A Appendix A

The clinical notes for all study participants were manually read and assessed by a team of trained clinical researchers. Data was then extracted from these clinical notes by the clinical researchers and entered into the proforma as required. All clinical notes are generated by the study participants' treating clinician/s as part of their standard care.

A.1 Demographics

Biological sex, and age. Current engagement in part- or full-time education or employment to determine Not in Education, Employment, or Training (NEET) status. NEET is assigned if there was no full- or part-time education, employment, training, or volunteer work.

A.2 Social and Occupational Functioning

The Social and Occupational Functioning Assessment Scale SOFAS;1 is a clinician-rated measure that assesses functioning on a 0–100 scale, with lower scores suggesting functional impairment. The instructions emphasise that the rater should avoid confounding the rating with clinical symptoms 1-3. A SOFAS score of below 70 is considered to be clinically-significant impairment 4.

A.3 Mental Disorder Diagnoses

Mental disorder diagnoses at each time point are classified according to DSM-5 criteria 5 and specified as either full- or sub-threshold. Diagnoses are also labelled as either primary, secondary, or tertiary based on judgement of which was the dominant presenting problem at that time point.

Mental disorder diagnosis is determined solely by the symptomology and/or diagnosis reported and recorded by the treating clinician/s as presented in the clinical notes of each study participant. Based on the information provided within these clinical notes, researchers determined whether DSM-5 criteria were met for a specific disorder at that time point. If symptomology recorded in the clinical notes indicated only some, but not all criteria being met for a specific disorder, then a sub-threshold classification was recorded. If symptomology indicated full DSM-5 criteria were met for that time point, then a full-threshold classification was recorded.

As per diagnosis, medication is also obtained from a review of the clinical notes as generated by the study participants' clinician/s. A certain medication is recorded if the clinical notes indicate that the study participant took that particular class of medication within the specified timeframe.

A.4 At-risk Mental States

Clusters of symptoms that have been previously indicated as risk factors for progression to more severe mental disorders 6-11 are recorded in all individuals regardless of diagnosis. This includes psychotic-like experiences (the presence of any psychotic symptoms including: perceptual abnormalities, bizarre ideas, disorganised speech, etc), manic-like experiences (the presence of any manic/hypomanic symptoms including: abnormally elevated mood or irritability; increased motor activity, speech, or sexual interest, etc), and circadian disturbance (the presence of significant disruption in sleep-wake or circadian cycles including the presence of a severe sleep-wake disorder or chronic fatigue). The presence or absence of these clusters of symptoms is determined solely by the symptomology reported and recorded by the treating clinician/s as presented in the clinical notes of each study participant. Similarly, the distinction between psychotic-like and manic-like symptoms is judged within the context of the clinical notes.

The threshold for mania like experiences and psychotic like experiences in this study is low. Conversely, the threshold for circadian disturbance in this study is high. More specifically, these experiences are rated based on their presence or absence and the nature (e.g. type, severity, frequency) of these experiences, and so stage 1a and stage 1b MLEs and PLEs are not necessarily different, but in some cases may differ in nature. The presence of these symptoms does not necessarily mean the participant currently has / or will go on to develop a serious mental health disorder. It is simply one of many risk factors that may exist. Moreover, the presence or absence of these symptoms do not, in and of themselves, determine the staging of a participant.

A.5 Self-harm and Suicidal Thoughts and Behaviours

The presence of suicidal ideation, suicide attempts, and self-harm is recorded. A suicide attempt is recorded when a young person has taken steps to take their own life. If an individual harms themselves via cutting, hitting themselves, burning themselves, or scratching with the intention to self-harm only and not to take their life, then this is included as self-harm and not a suicide attempt.

A.6 Physical Health Comorbidities

Any major physical illness is recorded. This includes (but is not limited to): diabetes, cancer, asthma, chronic pain, epilepsy and obesity.

A.7 Personal Mental Illness History

Known childhood-onset disorders (i.e. with clear onset prior to 12 years old) are recorded in addition to current diagnoses. Family history of a mental health disorder is ascertained via the treating clinician's assessment with the client. Please note, family history is only recorded when the client has reported a

mental health diagnosis of a first degree relative. Moreover, family history is only recorded if the client reported that the first-degree family member has a current of past diagnosed mental health disorder. Symptoms only, but no diagnosis, is not enough to meet criteria for this category.

A.8 Treatment Utilisation

Exposure to classes of medication (antidepressant, antipsychotic, mood stabiliser, or stimulant medication), and hospitalisation overnight or longer due to a mental health problem are recorded.

B Random effects modelling

In this section, we present technical details of the random effects model derived for this paper's statistical analysis. To represent the data, let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_J]^\top$ denote J individual response trajectories of length T, that is:

$$\mathbf{y}_j = [y_{j1}, \ldots, y_{jT}]^\top \in \mathbb{R}^T, \quad j \in \{1, \ldots, J\}\,. \tag{1}$$

For each individual j, we have access to their covariates $\mathbf{z}_j \in \mathbb{R}^D$ recorded at their first visit to the clinic (baseline). Each individual's response will be associated with a category, or cluster, $c_j \in \{1, \ldots, K\}$, where K is pre-specified.

