unique combinations of cluster labels across the 3 methods. We retained core clusters with the highest overlapping membership across the 3 methods, selecting a set of combined results where each cluster label from each method was represented only once in the combined labels to generate core clusters.

This process was undertaken to identify well-separated subgroups of individuals and enable inferences about members of these core clusters that are more robust and less sensitive to the choice of clustering method. This also enabled identification of clearly differentiated subgroups with distinct profiles of EEG characteristics. There are some limitations to this simple approach for integrating multiple sets of clustering results, in that individuals who are not consistently clustered across different algorithms (representing borderline or edge cases located between the final core clusters) are not allocated to the final set of core clusters. For this application

we decided this was a reasonable method, as internal validation criteria supported clustering solutions with a similar structure across the three algorithms and only a small number of participants ($n = 11$) were not allocated to the core clusters. This enabled us to identify final subgroups with clearly differentiated profiles of EEG characteristics. Further details on the intersection of clustering results and selection of core groups are provided in the Supplementary Materials.

3.2.6 Testing for differences in cognitive function & health measures between core clusters

To examine patterns of difference in levels of psychological distress, wellbeing, sleep quality and cognitive function across the final EEG-based clusters, we implemented Bayesian regression models using the "brms" package in R, version 2.15.0 (Bürkner, 2018). These models were designed to predict scores on each measure using core cluster allocation as a categorical predictive variable, with the largest cluster (cluster 1) as the reference category and including sex as a covariate. Prior sensitivity analyses were conducted for all Bayesian models, comparing the default improper flat prior on regression coefficients used in "brms" with two alternatives. The alternative priors tested were a normal distribution with mean 0 and standard deviation 10, and a Student's T distribution with 2 degrees of freedom, location 0 and scale 10. These priors each provided appropriate support, covering the anticipated range of the parameter values for the regression coefficients.

From these Bayesian regression models we calculated posterior probabilities for each outcome measure being greater in one cluster than another, for all pairwise combinations of the clusters. We also calculated expected marginal means for pairwise contrasts between clusters using the "emmeans" package in R, version 1.5.5-1 (Lenth, 2021). This allowed us to identify posterior credible intervals for the magnitude of the differences in these health measures between all pairwise combinations of the clusters.

4 Results

From cross-sectional baseline data for 12-year-old ($M = 12.64$, $SD = 0.32$) participants that had been recruited into LABS when this article was written, data from 59 individuals were included after removing individuals with missing or corrupted EEG data. Figure 2 presents examples of raw and pre-processed EEG data and a corresponding PSD for one participant in this dataset.

4.1 Principal Component Analysis

Based on visual inspection of the scree plot and identifying components with an eigenvalue greater than 1, the first 3 principal components were retained which together explain 80.6% of the total variance. Initial visual inspection of scatter plots of individuals by principal component scores indicated that individuals were well dispersed across these 3 components, and some areas of higher and lower density were apparent — suggesting the opportunity for unsupervised clustering methods to identify useful subgrouping structure in this dataset. Further details including eigenvalues, proportions of variance explained and a scree plot are available in the Supplementary Materials.

4.1.1 PCA Interpretation

Table 1 presents coefficients for each of the 8 input frequency features on the 3 retained components. Choosing a heuristic threshold magnitude of 0.4 for these coefficients, we can identify the most influential variables contributing to the variance explained in each component.

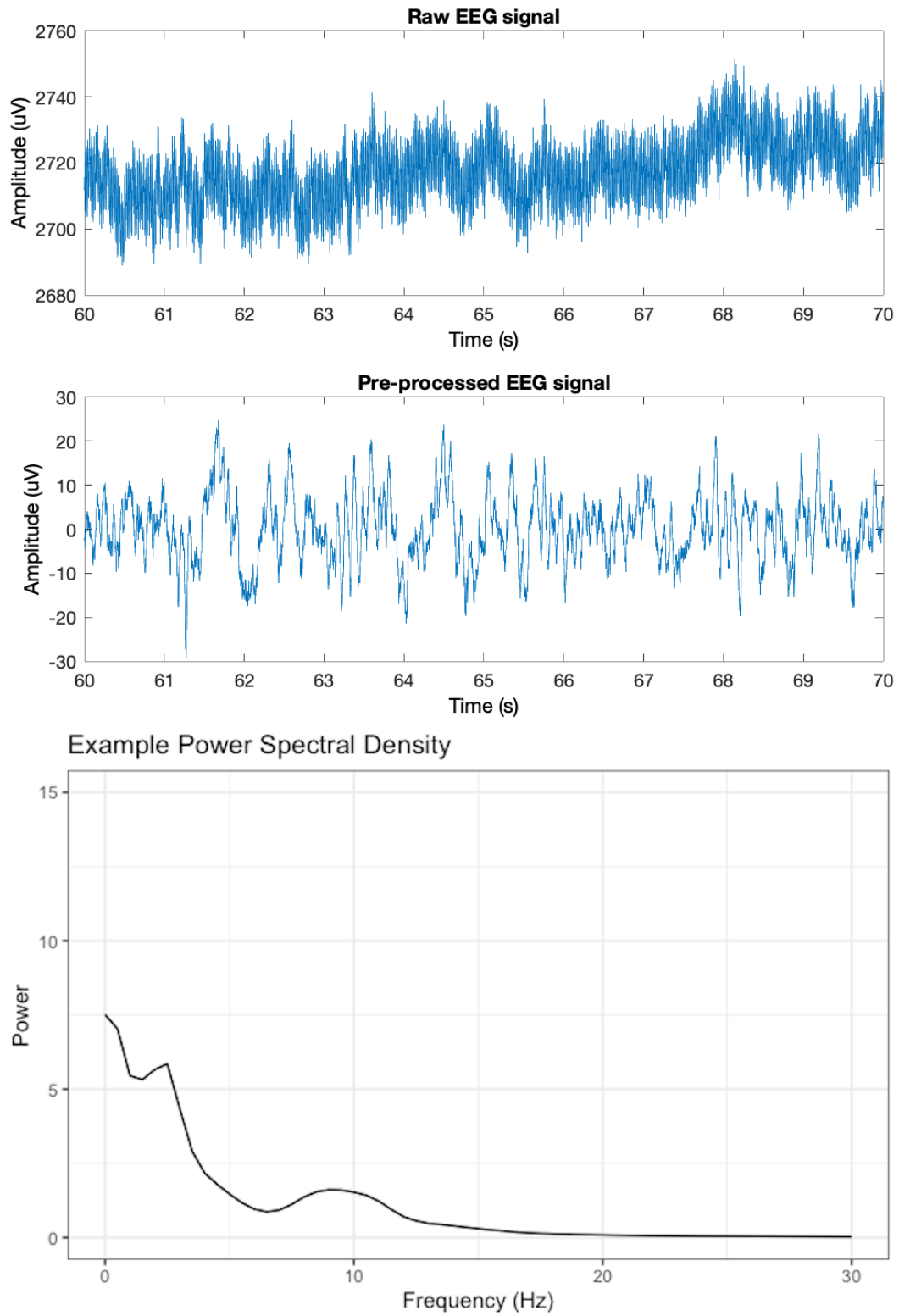


Figure 2: Examples of raw and pre-processed EEG data, and corresponding Power Spectral Density

	PC1	PC2	PC3
IAF	-0.2339	0.5657	-0.0179
SEF95	0.4128	0.0927	0.1925
Total Power	-0.3896	-0.1587	-0.1604
Delta (0-4 Hz)	0.5031	-0.0360	-0.1218
Theta (4-8 Hz)	-0.0973	-0.6179	0.2544
Alpha (8-12 Hz)	-0.4789	0.2204	-0.2257
Low Beta (12-16 Hz)	0.0885	0.4334	0.6317
Spectral Entropy	0.3512	0.1635	-0.6383
Proportion of Variance Explained	0.4126	0.2660	0.1275

Table 1: Component coefficients for PCA of frequency features. Coefficients with magnitude > 0.4 are in bold.

IAF = individual alpha frequency; SEF95 = 95% Spectral Edge Frequency. 4 Hz power bands are as a proportion of total power.

Component 1 explains 41% of the overall variance, and represents an average of 95% spectral edge frequency (SEF95) and fractional power in the delta range (0-4 Hz), contrasted with fractional power in the alpha range (8 - 12 Hz). Accordingly individuals with higher scores on this component would typically have a higher SEF95, a higher proportion of power at low frequencies between 0 and 4 Hz, contrasted with a lower proportion power between 8 and 12 Hz. This component can be thought of as describing the level of dispersion and delta power, contrasted with alpha power.

Component 2 explains 27% of the overall variance, and represents an average of Individual Alpha Frequency (IAF) and fractional power in the low beta range (12-16 Hz), contrasted with relative power in the theta range (4-8 Hz). Higher scores on this component identify individuals with higher frequency alpha peaks, higher relative beta power and lower relative theta power. Given the variability of IAF between 7-14 Hz, overlapping with the theta (4-8 Hz) and lower beta (12-16 Hz) bandwidths, this component therefore appears to mainly capture variability related to the frequency location of the alpha peak for each individual.

Component 3 explains 13% of the overall variance, and represents a contrast of relative power in the low beta range (12-16 Hz) with spectral entropy. This component captures the disparity between these variables — indicating that individuals in our dataset with higher relative power between 12-16Hz will tend to have a less complex PSD with lower spectral entropy, and vice versa.

4.2 Unsupervised clustering methods

Results from each clustering algorithm are presented below, before identifying integrated core clusters of individuals that are consistently clustered together across clustering methods. These core clusters are then described in terms of their average PSD, principal component scores, and original frequency feature values.

On the basis of the internal validation criteria introduced above, alongside the goals of model parsimony and interpretability for the final integrated clusters, we chose to implement a 5-cluster solution in each of the three individual clustering methods. Examining results from multiple internal validation indices, we identified support for a 5-cluster solution as the optimal choice for the k -means and GMM algorithms. For hierarchical clustering, we found marginal support for a 5-cluster or 7-cluster solution being optimal. For the sake of parsimony and inter-

Cluster label	1	2	3	4	5
<i>k</i> -means	18	15	13	10	3
HC	17	15	9	15	3
GMM	21	13	13	9	3
Core clusters	15	12	9	9	3

Table 2: Cluster membership for different clustering algorithms & core clusters, $K=5$

pretability in the final set of core clusters, we decided to implement a 5-cluster solution for all three algorithms to have a cohesive approach across the methods. This allowed us to identify the most consistent structure across results from the different algorithms, and maximise the number of individuals allocated to the integrated core clusters. Further details on selecting the number of clusters K for each method are provided in the Supplementary Materials.

Cluster stability metrics were calculated based on several resampling schemes including bootstrapping, subsetting and jittering, and indicated acceptable stability for our results. All three of our original clustering solutions demonstrated acceptable clustering stability. Mean Jaccard coefficients across the 5 clusters were: 0.65 for *k*-means; 0.72 for HC; 0.84 for GMM. We recognise that particularly for *k*-means and hierarchical clustering with Jaccard coefficients between 0.6 and 0.75, this indicates evidence of clustering pattern being present with some uncertainty about the exact allocations to specific clusters (Hennig, 2007). These indications of moderate uncertainty related to cluster stability for *k*-means and HC supported the value of our process to identify core clusters, in order to preserve stable well-separated groups of individuals who were consistently clustered together across the three algorithms.

Table 2 presents the number of individuals assigned to each cluster label for the three clustering algorithms, and for overlapping ‘core’ clusters. For each algorithm, cluster labels (1-5) have been assigned by decreasing cluster size except for HC clusters 3 and 4, for which labels were switched for the sake of clearer visual comparison across plots.

Figure 3 presents 2D scatter plots of clustering results for each clustering method, displayed according to each pairwise combination of the 3 retained principal components. 3D scatter plots are available in the Supplementary Materials.

4.3 Core clusters across methods

As discussed above, we aimed to retain the number of core clusters equal to the minimum K chosen across clustering methods, where each core cluster represents an intersection across a unique combination of cluster labels from each method, and cluster labels from each method were represented only once in the core clusters. In total we retained 5 core clusters which included a total of 48 individuals, and 11 individuals were not allocated.

Figure 4 presents 2D scatter plots of individuals by principal component scores coloured by core cluster membership after identification of overlapping clustering results between the algorithms, displayed by each combination of the 3 retained principal components. Individuals who were not included in any core cluster as they were not clustered consistently between methods ($n = 11$) have been removed from these plots.

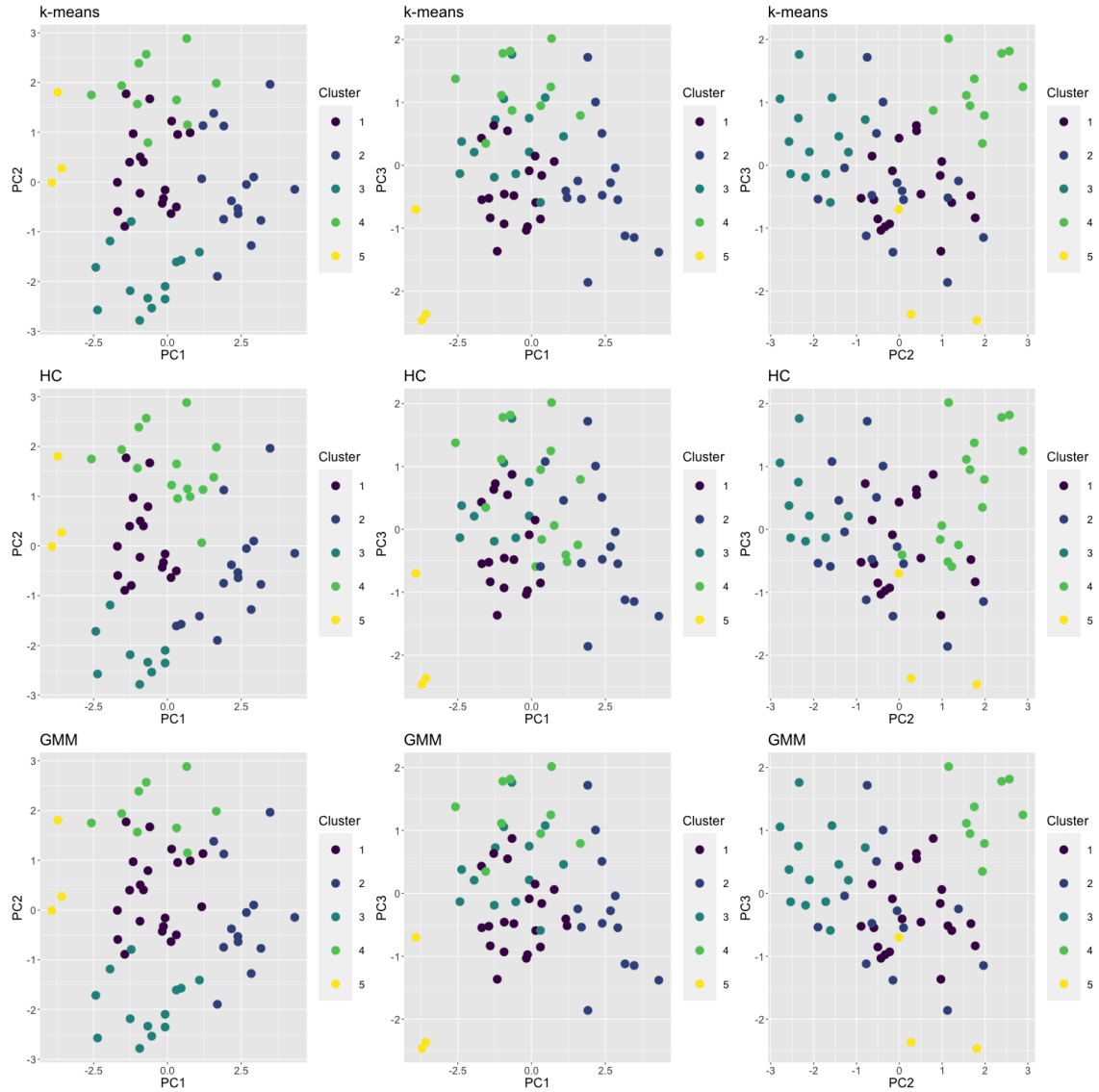


Figure 3: 2D scatter plots of individuals by component scores, coloured by cluster membership, $K = 5$.

Top row = k -means; middle row = HC; bottom row = GMM. Left column = PC1 v PC2; middle column = PC1 v PC3; right column = PC2 v PC3.

Core cluster	PC1	PC2	PC3
1	-0.7938 (0.6658)	0.1293 (0.8179)	-0.4220 (0.6231)
2	2.6522 (0.7516)	-0.2716 (1.0223)	-0.3466 (1.0275)
3	-1.1448 (0.9158)	-2.1971 (0.4879)	0.4344 (0.6506)
4	-0.3940 (1.3165)	1.9860 (0.5424)	1.2704 (0.5398)
5	-3.7381 (0.1681)	0.6892 (0.9771)	-1.8443 (0.9915)

Table 3: Core cluster profiles: Principal component scores. Mean (SD)

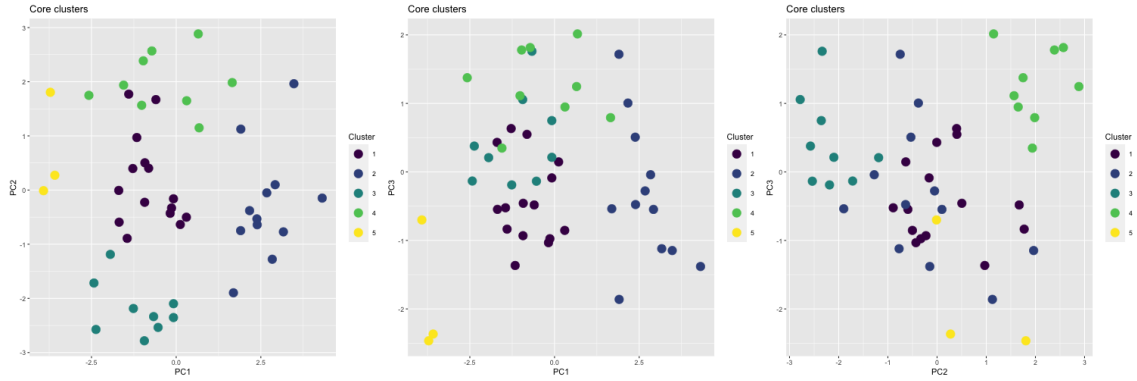


Figure 4: 2D scatter plots of individuals by component scores, coloured by core cluster membership.

Individuals not allocated to core clusters have been removed. Left panel = PC1 v PC2; middle panel = PC1 v PC3; right panel = PC2 v PC3.

4.3.1 Core clusters - principal component scores

Looking at the cluster locations in Figure 4 and the mean principal component scores for each core cluster in Table 3, we can identify the frequency characteristics that are expected to differentiate the core clusters.

Cluster 1 was the most central, with small scores on every PC, so it was not obvious from the PC scores how frequency characteristics for this cluster were expected to differ from the others. Clusters 2, 3 and 5 had high magnitude scores on PC1, and were expected to differ by their level of frequency dispersion and low-frequency power, contrasted with the magnitude of the individual alpha peak. Clusters 3 and 4 had high magnitude scores on PC2, differentiated by the frequency location of the individual alpha peak, and by higher frequency activity in the range 12-16Hz. Clusters 4 and 5 had high magnitude scores on PC3, differing by the level of contrast between relative power in the range 12-16 Hz and spectral entropy.

We can further investigate these patterns in cluster principal component scores by looking at mean scores on the original frequency features (Table 4), and mean power spectral density plots (Figure 5) for each core cluster.

4.3.2 Core clusters - EEG frequency features

From Table 4 and Figure 5, and as expected from the principal component scores, cluster 1 (n=15) sits close to the middle of the range for each of the frequency features calculated. It has a well defined alpha peak and a moderate IAF (8.77) relative to other clusters.

We can see that cluster 2 (n=12) is characterised by noticeably lower theta and alpha activity in the range 4-12 Hz, high power at low frequencies, high dispersion, and low total power. Members of this cluster have high relative power at low frequencies compared to other clusters, with the lowest IAF (7.33) and more than half of their relative power between 0-4 Hz (0.52), contrasted with having the lowest relative alpha power between 8-12 Hz (0.12). Cluster 2 is also characterised by high dispersion with the highest SEF95 (32.79) and spectral entropy (0.71) indicating a complex frequency distribution with power distributed across a wide range. This cluster also has the lowest total power (58.03).

Cluster 3 (n=9) has high power at low frequencies, with low dispersion and a pronounced alpha peak. With a low IAF (7.50) and the highest relative power in the range 4-8 Hz (0.32),

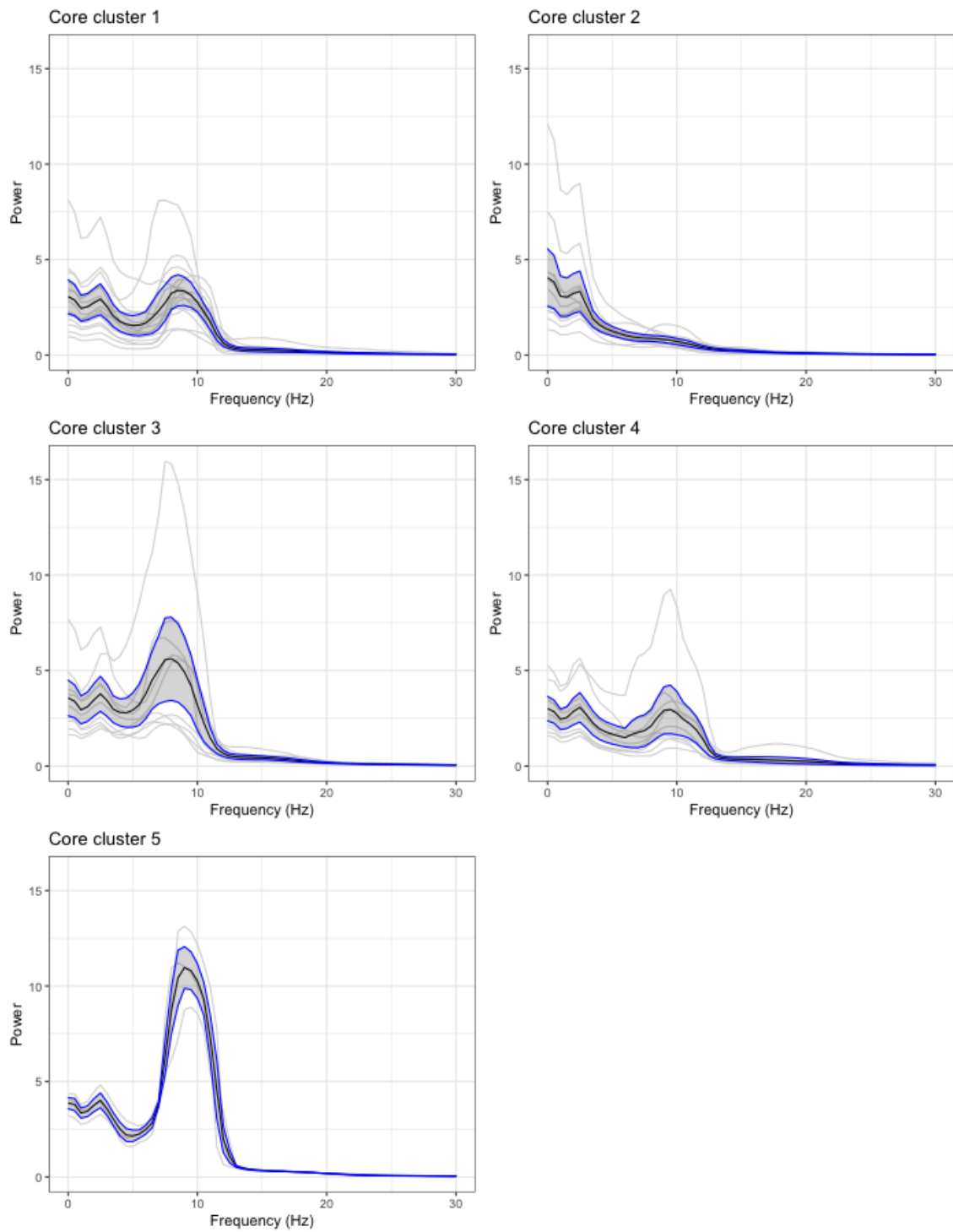


Figure 5: Mean PSDs for each core cluster. Shaded regions represent standard deviations. Light grey lines indicate PSDs for individuals allocated to each cluster.

we can identify high theta activity, and a pronounced alpha peak at the low range for IAF. A low SEF95 (19.39) further indicates that most of the power is concentrated at low frequencies for this cluster. Cluster 3 also has low spectral entropy (0.63) and the equal lowest relative power between 12-16 Hz (0.04), indicating a less complex frequency distribution that is concentrated at lower frequencies.

Cluster 4 ($n = 9$) has a high frequency alpha peak with the highest IAF (9.50) and the lowest relative power in the range 4-8 Hz (0.16), indicating low theta activity. It also distinguished by having the highest relative power in the range 12-16 Hz (0.09), demonstrating higher activity than other clusters in the upper alpha - lower beta range.

Cluster 5, the smallest cluster ($n = 3$), has high scores on principal component 1 and is characterised by a high magnitude, high frequency alpha peak, low dispersion, and high total power. It has a high IAF (9.00) and the highest activity in the alpha range 8-12 Hz (0.57), with over half its relative power occurring in this range. This cluster also has the highest total power with more than double the mean total power across all participants (208.38), the lowest relative delta power of any cluster in the range 0-4 Hz (0.18) and the lowest SEF95 (12.67), indicating a distinct and high degree of power concentrated in the alpha range.

4.4 Comparing cognitive function & mental health measures across clusters

Descriptive statistics for EEG frequency features in each cluster are presented in Table 4. Descriptive statistics for mental health and cognitive function measures in each core cluster are presented in Table 5. Figure 6 presents column graphs showing Z-scores based on the global mean and standard deviation for each EEG feature (top panel) and each health and behavioural measure (bottom panel) across the 5 clusters, allowing a comparison of the EEG and health characteristics in each group. A Z-score measures the number of standard deviations by which the value of a raw datapoint is above or below the mean value. In Figure 6, using Z-scores to standardise health measures and EEG frequency features aids comparison among clusters for variables which have different units or magnitudes. Violin plots showing the distribution of scores for health measures in each cluster are available in Supplementary Figure 7.

Bayesian regression models were applied to investigate detailed patterns of difference in cognitive function and health measures between clusters. From sensitivity analyses, the results were not substantially different according to the prior used. Figure 7 presents posterior probabilities for these outcome measures being different between each pairwise combination of clusters. These posterior probabilities are approximated from the models results by calculating the proportion of Markov Chain Monte Carlo iterations in which the condition is true. In Figure 7, red cells indicate greater than 90% probability that the corresponding measure was higher in cluster A than cluster B ($A > B$). Blue cells indicate greater than 90% probability that the measure was lower in cluster A than cluster B ($A < B$). Expected marginal means from these models for contrasts between clusters are presented in the Supplementary Materials (Supplementary Table 2), providing information about the magnitude of the differences highlighted in Figure 7. Bold typeface identifies cells with greater than 80% or less than 20% probability.

For suicidality, cluster 3 appeared to have a higher score on the SIDAS scale compared to every other cluster. However, this signal disappeared when one outlying subject in cluster 3 was removed from the analysis, who had the maximum possible SIDAS score, and was also an outlier on reaction time for the CogState detection task (DET - RT). Therefore, findings on between-cluster differences for suicidality (SIDAS) and reaction time on the detection task (DET - RT) should be interpreted with caution, as they are sensitive to the contribution of this one individual. Given that this is a population sample and instances of extreme responding and individuals lying in the tails of distributions are expected occurrences, we decided to include this individual for analyses in the body of this paper.

4.4.1 Cluster Profiles

Looking at the differences in EEG features, health and cognitive function measures across clusters, we could identify tentative EEG and health profiles that may be meaningful in the context of risk prediction for mental health and cognitive development based on characteristic resting state EEG profiles in adolescents. These EEG-based clusters were substantively different in terms of their self-reported psychological distress, wellbeing, sleep quality, and several cognitive function measures. Figures 5-6 and Tables 4-5 show how these groups exhibit meaningful differences in their resting state EEG profiles alongside health measures of interest.

Cluster 1 was the largest cluster, and sat close to the global mean for this sample on most EEG features (Figure 6). From our Bayesian regression models (Figure 7), we identified a high ($> 90\%$) probability that members of cluster 1 had lower distress than cluster 2 (K10 and SPHERE-12), lower distress and suicidality than cluster 3 (SPHERE-12 and SIDAS), and slower reactions on all three reaction time tasks than clusters 2 and 4.

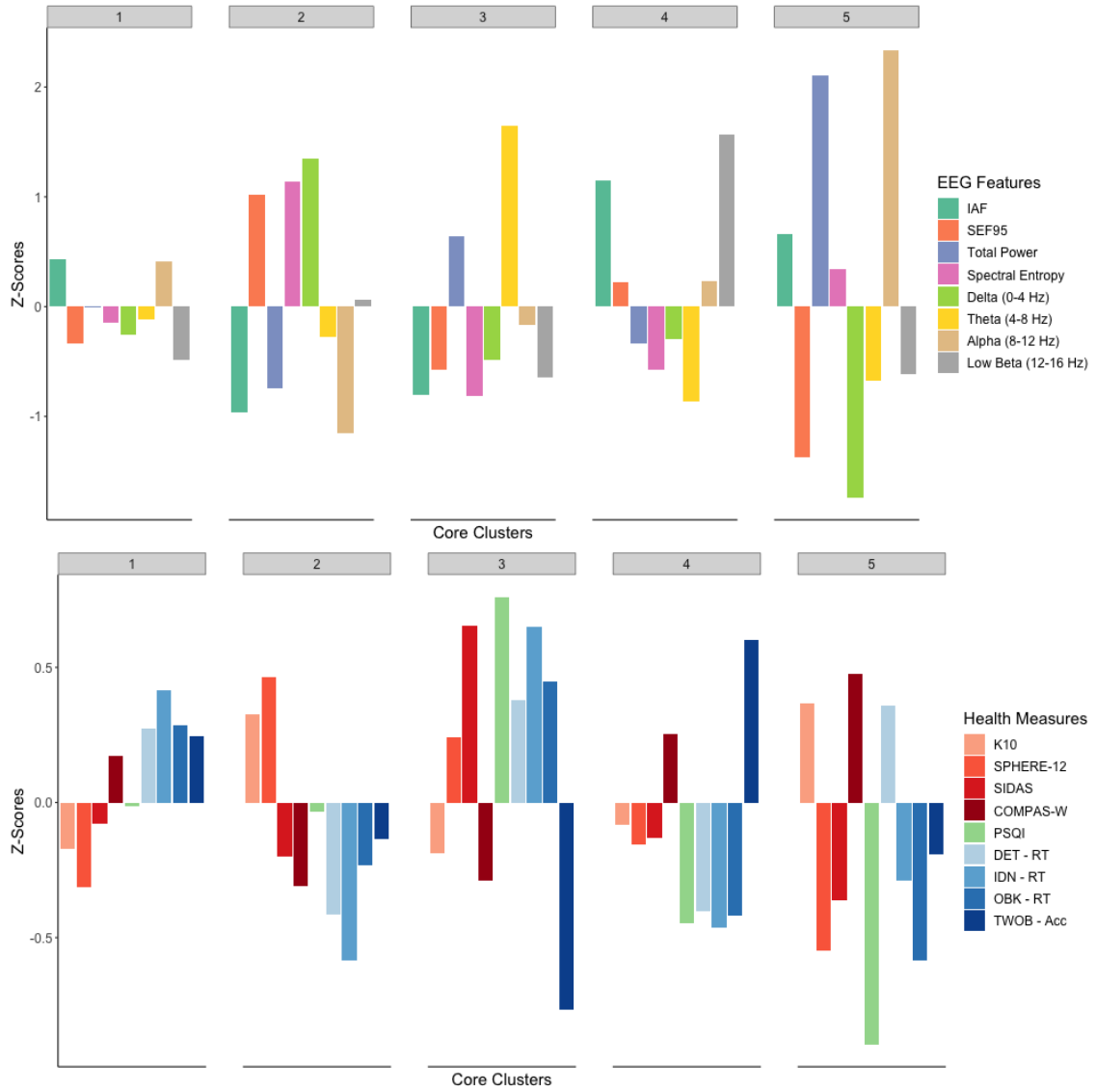


Figure 6: Bar plots showing Z-scores for each EEG feature (top panel) and each health and behavioural (cognitive) measure (bottom panel) across core clusters.

IAF = individual alpha frequency; SEF95 = 95% Spectral Edge Frequency. 4 Hz power bands are as a proportion of total power. K10 = Kessler Psychological Distress Scale; SPHERE12 = SPHERE-12 Screening Questionnaire; SIDAS = Suicidal Ideation Attributes Scale; COMPAS-W = COMPAS-W Wellbeing Scale; PSQI = Pittsburgh Sleep Quality Index; DET - RT = CogState Detection Task Reaction Time; IDN - RT = CogState Identification Task Reaction Time; OBK - RT = CogState One Back Task Reaction Time; TWOB - Acc = CogState Two Back Task Accuracy Score.

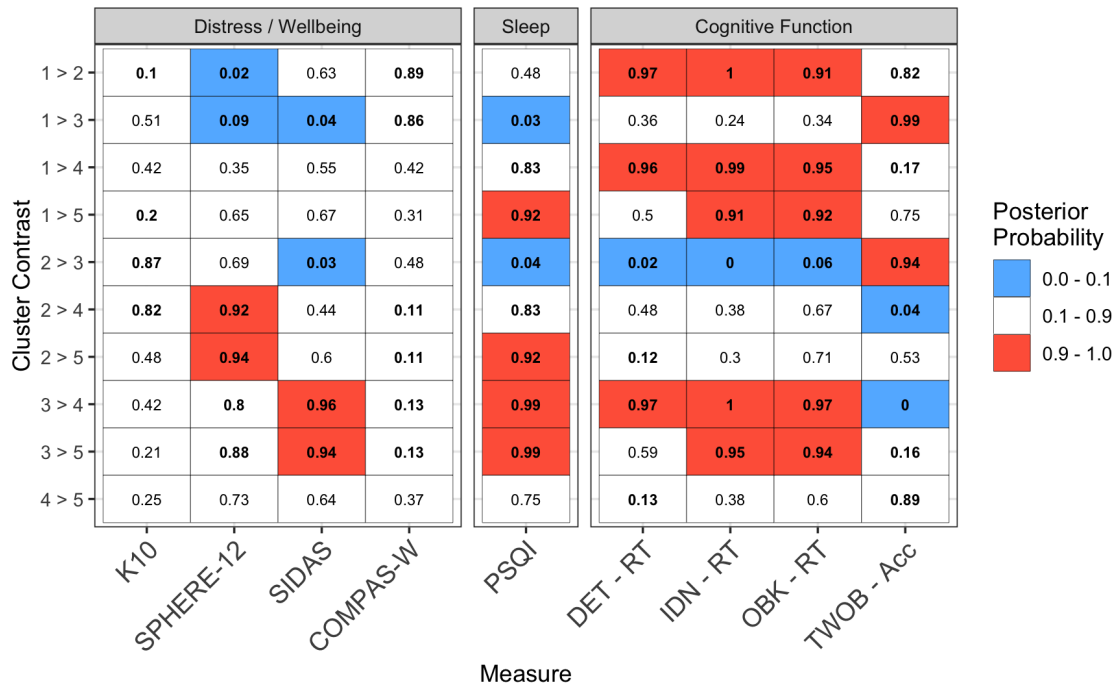


Figure 7: Posterior probabilities for differences in health measures between clusters, calculated from Bayesian regression models. Red cells = greater than 90% probability that measure is higher in cluster A than cluster B ($A > B$). Blue cells = greater than 90% probability that measure is higher in cluster B than cluster A ($B > A$). Bold typeface identifies cells with greater than 80% or less than 20% probability.

K10 = Kessler Psychological Distress Scale; SPHERE12 = SPHERE-12 Screening Questionnaire; SIDAS = Suicidal Ideation Attributes Scale; COMPAS-W = COMPAS-W Wellbeing Scale; PSQI = Pittsburgh Sleep Quality Index; DET - RT = CogState Detection Task Reaction Time; IDN - RT = CogState Identification Task Reaction Time; OBK - RT = CogState One Back Task Reaction Time; TWOB - Acc = CogState Two Back Task Accuracy Score.

	Overall ($n = 59$)	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Un-allocated
IAF	8.34 (1.00)	8.77 (0.53)	7.33 (0.78)	7.50 (0.50)	9.50 (0.50)	9.00 (0.50)	8.41 (0.89)
SEF95	23.93 (7.83)	21.37 (4.03)	32.79 (9.65)	19.39 (3.72)	26.11 (6.45)	12.67 (0.58)	22.77 (4.38)
Total Power	95.50 (48.68)	97.08 (32.55)	58.03 (23.50)	130.96 (45.09)	79.57 (45.64)	208.38 (66.16)	87.46 (23.31)
Delta (0-4 Hz)	0.38 (0.10)	0.34 (0.05)	0.52 (0.06)	0.32 (0.07)	0.34 (0.07)	0.18 (0.03)	0.44 (0.05)
Theta (4-8 Hz)	0.22 (0.06)	0.21 (0.03)	0.20 (0.04)	0.32 (0.03)	0.16 (0.03)	0.18 (0.06)	0.20 (0.04)
Alpha (8-12 Hz)	0.26 (0.12)	0.33 (0.06)	0.12 (0.03)	0.25 (0.07)	0.30 (0.09)	0.57 (0.10)	0.21 (0.05)
Low Beta (12-16 Hz)	0.05 (0.02)	0.04 (0.01)	0.05 (0.02)	0.04 (0.01)	0.09 (0.02)	0.04 (0.02)	0.05 (0.01)
Spectral Entropy	0.66 (0.04)	0.66 (0.02)	0.71 (0.05)	0.63 (0.02)	0.64 (0.03)	0.68 (0.02)	0.66 (0.03)

Table 4: Descriptive statistics for EEG features. The first column presents statistics for all participants ($n = 59$), and the remaining columns present statistics for individuals allocated to each of the final core clusters. IAF = individual alpha frequency; SEF95 = 95% Spectral Edge Frequency. 4 Hz power bands are as a proportion of total power.

Measure	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
N	59	15	12	9	9	3
Age (weeks), mean (sd)	657.11 (16.48)	655.71 (15.49)	659.19 (17.26)	658.81 (16.52)	659.05 (19.12)	653.14 (23.94)
Female, n (%)	28 (47.46)	8 (53.33)	4 (33.33)	6 (66.66)	3 (33.33)	1 (33.33)
K10, mean (sd)	15.66 (4.63)	14.40 (4.29)	16.50 (4.40)	14.33 (2.83)	14.78 (5.54)	16.67 (2.31)
SPHERE12, mean (sd)	4.02 (3.63)	2.80 (3.10)	5.42 (3.26)	4.67 (3.32)	3.33 (4.00)	2.00 (1.73)
SIDAS, mean (sd)	3.17 (8.44)	2.47 (6.36)	1.42 (4.01)	8.89 (17.00)	2.00 (4.69)	0.00 (0.00)
COMPAS-W, mean (sd)	102.17 (10.95)	105.00 (10.11)	99.67 (11.98)	99.89 (12.94)	105.89 (10.35)	108.33 (6.43)
PSQI, mean (sd)	4.95 (2.81)	4.80 (2.27)	4.75 (1.66)	6.67 (2.92)	3.75 (2.43)	2.67 (1.15)
DET - RT, mean (sd)	2.54 (0.09)	2.57 (0.08)	2.50 (0.03)	2.58 (0.19)	2.50 (0.03)	2.58 (0.01)
IDN - RT, mean (sd)	2.73 (0.07)	2.76 (0.08)	2.69 (0.04)	2.78 (0.08)	2.70 (0.05)	2.71 (0.03)
OBK - RT, mean (sd)	2.89 (0.09)	2.92 (0.09)	2.87 (0.11)	2.94 (0.11)	2.85 (0.07)	2.83 (0.05)
TWOB - Acc, mean (sd)	1.25 (0.19)	1.29 (0.18)	1.21 (0.14)	1.09 (0.27)	1.36 (0.15)	1.20 (0.12)

Table 5: Descriptive statistics for mental health and cognitive function measures. The first column presents statistics for all participants ($n = 59$), and the remaining columns present statistics for individuals allocated to each of the final core clusters.

K10 = Kessler Psychological Distress Scale; SPHERE12 = SPHERE-12 Screening Questionnaire; SIDAS = Suicidal Ideation Attributes Scale; COMPAS-W = COMPAS-W Wellbeing Scale; PSQI = Pittsburgh Sleep Quality Index; DET - RT = CogState Detection Task Reaction Time; IDN - RT = CogState Identification Task Reaction Time; OBK - RT = CogState One Back Task Reaction Time; TWOB - Acc = CogState Two Back Task Accuracy Score.

Cluster 2 was characterised by low relative theta and alpha power (4-12 Hz), a low frequency alpha peak, and low total power. We found a high probability that members of cluster 2 were more distressed than clusters 1, 4 and 5 (SPHERE-12), and had faster reaction times on all three reaction time tasks than clusters 1 and 3.

Cluster 3 was characterised by a low frequency alpha peak, high relative theta power, and a concentration of power at lower frequencies. Members of cluster 3 had the most concerning sleep and cognitive function profile, with substantially poorer sleep quality than members of all other clusters, slower reaction times than clusters 2, 4 and 5 on reaction time measures, and poorer accuracy on the Two-Back task than clusters 1, 2 and 4. This cluster appeared to have higher suicidality (SIDAS) than other clusters, though as discussed above this signal disappeared when removing one outlying participant, so should be interpreted here with caution. The signals for differences in psychological distress and wellbeing measures were not as strong, but there was moderate evidence (>80% probability) to suggest elevated distress (SPHERE-12) and lower self-reported wellbeing (COMPAS-W) in this group relative to clusters 1, 4 and 5. This could suggest that the characteristic resting state EEG profile of cluster 3 is associated with elevated risk of poor outcomes in the domains of mental health, sleep quality and cognitive function in early adolescence.

Cluster 4 was characterised by a high frequency alpha peak, low relative theta power (4-8 Hz), and high relative power in the upper alpha/lower beta range (12-16 Hz). We found moderate to high probability that cluster 4 had lower distress (SPHERE-12) and higher wellbeing (COMPAS-W) than individuals in clusters 2 and 3. Cluster 4 also had moderate (> 80%) probability of reporting better sleep quality than clusters 1, 2 and 3. Members of cluster 4 had faster reaction times than clusters 1 and 3 on all three reaction time measures, as well as having higher accuracy than clusters 2 and 3 on the Two-Back task. This could suggest that the EEG characteristics of cluster 4 are associated with improved outcomes in the domains of mental health, sleep quality and cognitive function.

Cluster 5, having only 3 members, shows more variability across these measures. It is characterised by a uniquely high magnitude alpha peak, low dispersion, and high total power. Particular caution should be taken for interpreting results about differences in health and cognitive function for this group, given the very small number of individuals represented. Our models identified high (>90%) probability that this cluster had better sleep quality than clusters 1, 2 and 3, and faster reaction times than clusters 1 and 3 on two of the reaction time tasks.

It is important to note that the relationships between EEG characteristics and health measures were not directly modeled in a regression framework here, and instead the patterns of difference in health measures here are related to the particular combinations of EEG characteristics that distinguish these clusters. The relationships found between individual EEG features and health measures might be different if modelling them together as predictor and response variables respectively, but that was not the goal of the present analysis. These are interesting initial findings that warrant further research and demonstrate the utility of the data-driven method implemented here to identify data-driven clusters with distinct neurophysiology characteristics.

5 Discussion

In this work we have addressed several main objectives: to develop and validate a flexible analysis pipeline for identifying data-driven subgroups of adolescents based on resting state EEG data; to interrogate the patterns of EEG characteristics that distinguish these clusters; to investigate patterns of difference in mental health and cognitive function measures between

clusters; and to identify preliminary risk and protective profiles of health and functional measures associated with EEG-based clusters.

With the analysis pipeline developed here, using data-driven unsupervised clustering we were able to identify well-separated subgroups of 12-year-olds on the basis of frequency content from resting state, eyes closed EEG data. While these results are based on measurements from a single EEG electrode, this pipeline has been developed with flexibility and scalability in mind, and can easily accommodate a larger and/or different set of EEG features calculated across multiple electrodes. The implementation of a Bayesian approach for comparison of health measures between clusters enabled us to make probabilistic statements about between-group differences with quantified uncertainty. This has attractive benefits for application in a clinical setting, where intuitive probabilistic interpretation of risk is desirable for credibility and interpretation of the models for clinicians, and communication of probabilities to health-care consumers.

The dimensionality reduction stage we have implemented here with PCA enables this analysis framework to accommodate working with data from a small number of individuals, and a large number of features calculated from EEG data. In future applications this dimensionality reduction stage will also allow us to look at EEG characteristics across a number of electrodes, examining spatial variation in activity across different brain regions while avoiding some of the challenges arising when analysing high dimensional data. Working in a dimension-reduced space also improves model parsimony and interpretability for unsupervised clustering results.

Particularly influential EEG variables that we identified as distinguishing individuals included: the level of dispersion of power across frequencies, calculated with the 95% spectral edge frequency; relative power in 4 Hz width frequency bands (delta, theta, alpha and beta) up to 16 Hz; and the location of the individual's alpha peak frequency. We do not claim that in every case an EEG-based cluster will be associated with pronounced differences in mental health or cognitive function, as there is scope for biology and neurophysiology to vary between individuals without necessarily being associated with substantial health or behavioural differences. However, we have found some patterns of association between characteristic profiles of resting state EEG frequency content and health outcomes which warrant further investigation. For instance the indications of a potential at-risk profile associated with the EEG characteristics of cluster 3, or a protective profile associated with cluster 4, support the utility of these methods and suggest avenues for future investigation. It is also interesting that these two clusters which have contrasting and distinct health and cognitive function profiles also exhibit contrasts on several key EEG features. For example compared to other clusters, cluster 3 with higher distress and poorer cognitive testing results exhibits a low frequency alpha peak (IAF), high relative theta (4-12 Hz) power, and low relative power in the upper alpha - lower beta range (12-16 Hz). Contrastingly, cluster 4 with lower distress and better cognitive function has a high IAF, low relative theta power, and high relative beta power. These observations are only descriptive at this stage, but future work will allow us to further investigate whether these combined patterns of EEG frequency features could be useful as indicative biomarkers for risk for mental health and cognitive function in young people.

Given the small sample size in the present study, it is important not to over-interpret these results or to assume that they are able to be broadly generalised. Overall, the strength of the signals for between-cluster differences was weaker for self-report distress and wellbeing measures compared to cognitive testing tasks — which is to be expected, given the known ability of cognitive tests to have better inter-subject reliability and power to detect granular differences between individuals than self-report scales (Rabbitt and Abson, 1990; Williams et al., 2017; Herreen and Zajac, 2018). However, the large number of substantial signals identified here relating to differences in mental health and cognitive function between EEG-based clusters supports the utility of the methodological pipeline we have implemented.

The present findings demonstrate the potential for this approach to offer new and important insights around EEG profiles and biomarkers that may be useful in a clinical context. This modelling framework enables insight into detailed patterns of difference between distinctive profiles of resting state EEG activity, and associations with sleep quality and cognitive function measures.

We have foregrounded empirical assessment of brain data, creating opportunities for novel insights into the relationships between neurophysiology, mental health and cognitive function. Our approach offers findings that would not be accessible from a traditional case-control approach of comparing neurophysiology between groups of individuals based on clinical diagnostic categories. Case-control approaches to researching the patterns between brain activity and mental health are especially limited for research in adolescence, when patterns of dimensional symptoms are associated with pluripotent risk for different mental disorder outcomes and often overlap across traditional diagnostic categories (McGorry and Mei, 2018). Considering the pluripotential nature of mental health problems as they are understood to develop and emerge during adolescence, it was beneficial to use the empirical data-driven approach that we implemented here, studying psychopathology measures on a continuum rather than between diagnostic categories. Other literature has demonstrated that this approach offers reduced constraints from separating and contrasting individuals according to clinical diagnostic labels. Stepping beyond the case-control paradigm enables novel understanding of how different characteristics of neurophysiology and mental health outcomes relate to each other. This is advantageous for novel hypothesis generation and improving upon known issues with traditional diagnostic categories including comorbidity and difficulties with differential diagnosis, especially during adolescence (Raballo and Poletti, 2020; Kotov et al., 2017).

The present study contributes to a growing body of research has demonstrated important interconnections between cognitive function, brain characteristics, and mental health throughout adolescent development. There is some evidence that greater cognitive ability in childhood is associated with decreased risk of internalising mental health symptoms through adolescence, and can attenuate the impact of external stressors on risk of mental health problems (Weeks et al., 2014). Higher childhood cognitive ability is also associated with fewer symptoms of anxiety and depression in adult females and increases risk for potential alcohol abuse in adulthood (Hatch et al., 2007). Some recent findings from longitudinal data show some evidence of a causal relationship between depressive symptoms in adolescence predicting poorer cognitive ability in adulthood. This body of literature suggests important interrelationships between cognitive function, psychopathology and brain health in adolescence (Steinberg, 2005; Moberget et al., 2019). Our work highlights that to develop a stronger understanding of these interrelationships, cognitive function and psychopathology symptoms are both important factors to measure and study together in the context of the developing adolescent brain.

An alternative to the statistical methods employed here for cluster comparison could be to take a frequentist approach, testing for between-cluster differences in mental health and cognitive function measures by applying analyses of variance or Kruskal-Wallis tests, with post-hoc testing using Tukey’s Honestly Significant Difference or similar (Kruskal and Wallis, 1952; Tukey, 1949). When we implemented these methods, we found that the post-hoc analyses were not definitive, and that these methods were underpowered when correcting for multiple comparisons given the small sample size used in the present study. The Bayesian approach used here avoids these barriers and provides more inferentially informative findings, giving direct posterior probabilities for differences in mental health and cognitive function measures between clusters (Mengersen et al., 2016).

There are several limitations to the present study. The process of manual EEG feature selection, followed by dimensionality reduction with PCA, faces an obstacle in the number of

features we can analyse across a limited number of electrodes, given the small sample sizes typically available in EEG studies. This is because like other popular dimension reduction methods, PCA is known to perform poorly and encounter difficulties as the number of features approaches or exceeds the number of observations — other dimensionality reduction approaches would need to be considered for situations where the number of features of interest exceeds the number of participants for whom data is available. Here we chose PCA as an appropriate dimension reduction method given the absence of nonlinear relationships between the input variables. Depending on the presence of nonlinearity and the other goals of the analysis, other dimension reduction techniques could be considered such as an autoencoder or factor analysis, among other methods. Another limitation of this approach is that it requires manual selection of known or commonly used features calculated from EEG data, which may limit the ability of this approach to find novel patterns or features in EEG signals that are associated with important differences between individuals and subgroups. To improve on these shortcomings, planned extensions to this work include longitudinal analyses of data from a larger number of individuals, and we will investigate alternative approaches to dimension reduction such as functional PCA or manifold learning which may have more flexibility to capture rich information from EEG time series while avoiding some of the challenges around high dimensionality and manual feature selection.

Since mental health and cognitive measures were collected in the same session as EEG data in this cross-sectional sample, it is possible that patterns in EEG and health measures may be affected by situational factors such as recent sleep quality or transient stressors. This is a drawback of working with a small sample size of cross-sectional data. Once more longitudinal data becomes available, further analyses will allow us to determine whether there is a robust association between these EEG characteristics and patterns of risk for mental health and cognitive function, and have more confidence about the strength of these associations cross-sectionally and over time.

The simple approach implemented here to combined multiple sets of results from different clustering algorithms meant that we could not make inferences about $n = 11$ individuals who are not allocated to final clusters due to being inconsistently clustered across the algorithms). This was suitable given the goal of this work to identify distinct EEG-based subgroups without aiming to support inferences about every individual participant based on their resting state EEG data. An alternative approach in future would be to combine the results using Bayesian model averaging, allowing us to make probabilistic statements about cluster membership and relating this to probability of differences between clusters, taking into account the strength of allocations to the clusters.

In the situation that the appropriate type of clustering structure is known *a priori* for a given dataset, combining results from multiple clustering algorithms could be detrimental by ‘diluting’ the appropriate solution with less appropriate clustering types. However, in this case we did not have any *a priori* reason to believe that there was one particular sort of clustering structure that would be most appropriate for this dataset. Accordingly, combining clustering results from multiple models allowed us to identify classifications that were robust to model choice, and to reserve judgement on those individuals whose classification was sensitive to model choice.

While we chose to use these three clustering algorithms on the basis that they are widely used and provide coverage of common attributes of existing clustering algorithms, they are not the only reasonable choices in this setting. Many other clustering algorithms could be considered for similar work in future, such as fuzzy C-means to accommodate ambiguous allocation in between clusters (Bezdek et al., 1984), or Deep Embedded Clustering with Data Augmentation (DEC-DA), which is well suited to small datasets with high dimensionality (Guo et al., 2018).

In this study the health characteristics considered were measured using dimensional self-report and testing scores rather than diagnostic categories, in order to explore more granular patterns of variation in continuous variables for health. The clinical utility of these findings will be dependent on replication of the findings here, including the application of EEG in clinical settings and assessing the relationships between cluster allocation and clinical diagnostic categories. Accordingly, future extensions to this work will need to undertake further sensitivity analyses, and to demonstrate external validity and replicability of these findings in clinical settings and other population samples.

Extending this approach to include longitudinal health outcomes will allow us to explore if and how these distinctive patterns of resting state EEG activity between clusters at age 12 can be useful to support longitudinal prediction of cognitive function and mental health outcomes at older ages. Our findings to date suggest potential utility to be explored in developing risk profiles that could be used in a clinical context of risk prediction and early intervention. For instance, when analysing longitudinal data we could assess the relationships between EEG-based cluster allocation at age 12 and the incidence of adverse health or cognitive outcomes at later ages. This could enable the development of health risk prediction tools to support clinicians in tailored care planning with the assistance of insights from EEG measurements. More work will be needed to investigate whether the distinctive EEG characteristics of these clusters are useful as an indicator of risk for cognitive development and mental health, both cross-sectionally and over time through adolescent development.

6 Author Contributions

Owen Forbes: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft. **Paul E. Schwenn:** Conceptualization, Data curation, Project administration, Methodology, Investigation, Supervision, Writing - Reviewing and Editing. **Paul Pao-Yen Wu:** Conceptualization, Methodology, Supervision, Writing - Reviewing and Editing. **Edgar Santos-Fernandez:** Conceptualization, Methodology, Supervision, Writing - Reviewing and Editing. **Hong-Bo Xie:** Conceptualization, Methodology, Supervision, Writing - Reviewing and Editing. **Jim Lagopoulos:** Funding acquisition, Data curation, Project administration, Methodology, Investigation, Writing - Reviewing and Editing. **Larisa T. McLoughlin:** Data curation, Investigation, Writing - Reviewing and Editing. **Dashiell D. Sacks:** Data curation, Investigation, Writing - Reviewing and Editing. **Kerrie Mengersen:** Conceptualization, Methodology, Supervision, Writing - Reviewing and Editing. **Daniel F. Hermens:** Funding acquisition, Data curation, Project administration, Methodology, Investigation, Supervision, Writing - Reviewing and Editing.

7 Acknowledgements

We thank the Longitudinal Adolescent Brain Study (LABS) participants and their caregivers.

The Longitudinal Adolescent Brain Study is supported by a grant from the Prioritising Mental Health Initiative, Australian Commonwealth Government.

This research was supported by the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (project number CE140100049) and funded in part by the Australian Government. <https://acems.org.au/>

O. Forbes is supported by PhD scholarships from the Queensland University of Technology, and the International Biometrics Society - Australasian Region.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

8 Code & Data Availability Statement

All code written in support of this publication is publicly available at https://github.com/of2/EEG_clustering_public

The conditions of the ethics approval do not permit public archiving of anonymised study data. Datasets from the Longitudinal Adolescent Brain Study are available on request from the data custodians at the Thompson Institute. Access will be granted to named individuals in accordance with ethical procedures governing the reuse of clinical data, including completion of a formal data sharing agreement. Contact LABSscmnti(at)usc.edu.au - more details at <https://www.usc.edu.au/thompson-institute/research/youth-mentalhealth/longitudinal-adolescent-brain-study/contact-labs>

9 Competing Interests Statement

The authors have no competing interests to declare.

10 References

References

- Acharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K.-H., and Suri, J. S. (2012). Automated diagnosis of epileptic EEG using entropies. *Biomedical Signal Processing and Control*, 7(4):401–408.
- Aggarwal, C. C. and Reddy, C. K. (2014). *Data clustering: Algorithms and applications*. Chapman & Hall/CRC Data mining and Knowledge Discovery Series. London.
- Alizadeh, V. and Dehzangi, O. (2016). The impact of secondary tasks on drivers during naturalistic driving: Analysis of EEG dynamics. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2493–2499. IEEE.
- Arns, M., Conners, C. K., and Kraemer, H. C. (2013). A decade of EEG theta/beta ratio research in ADHD: a meta-analysis. *Journal of attention disorders*, 17(5):374–383.
- Babadi, B. and Brown, E. N. (2014). A review of multitaper spectral analysis. *IEEE Transactions on Biomedical Engineering*, 61(5):1555–1564.
- Barry, R. J., De Blasio, F. M., and Karamacoska, D. (2019). Data-driven derivation of natural EEG frequency components: An optimised example assessing resting EEG in healthy ageing. *Journal of neuroscience methods*, 321:1–11.
- Baskaran, A., Milev, R., and McIntyre, R. S. (2012). The neurobiology of the EEG biomarker as a predictor of treatment response in depression. *Neuropharmacology*, 63(4):507–513.
- Beaudequin, D., Schwenn, P., McLoughlin, L. T., Parker, M. J., Broadhouse, K., Simcock, G., Boyes, A., Kannis-Dymand, L., Wood, A., and Lagopoulos, J. (2020). Using measures of intrinsic homeostasis and extrinsic modulation to evaluate mental health in adolescents: Preliminary results from the longitudinal adolescent brain study (LABS). *Psychiatry research*, 285:112848.
- Bezdek, J. C., Ehrlich, R., and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203.
- Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., and Mitra, P. P. (2010). Chronux: a platform for analyzing neural signals. *Journal of neuroscience methods*, 192(1):146–151.
- Bosl, W. J., Tager-Flusberg, H., and Nelson, C. A. (2018). EEG analytics for early detection of autism spectrum disorder: a data-driven approach. *Scientific reports*, 8(1):1–20.
- Buyssse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213.
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1):395–411.
- Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.
- Campisi, P. and La Rocca, D. (2014). Brain waves for automatic biometric-based user recognition. *IEEE transactions on information forensics and security*, 9(5):782–800.
- Catarino, A., Churches, O., Baron-Cohen, S., Andrade, A., and Ring, H. (2011). Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. *Clinical neurophysiology*, 122(12):2375–2383.

- Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., and Jung, T.-P. (2019). Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings. *IEEE Transactions on Biomedical Engineering*, 67(4):1114–1121.
- Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.
- Craddock, K. E., Zhou, X., Liu, S., Gochman, P., Dickinson, D., and Rapoport, J. L. (2018). Symptom dimensions and subgroups in childhood-onset schizophrenia. *Schizophrenia research*, 197:71–77.
- Cragg, L., Kovacevic, N., McIntosh, A. R., Poulsen, C., Martinu, K., Leonard, G., and Paus, T. (2011). Maturation of EEG power spectra in early adolescence: a longitudinal study. *Developmental science*, 14(5):935–943.
- Crouse, J. J., Chitty, K. M., Iorfino, F., Carpenter, J. S., White, D., Nichles, A., Zmicerevska, N., Tickell, A. M., Lee, R. S., and Naismith, S. L. (2020). Transdiagnostic neurocognitive subgroups and functional course in young people with emerging mental disorders: a cohort study. *BJPsych open*, 6(2).
- Davies, D. (1979). L., Bouldin, D., W., A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1 (2)*, pages 224–227.
- Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific reports*, 5:16743.
- Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PloS one*, 7(2):e30135.
- Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4):1443–1449.
- Desgraupes, B. (2018). *clusterCrit: Clustering Indices*. R package version 1.2.8.
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1):95–104.
- Forbes, M. K., Sunderland, M., Rapee, R. M., Batterham, P. J., Cate, A. L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A. J., Lynch, S. J., et al. (2021). A detailed hierarchical model of psychopathology: From individual symptoms up to the general factor of psychopathology. *Clinical Psychological Science*, 9(2):139–168.
- Gatt, J. M., Burton, K. L., Schofield, P. R., Bryant, R. A., and Williams, L. M. (2014). The heritability of mental health and wellbeing defined using COMPAS-W, a new composite measure of wellbeing. *Psychiatry Research*, 219(1):204–213.
- Goerg, G. M. (2013). Forecastable Component Analysis. In *JMLR, W&CP*, volume 28, pages 64–72.
- Goerg, G. M. (2020). *ForeCA: An R package for Forecastable Component Analysis*. R package version 0.2.7.
- Gorji Sefidmazgi, M., Sayemuzzaman, M., Homaifar, A., Jha, M. K., and Liess, S. (2014). Trend analysis using non-stationary time series clustering based on the finite element method. *Nonlinear Processes in Geophysics*, 21(3):605–615.

- Guo, X., Zhu, E., Liu, X., and Yin, J. (2018). Deep embedded clustering with data augmentation. In *Asian conference on machine learning*, pages 550–565. PMLR.
- Hartmann, J. A., Nelson, B., Spooner, R., Paul Amminger, G., Chanen, A., Davey, C. G., McHugh, M., Ratheesh, A., Treen, D., and Yuen, H. P. (2019). Broad clinical high-risk mental state (CHARMS): methodology of a cohort study validating criteria for pluripotent risk. *Early Intervention in Psychiatry*, 13(3):379–386.
- Harvill, J. L., Kohli, P., and Ravishanker, N. (2017). Clustering nonlinear, nonstationary time series using BSLEX. *Methodology and Computing in Applied Probability*, 19(3):935–955.
- Hatch, S. L., Jones, P. B., Kuh, D., Hardy, R., Wadsworth, M. E., and Richards, M. (2007). Childhood cognitive ability and adult mental health in the British 1946 birth cohort. *Social science & medicine*, 64(11):2285–2296.
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52(1):258–271.
- Hennig, C. (2020). *fpc: Flexible Procedures for Clustering*. R package version 2.2-9.
- Hermens, D. F., Soei, E. X., Clarke, S. D., Kohn, M. R., Gordon, E., and Williams, L. M. (2005). Resting EEG theta activity predicts cognitive performance in attention-deficit hyperactivity disorder. *Pediatric neurology*, 32(4):248–256.
- Herreen, D. and Zajac, I. T. (2018). The reliability and validity of a self-report measure of cognitive abilities in older adults: More personality than cognitive function. *Journal of Intelligence*, 6(1):1.
- Hickie, I. B., Davenport, T. A., Hadzi-Paviovic, D., Koschera, A., Naismith, S. L., Scott, E. M., and Wilhelm, K. A. (2001). Development of a simple screening tool for common mental disorders in general practice. *Medical Journal of Australia*, 175:S10–S17.
- Jamieson, D., Broadhouse, K. M., McLoughlin, L. T., Schwenn, P., Parker, M. J., Lagopoulos, J., and Hermens, D. F. (2020). Investigating the association between sleep quality and diffusion-derived structural integrity of white matter in early adolescence. *Journal of Adolescence*, 83:12–21.
- Jaramillo-Jimenez, A., Suarez-Revelo, J. X., Ochoa-Gomez, J. F., Arroyave, J. A. C., Boccanegra, Y., Lopera, F., Buriticá, O., Pineda-Salazar, D. A., Gómez, L. M., Quintero, C. A. T., et al. (2021). Resting-state EEG alpha/theta ratio related to neuropsychological test performance in Parkinson’s Disease. *Clinical Neurophysiology*, 132(3):756–764.
- Keizer, A. W. (2019). Standardization and Personalized Medicine Using Quantitative EEG in Clinical Settings. *Clinical EEG and neuroscience*, page 1550059419874945.
- Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S.-L. T., Manderscheid, R. W., and Walters, E. E. (2003). Screening for serious mental illness in the general population. *Archives of general psychiatry*, 60(2):184–189.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3):169–195.
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., and Clark, L. A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of abnormal psychology*, 126(4):454.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

- Latzman, R. D. and DeYoung, C. G. (2020). Using empirically-derived dimensional phenotypes to accelerate clinical neuroscience: The Hierarchical Taxonomy of Psychopathology (HiTOP) framework. *Neuropsychopharmacology*, 45(7):1083–1085.
- Lehembre, R., Bruno, M.-A., Vanhaudenhuyse, A., Chatelle, C., Cologan, V., Leclercq, Y., Soddu, A., Macq, B., Laureys, S., and Noirhomme, Q. (2012). Resting-state EEG study of comatose patients: a connectivity and frequency analysis to find differences between vegetative and minimally conscious states. *Functional neurology*, 27(1):41.
- Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.5.5-1.
- Liedorp, M., Van Der Flier, W., Hoogervorst, E., Scheltens, P., and Stam, C. (2009). Associations between patterns of EEG abnormalities and diagnosis in a large memory clinic cohort. *Dementia and geriatric cognitive disorders*, 27(1):18–23.
- Loo, S. K., Lenartowicz, A., and Makeig, S. (2016). Research review: Use of EEG biomarkers in child psychiatry research—current state and future directions. *Journal of Child Psychology and Psychiatry*, 57(1):4–17.
- Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., and Pietrzak, R. H. (2009). Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives of Clinical Neuropsychology*, 24(2):165–178.
- McGorry, P. D. and Mei, C. (2018). Early intervention in youth mental health: progress and future directions. *Evidence-based mental health*, 21(4):182–184.
- McGorry, P. D., Purcell, R., Goldstone, S., and Amminger, G. P. (2011). Age of onset and timing of treatment for mental and substance use disorders: implications for preventive intervention strategies and models of care. *Current opinion in psychiatry*, 24(4):301–306.
- McVoy, M., Lytle, S., Fulchiero, E., Aebi, M. E., Adeleye, O., and Sajatovic, M. (2019). A systematic review of quantitative EEG as a possible biomarker in child psychiatric disorders. *Psychiatry research*, 279:331–344.
- Mengersen, K. L., Drovandi, C. C., Robert, C. P., Pyne, D. B., and Gore, C. J. (2016). Bayesian estimation of small effects in exercise and sports science. *PloS one*, 11(4):e0147311.
- Moberget, T., Alnæs, D., Kaufmann, T., Doan, N. T., Córdova-Palomera, A., Norbom, L. B., Rokicki, J., van der Meer, D., Andreassen, O. A., and Westlye, L. T. (2019). Cerebellar gray matter volume is associated with cognitive function and psychopathology in adolescence. *Biological psychiatry*, 86(1):65–75.
- Mouselimis, L. (2020). *ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering*. R package version 1.2.2.
- Ousley, O. and Cermak, T. (2014). Autism spectrum disorder: defining dimensions and subgroups. *Current developmental disorders reports*, 1(1):20–28.
- Paus, T., Keshavan, M., and Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature reviews neuroscience*, 9(12):947–957.
- Pillay, K., Dereymaeker, A., Jansen, K., Naulaers, G., and De Vos, M. (2020). Applying a data-driven approach to quantify eeG maturational deviations in preterms with normal and abnormal neurodevelopmental outcomes. *Scientific reports*, 10(1):1–14.
- Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197.

- Prerau, M. J., Brown, R. E., Bianchi, M. T., Ellenbogen, J. M., and Purdon, P. L. (2017). Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. *Physiology*, 32(1):60–92.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raballo, A. and Poletti, M. (2020). Advances in early identification of children and adolescents at risk for psychiatric illness. *Current Opinion in Psychiatry*, 33(6):611–617.
- Rabbitt, P. and Abson, V. (1990). ‘Lost and Found’: Some logical and methodological limitations of self-report questionnaires as tools to study cognitive ageing. *British Journal of Psychology*, 81(1):1–16.
- Reddy, A. G. and Narava, S. (2013). Artifact removal from EEG signals. *International Journal of Computer Applications*, 77(13).
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.
- Saby, J. N. and Marshall, P. J. (2012). The utility of EEG band power analysis in the study of infancy and early childhood. *Developmental neuropsychology*, 37(3):253–273.
- Sawyer, M. G., Reece, C. E., Sawyer, A. C., Johnson, S. E., and Lawrence, D. (2018). Has the prevalence of child and adolescent mental disorders in Australia changed between 1998 and 2013 to 2014? *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(5):343–350. e5.
- Smit, C. M., Wright, M. J., Hansell, N. K., Geffen, G. M., and Martin, N. G. (2006). Genetic variation of individual alpha frequency (IAF) and alpha power in a large adolescent twin sample. *International Journal of Psychophysiology*, 61(2):235–243.
- Steinbach, M., Ertöz, L., and Kumar, V. (2004). *The challenges of clustering high dimensional data*, pages 273–309. Springer.
- Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in cognitive sciences*, 9(2):69–74.
- Thomas, E., Temko, A., Lightbody, G., Marnane, W., and Boylan, G. (2010). Gaussian mixture models for classification of neonatal seizures using EEG. *Physiological Measurement*, 31(7):1047.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.
- Tian, Y., Zhang, H., Xu, W., Zhang, H., Yang, L., Zheng, S., and Shi, Y. (2017). Spectral entropy can predict changes of working memory performance reduced by short-time training in the delayed-match-to-sample task. *Frontiers in human neuroscience*, 11:437.
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, pages 99–114.
- Van Spijker, B. A., Batterham, P. J., Calear, A. L., Farrer, L., Christensen, H., Reynolds, J., and Kerkhof, A. J. (2014). The Suicidal Ideation Attributes Scale (SIDAS): Community-based validation study of a new scale for the measurement of suicidal ideation. *Suicide and Life-Threatening Behavior*, 44(4):408–419.
- Walden, A. T., Percival, D. B., and McCoy, E. J. (1998). Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*, 46(12):3153–3165.

- Weeks, M., Wild, T., Ploubidis, G., Naicker, K., Cairney, J., North, C., and Colman, I. (2014). Childhood cognitive ability and its relationship with anxiety and depression in adolescence. *Journal of affective disorders*, 152:139–145.
- Williams, P. G., Rau, H. K., Suchy, Y., Thorgusen, S. R., and Smith, T. W. (2017). On the validity of self-report assessment of cognitive abilities: Attentional control scale associations with cognitive performance, emotional adjustment, and personality. *Psychological Assessment*, 29(5):519.
- Xie, X. L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 13(8):841–847.
- Zhao, L. and He, Y. (2013). Power spectrum estimation of the welch method based on imagery EEG. In *Applied Mechanics and Materials*, volume 278, pages 1260–1264. Trans Tech Publ.

1 Supplementary Materials

All code available on GitHub at https://github.com/of2/EEG_clustering_public

1.1 Data Cleaning & Pre-processing

The steps implemented in the pre-processing pipeline are as follows:

1. Setup: import data & channel location information, and select EEG Channels. Trim EEG session to start and end event tags (remove leading and trailing parts of the recording).
2. Downsample signal to 512Hz (half the original sampling rate of 1024 Hz) - this will enable the ICA to run faster, and will still allow for high frequency signals up to the corresponding Nyquist frequency of 256Hz.
3. High-pass filter at 1Hz to remove DC drift (Voipio et al., 2003).
4. Use EEGLAB function `cleanLineNoise` to remove AC electrical background noise at 50Hz and harmonics at 100, 150, 200 Hz.
5. Remove bad channels identified using EEGLAB function `clean_rawdata()` (e.g. electrodes with poor contact/faulty connection).
6. Automatically remove artifacts in continuous data using Artifact Subspace Reconstruction (ASR) (Kothe and Makeig, 2013; Chang et al., 2019). ASR is an automated technique that removes artifacts by: decomposing the signal into principal components in a sliding window (here of width 0.5 seconds); comparing the standard deviation of these components to the standard deviation of an identified 'clean' section of the data; rejecting 'bad' components which have a SD greater than some threshold multiple of the SD in the 'clean' section (here we used a threshold of 20); and reconstructs the signal in the sliding window by combining PCs remaining after any removal. This tool is appropriate for removing short-lasting artifacts, for instance from transient connection problems or touching of EEG electrodes (Miyakoshi et al., 2020).
7. Interpolate removed bad channels using adjacent remaining channels, using spherical interpolation with EEGLAB function `pop_interp()`.
8. Re-reference the data to the average signal.
9. Run Adaptive Mixture Independent Components Analysis (AMICA), which has been demonstrated to outperform other ICA methods for this application of EEG artifact removal (Delorme et al., 2012).
10. Reject independent components (ICs) which are labeled $< 70\%$ likely to be 'brain' signal components, using EEGLAB function `iclabel()`. This function uses an algorithm trained on expert-labeled EEG data to probabilistically characterise ICs which represent brain signal, eye blinks/movements, muscle movements, and other biosignal artifacts (Pion-Tonachini et al., 2019).
11. Post-processing to retain selected ICs and save the output.

1.2 Further Details on Unsupervised Clustering Algorithms

***k*-means**

The *k*-means algorithm is one of the most popular and widely used methods for unsupervised clustering applications. It is a 'hard clustering' approach in which each cluster must be assigned at least one data point, and each data point is assigned to one cluster with probability 1, and 0 probability of membership in all other clusters (Hartigan and Wong, 1979;

Ghosal et al., 2020). By the standard use of the Euclidean distance metric, k -means implicitly assumes that each cluster has spherical variance, and an approximately equal number of members (Raykov et al., 2016). Accordingly this algorithm will tend to find evenly sized partitions of data which are approximately circular or spherical around their centroid points. We calculated results for k -means using the `kmeans()` function with default settings from the base package 'stats' in R (R Core Team, 2021).

Hierarchical clustering (Ward's method)

Hierarchical agglomerative clustering begins with each point in a dataset assigned to its own cluster, and at each step iteratively combines the two most similar clusters (with similarity defined according to some chosen metric), until all data points are assigned to one cluster (Hartigan, 1975). The input used for the hierarchical clustering algorithm in this case is a distance matrix, containing the Euclidean distances between data points. The present study implements hierarchical clustering using Ward's minimum variance method, which agglomerates clusters based on a sum-of-squares criterion, using multivariate Euclidean distance (Ward Jr, 1963; Murtagh and Legendre, 2014). In this way this algorithm produces groups that minimise the sums of squared distances between data points and their cluster centroids, with centroids calculated as the arithmetic mean of all points x_{ij} in cluster c_i (Matani et al., 2003). We calculated results for hierarchical clustering using the `hclust()` function with method "ward.D2" from the base package 'stats' in R (R Core Team, 2021).

Gaussian Mixture Model

Another popular clustering method is the Gaussian mixture model (GMM), which applies the assumption that data is generated from a mixture of underlying Gaussian distributions, and assigns each data point a probability of allocation to each component distribution (Aggarwal and Reddy, 2014). For application to clustering, the groups of data points that have the highest probability of belonging to each component distribution can be defined as clusters. Maximum likelihood estimation (MLE) can be implemented with Expectation-Maximisation (EM) to infer the parameters of the component distributions that provide the best fit to the data. Compared to the other two clustering algorithms used in this work, the GMM is likely to find clusters with more flexible boundary shapes than k -means (Raykov et al., 2016), and is also unique in having probabilistic cluster assignment. For the applications in this work we use the crisp projection of GMM results with allocation of each data point to the mixture component in which it is most likely to belong. The present study implements GMM with default settings including Euclidean distance and a diagonal covariance matrix using the package 'ClusterR' in R, version 1.22 (Mouselimis, 2020).

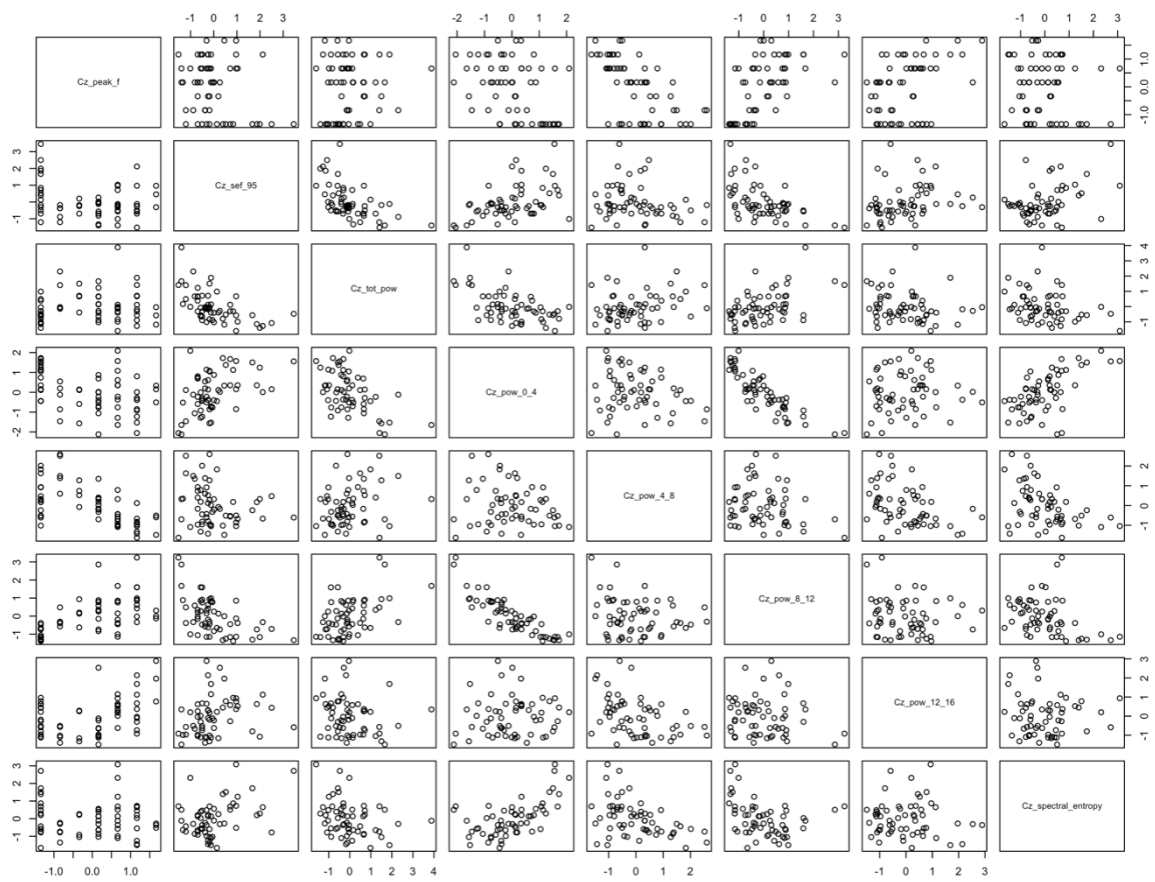
1.3 Testing for Linear Relationships Among Variables for PCA

Bartlett's test of sphericity returned a significant result, indicating the presence of linear correlations among the 8 variables calculated as summary features of EEG frequency content: $\chi^2(28) = 399.44, p = 2.53 \times 10^{-6}$.

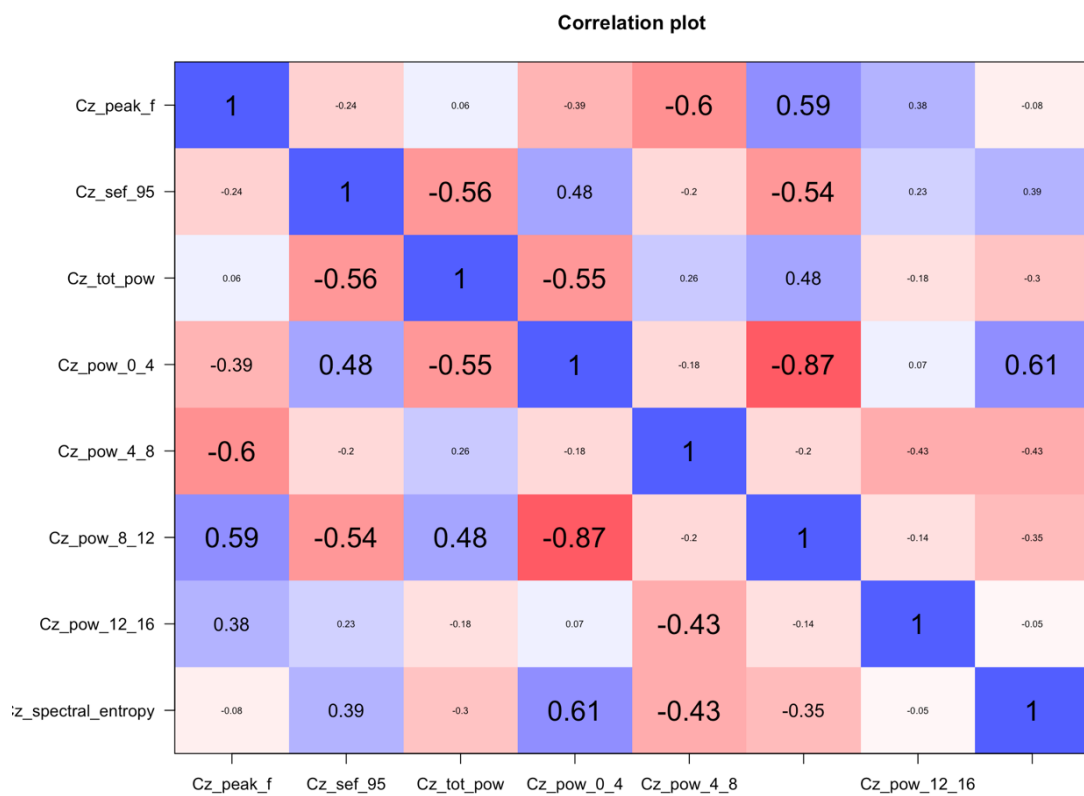
From visual inspection of pairwise scatterplots of these 8 input features (Supplementary Figure 1), correlation plots for all pairwise combinations of features (Supplementary Figure 2), and Bartlett's test of sphericity to assess presence of linear correlations between variables, we identified the presence of linear correlations between these variables with no obvious nonlinear relationships, indicating that PCA was an appropriate approach for dimension reduction.

1.4 PCA - Eigenvalues & Variance Explained

Supplementary Table 1 presents eigenvalues, standard deviations, proportion of variance explained, and cumulative proportion of variance explained for the principal components. Sup-



Supplementary Figure 1: Pairwise scatterplots of EEG frequency features

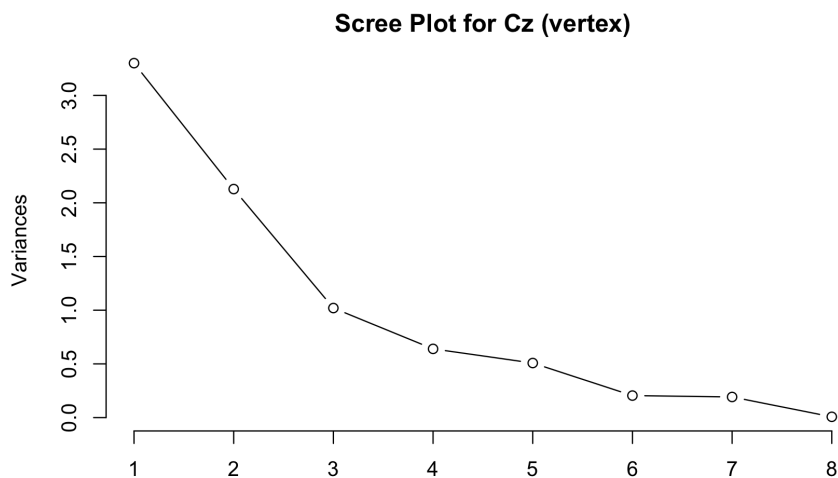


Supplementary Figure 2: Pairwise correlations between EEG frequency features

plementary Figure 3 presents the scree plot for this PCA. It can be seen that the proportion of variance explained substantially decreases after the third component.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Eigenvalues	3.3004	2.1280	1.0199	0.6398	0.5079	0.2048	0.1920	0.0072
Standard Deviation	1.8167	1.4588	1.0099	0.7999	0.7127	0.4526	0.4382	0.0847
Proportion of Variance	0.4126	0.2660	0.1275	0.0800	0.0635	0.0256	0.0240	0.0009
Cumulative Proportion	0.4126	0.6785	0.8060	0.8860	0.9495	0.9751	0.9991	1.0000

Supplementary Table 1: Eigenvalues & variance explained for PCA of EEG frequency features



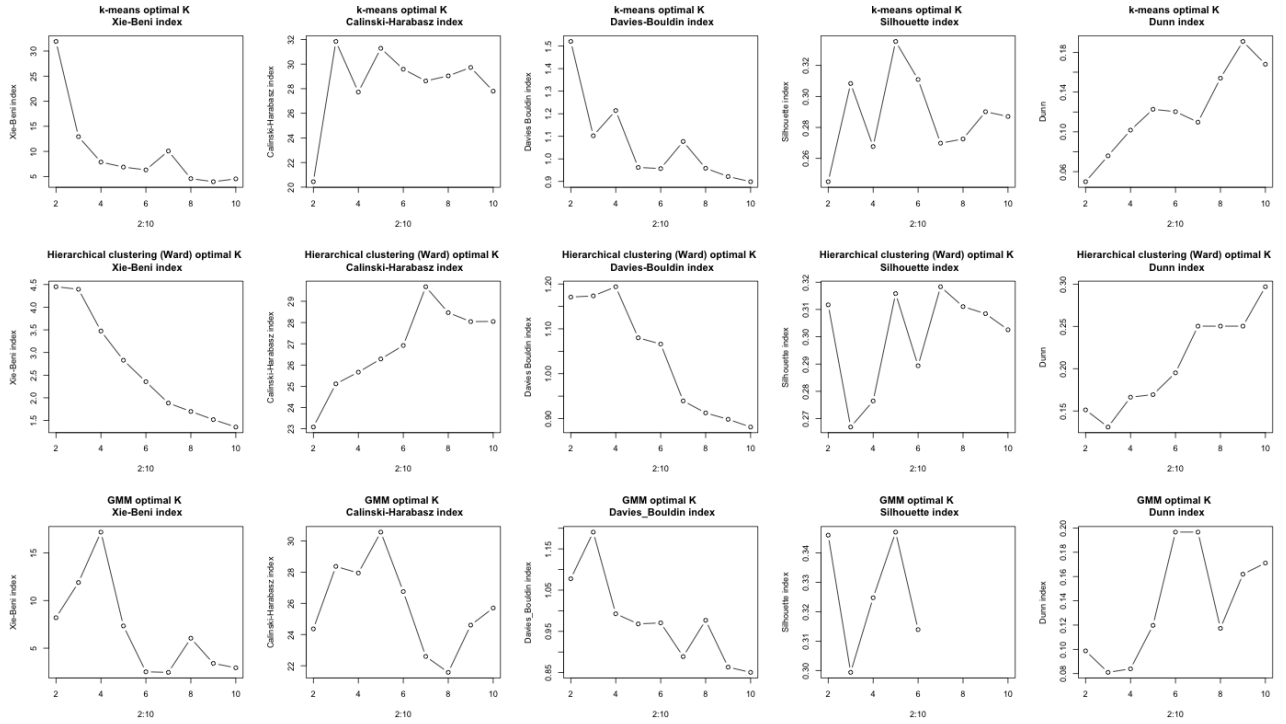
Supplementary Figure 3: Scree plot for PCA of frequency features

1.5 Selecting K and Identifying Core Clusters

For GMM, internal validation criteria were calculated from a crisp projection of the results, assigning each data point to the mixture component with the highest associated probability.

Supplementary Figure 4 presents internal validation indices results for between 2 and 10 clusters calculated for each method. For each index, our approach included: to assess the point at which the creation of new additional clusters had diminishing returns (addition of new clusters had decreasing benefit to index value), appearing as an elbow (for indices where the value should be minimised - Xie-Beni, Davies-Bouldin) or shoulder/peak (for indices where the value should be maximised - Calinski-Harabasz, Silhouette, Dunn); and to look for consistency in the optimal number of clusters across the different internal validation metrics.

For k-means (top row), across these metrics we identified four votes for $K = 5$ (Calinski-Harabasz, Davies-Bouldin, Silhouette and Dunn), and one vote for $K = 4$ (Xie-Beni), and so we chose a solution with $K = 5$ for k-means. For HC (middle row), across these metrics we identified one neutral vote (Xie-Beni monotonically decreasing with increasing K), two marginal votes for $K = 5$ or 7 (Davies-Bouldin and Silhouette), and two votes for $K = 7$ (Calinski-Harabasz and Dunn). For GMM (bottom row), across these metrics we identified three votes for $K = 5$ (Calinski-Harabasz, Davies-Bouldin and Silhouette), and two vote for



Supplementary Figure 4: Internal validation indices used to select optimal number of clusters K for each method.

Top row = k-means; middle row = hierarchical clustering; bottom row = GMM

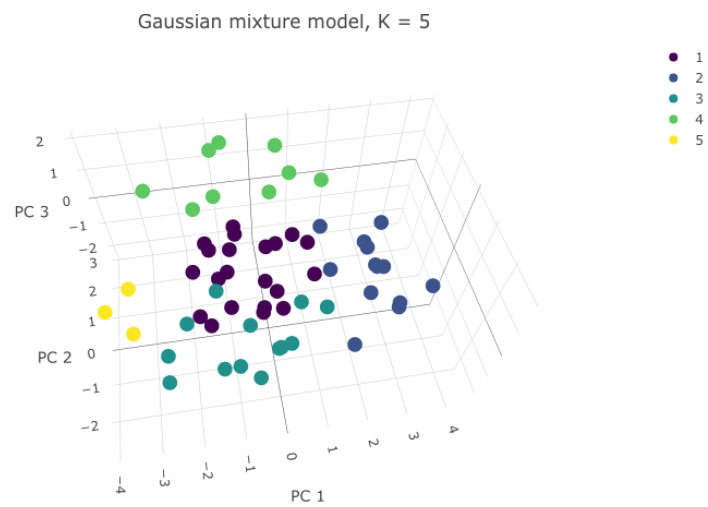
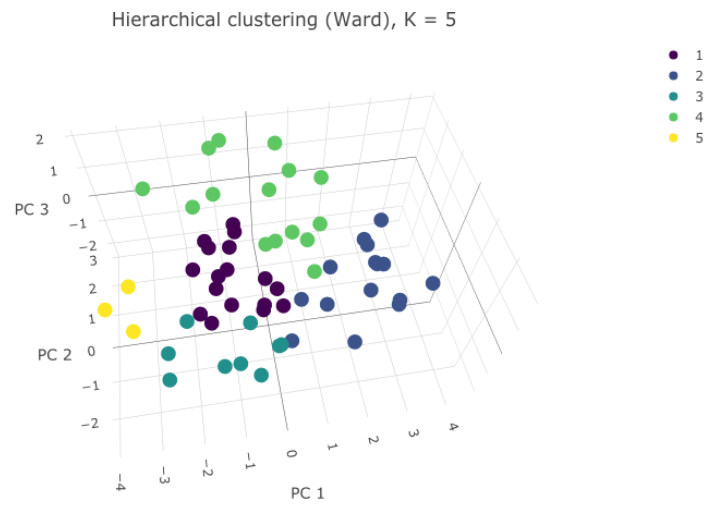
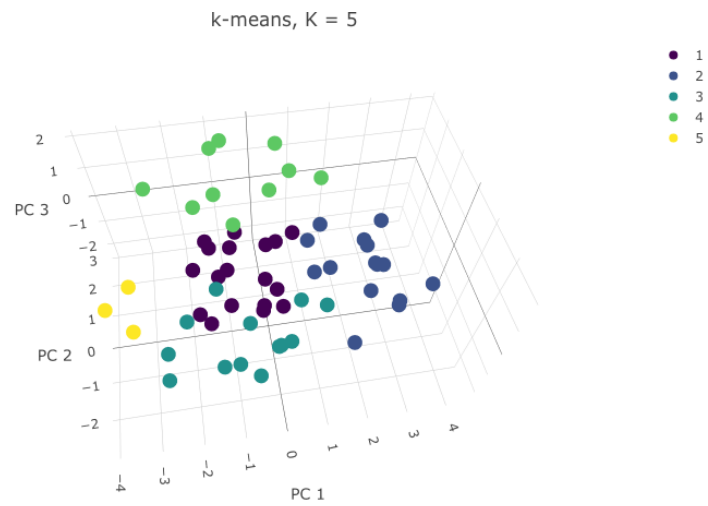
$K = 6$ (Xie-Beni and Dunn). We chose $K = 5$ for GMM. Of note here, the method used in the ClusterCrit package to calculate the Silhouette index is not defined for solutions with any clusters containing only one member, which is why Silhouette values do not appear for the GMM solutions of $K = 7$ and above.

We tested the effect of constructing integrated clusters based on combining 5-cluster solutions from k -means and GMM with the 7-cluster solution from HC. We found that the principal difference was that one of the integrated clusters from our final results (core cluster 4, $n=8$) was split into two equally sized sub-clusters ($n=4$), separated mainly in PC1. When we compared these two clusters we found that they were very similar across all frequency features except total power, and were otherwise similar in IAF, fractional power bands, SEF95 and spectral entropy. Since this split was only marginally supported by one of the clustering algorithms, we decided to keep them together in core cluster 4 for the final results.

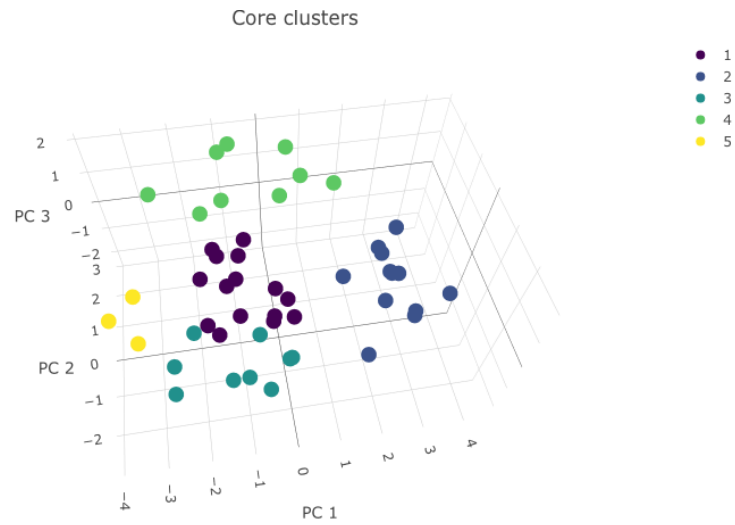
1.6 Core clusters - 3D Scatter Plots (Static)

Supplementary Figure 5 presents 3D scatter plots of clustering results for each clustering method.

Supplementary Figure 6 presents a 3D scatter plot of individuals by principal component scores coloured by core cluster membership.



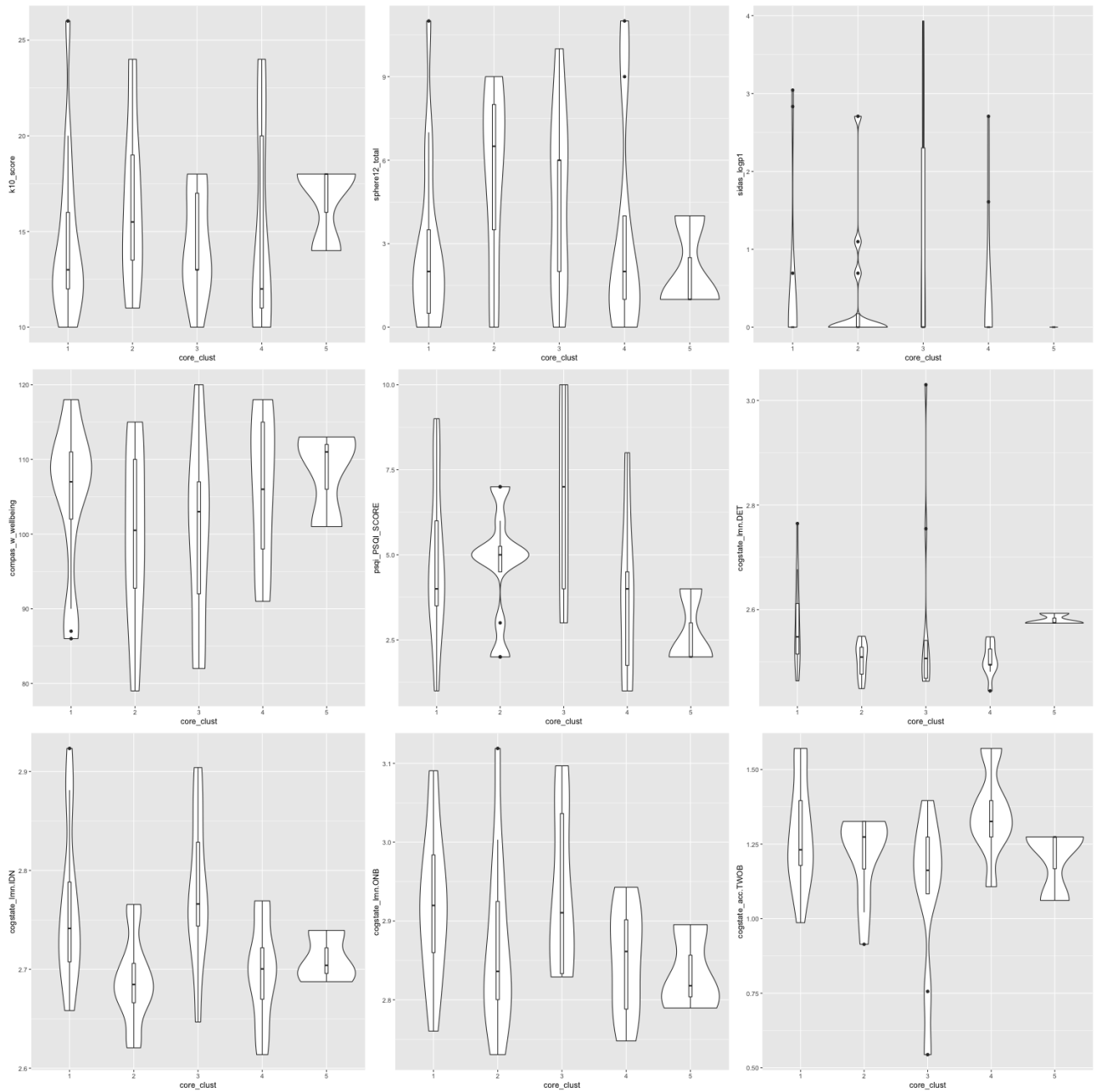
Supplementary Figure 5: 3D scatter plots of individuals by component scores, coloured by cluster membership, $K = 5$. Top panel = k -means; middle panel = HC; bottom panel = GMM.



Supplementary Figure 6: 3D scatter plot of individuals by component scores, coloured by core cluster membership (overlapping clustering results between the 3 separate algorithms presented in Supplementary Figure 5). Individuals not allocated to core clusters ($n = 11$) have been removed.

Interactive html 3d plots are available in the online Supplementary Materials

1.7 Violin plots for health measures across core clusters



Supplementary Figure 7: Violin plots for health & cognitive function measures across clusters. For SIDAS scores, panel C displays $\log(\text{SIDAS score} + 1)$ to account for positive skew. K10 = Kessler Psychological Distress Scale; SPHERE12 = SPHERE-12 Screening Questionnaire; SIDAS = Suicidal Ideation Attributes Scale; COMPAS-W = COMPAS-W Wellbeing Scale; PSQI = Pittsburgh Sleep Quality Index; DET - RT = CogState Detection Task Reaction Time; IDN - RT = CogState Identification Task Reaction Time; OBK - RT = CogState One Back Task Reaction Time; TWOB - Acc = CogState Two Back Task Accuracy Score.

1.8 Expected Marginal Means for Cluster Contrasts

Contrast	K10	SPHERE-12	SIDAS	COMPAS-W	PSQI	DET - RT	IDN - RT	OBK - RT	TWOB - Acc
1 > 2	-2.1 (-5.46, 1.18)	-2.63 (-5.15, 0.01)	1.03 (-5.31, 7.77)	5.31 (-3.33, 13.84)	-0.04 (-1.89, 1.7)	0.08 (0, 0.15)	0.08 (0.03, 0.13)	0.05 (-0.02, 0.13)	0.07 (-0.08, 0.22)
1 > 3	0.06 (-3.64, 3.61)	-1.86 (-4.68, 0.98)	-6.45 (-13.55, 0.72)	5.07 (-4.33, 14.43)	-1.82 (-3.8, 0.1)	-0.02 (-0.1, 0.07)	-0.02 (-0.08, 0.04)	-0.02 (-0.1, 0.06)	0.2 (0.04, 0.36)
1 > 4	-0.36 (-3.97, 3.28)	-0.54 (-3.36, 2.22)	0.41 (-6.62, 7.6)	-0.89 (-10.29, 8.65)	0.98 (-1.08, 2.96)	0.08 (-0.01, 0.16)	0.07 (0.01, 0.12)	0.07 (-0.01, 0.15)	-0.08 (-0.24, 0.08)
1 > 5	-2.27 (-7.64, 3.27)	0.8 (-3.29, 5.12)	2.46 (-8.09, 13.31)	-3.36 (-17.81, 10.7)	2.03 (-0.84, 4.97)	0 (-0.13, 0.13)	0.06 (-0.03, 0.14)	0.09 (-0.03, 0.21)	0.08 (-0.16, 0.32)
2 > 3	2.17 (-1.77, 5.89)	0.76 (-2.2, 3.74)	-7.5 (-15.19, -0.12)	-0.25 (-9.82, 9.59)	-1.77 (-3.86, 0.28)	-0.09 (-0.18, 0)	-0.1 (-0.16, -0.04)	-0.07 (-0.15, 0.02)	0.14 (-0.04, 0.31)
2 > 4	1.74 (-2.06, 5.47)	2.08 (-0.8, 5.06)	-0.6 (-8.46, 6.73)	-6.22 (-16.18, 3.57)	1.02 (-1.08, 3.11)	0 (-0.09, 0.09)	-0.01 (-0.07, 0.05)	0.02 (-0.07, 0.1)	-0.15 (-0.32, 0.02)
2 > 5	-0.17 (-3.75, 5.38)	3.46 (-0.88, 7.74)	1.44 (-9.28, 12.71)	-8.69 (-23.25, 5.68)	2.08 (-0.91, 4.98)	-0.08 (-0.21, 0.05)	-0.02 (-0.11, 0.06)	0.03 (-0.09, 0.16)	0.01 (-0.23, 0.26)
3 > 4	-0.44 (-4.45, 4.06)	1.33 (-1.82, 4.48)	6.89 (-1.05, 14.94)	-5.94 (-16.37, 4.98)	2.8 (0.38, 5.16)	0.09 (0, 0.19)	0.09 (0.02, 0.15)	0.09 (0, 0.18)	-0.28 (-0.46, -0.1)
3 > 5	-2.34 (-8.12, 3.43)	2.68 (-1.96, 6.93)	8.94 (-2.33, 20.32)	-8.44 (-23.35, 7.02)	3.86 (0.83, 6.97)	0.02 (-0.12, 0.15)	0.07 (-0.02, 0.16)	0.1 (-0.03, 0.23)	-0.13 (-0.38, 0.13)
4 > 5	-1.9 (-7.68, 3.88)	1.36 (-3.02, 5.89)	2 (-9.27, 13.34)	-2.51 (-17.43, 12.51)	1.05 (-2.05, 4.13)	-0.08 (-0.21, 0.06)	-0.01 (-0.1, 0.07)	0.02 (-0.11, 0.15)	0.16 (-0.09, 0.41)

Supplementary Table 2: Expected marginal means for pairwise contrasts between clusters, calculated from Bayesian regression models. Posterior mean (95% credible interval). Cells corresponding to contrasts with posterior probability less than 10% or greater than 90% (Table 9) are presented in bold. K10 = Kessler Psychological Distress Scale; SPHERE12 = SPHERE-12 Screening Questionnaire; SIDAS = Suicidal Ideation Attributes Scale; COMPAS-W = COMPAS-W Wellbeing Scale; PSQI = Pittsburgh Sleep Quality Index; DET - RT = CogState Detection Task Reaction Time; IDN - RT = CogState Identification Task Reaction Time; OBK - RT = CogState One Back Task Reaction Time; TWOB - Acc = CogState Two Back Task Accuracy Score.

2 References

References

- Aggarwal, C. C. and Reddy, C. K. (2014). *Data clustering: Algorithms and applications*. Chapman & Hall/CRC Data mining and Knowledge Discovery Series. London.
- Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., and Jung, T.-P. (2019). Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel eeg recordings. *IEEE Transactions on Biomedical Engineering*, 67(4):1114–1121.
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent eeg sources are dipolar. *PLoS one*, 7(2):e30135.
- Ghosal, A., Nandy, A., Das, A. K., Goswami, S., and Panday, M. (2020). A short review on different clustering techniques and their applications. *Emerging technology in modelling and graphics*, pages 69–83.
- Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, Inc.
- Hartigan, J. A. and Wong, M. A. (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1):100–108.
- Kothe, C. A. and Makeig, S. (2013). Bcilab: a platform for brain–computer interface development. *Journal of neural engineering*, 10(5):056014.
- Matani, A., Masuda, Y., Okubo, H., and Chihara, K. (2003). Hierarchical clustering and filtering in half-inverse space for meg and/or eeg hypothesis-free analysis. *IEEE transactions on signal processing*, 51(2):350–361.
- Miyakoshi, M., Jurgiel, J., Dillon, A., Chang, S., Piacentini, J., Makeig, S., and Loo, S. K. (2020). Modulation of frontal oscillatory power during blink suppression in children: Effects of premonitory urge and reward. *Cerebral Cortex Communications*, 1(1):tgaa046.
- Mouselimis, L. (2020). *ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering*. R package version 1.2.2.
- Murtagh, F. and Legendre, P. (2014). Ward’s hierarchical agglomerative clustering method: which algorithms implement ward’s criterion? *Journal of classification*, 31(3):274–295.
- Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raykov, Y. P., Boukouvalas, A., Baig, F., and Little, M. A. (2016). What to do when k-means clustering fails: a simple yet principled alternative algorithm. *PLoS one*, 11(9):e0162259.
- Voipio, J., Tallgren, P., Heinonen, E., Vanhatalo, S., and Kaila, K. (2003). Millivolt-scale dc shifts in the human scalp eeg: evidence for a nonneuronal generator. *Journal of neurophysiology*, 89(4):2208–2214.
- Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244.