B.1 Response model

Within each cluster $k \in \{1, \ldots, K\}$, we will assume a Gaussian linear model for the response trajectories of an individual $j \in \{1, \ldots, J\}$:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta}_{kj} + \boldsymbol{\epsilon}_j\,, \tag{2}$$

where $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ represents noise, linear regression coefficients $\boldsymbol{\beta}_{kj}$ differ by individual with a common prior per cluster, and $\mathbf{X}_j \in \mathbb{R}^{T\times F}$ is a design matrix based on the individual's observation time points. In particular, we adopt linear features:

$$\mathbf{X}_j := \begin{bmatrix} 1 & t_{j1} \\ \vdots & \vdots \\ 1 & t_{jT} \end{bmatrix}\,, \tag{3}$$

where $t_{j1}, \ldots, t_{jT}$ represent the individual's observation time points.

For any individual j, we then have a Gaussian likelihood model:

$$\mathbf{y}_j | k, \boldsymbol{\beta}_{kj}, \sigma^2 \sim \mathcal{N}(\mathbf{X}_T \boldsymbol{\beta}_{kj}, \sigma^2\mathbf{I})\,. \tag{4}$$

We apply an inverse-gamma prior to the noise variance parameter:

$$\sigma^2 \sim \Gamma^{-1}(a_\sigma, b_\sigma) \tag{5}$$

due to its conjugacy with a Gaussian likelihood. For the linear regression coefficients vector $\boldsymbol{\beta}_{kj}$, following the random effects approach, we place a cluster-

dependent prior, which is dependent on whether the intercept (i.e., the first component of the vector) and the slope (i.e., the second component of the vector) are assumed to be correlated or not. Both approaches use $K = 3$ clusters (i.e., *constant*, *up* and *down*) and $T = 2$ time points (with $t_1 = 0$ corresponding to the baseline).

B.1.1 Correlated-intercept response model

From preliminary experiments with real data, it was possible to observe that the initial score of an individual affects the cluster which they are assigned to. Individuals with a low initial score are more likely to improve, while higher scores indicate the individual is more likely to stay the same or somewhat deteriorate, at least during early stages. To model this behaviour, we can model the dependence of the slope on the intercept in their cluster-conditional priors. A principled approach for that is to use a multivariate prior for the linear regression coefficients. We would also like to encode constraints on monotonicity and limits of the initial score. Therefore, we propose the following model:

$$\boldsymbol{\alpha}_{jk} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha},k}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha},k}), \quad k \in \{1,2,3\},\ j \in \{1,\ldots,J\} \tag{6}$$

$$\beta_{jk,0} = \frac{1}{2}(y_{\min} + y_{\max}) + \frac{1}{2}(y_{\max} - y_{\min})\alpha_{jk,0} \tag{7}$$

$$\beta_{jk,1} = \begin{cases} 0, & \text{if } k = 1 \\ \exp(\alpha_{jk,1}), & \text{if } k = 2 \\ -\exp(\alpha_{jk,1}), & \text{if } k = 3 \end{cases} \tag{8}$$

This transform is invertible for both the intercept and the slope. However, the linearity of the intercept transform ensures $\beta_{jk,0}$ is still Gaussian, which allows for a closed-form solution for the posterior predictive distribution.

B.1.2 Hyper-priors

For this model, we specify independent hyper-priors for each prior parameter. The mean vector is assumed Gaussian:

$$\boldsymbol{\mu}_{\boldsymbol{\alpha},k} \sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}}), \quad k \in \{1,\ldots,K\}, \tag{9}$$

Hyper-parameter	$y_{\min}$	$y_{\max}$	$\widehat{\boldsymbol{\alpha}}$	$\boldsymbol{\sigma}_{\boldsymbol{\alpha}}$	$\mathbf{a}_\nu$	$\mathbf{b}_\nu$
Value	20	100	$[0,1]$	$[0.01, 0.01]$	$[10,10]$	$[1,2]$

Table 1: Hyper-parameter settings for the correlated-intercept response model

where the first component of $\widehat{\boldsymbol{\alpha}}$ is set to 0, while the second component is adjustable, and $\mathrm{diag}(\boldsymbol{\sigma}_{\boldsymbol{\alpha}})$ is a diagonal matrix with diagonal elements given by the vector $\boldsymbol{\sigma}_{\boldsymbol{\alpha}} \in \mathbb{R}^2$.

Care needed to be taken regarding the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha},k}$ for this re-

sponse model. Note that, if the slope and intercept coefficients are not correlated according to the prior, knowing the intercept provides no information about the slope. Therefore, one needs to ensure that the correlation between these two coefficients is non-zero according to the prior. One approach for that is to infer the prior covariance matrix, placing a hyper-prior on it which allows for correlation, such as a Wishart or a LKJ prior (Lewandowski et al., 2009). In our case, we use the latter combined with a separate inverse-Gamma hyper-prior for the diagonal covariance elements for each cluster. We then decompose the covariance matrices prior into a LKJ prior with concentration parameter set to 1 for the correlations, which makes it a non-informative, uniform prior over the space of correlation matrices (Lewandowski et al., 2009), and inverse-gamma priors for the scaling of each row. Namely, for $k \in \{1, \ldots, K\}$, we set:

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha},k} = \operatorname{diag}(\boldsymbol{\nu}_k^{1/2}) \mathbf{C}_k \operatorname{diag}(\boldsymbol{\nu}_k^{1/2}) \tag{10}$$

$$\boldsymbol{\nu}_k \sim \Gamma^{-1}(\mathbf{a}_{\boldsymbol{\nu}}, \mathbf{b}_{\boldsymbol{\nu}}) \tag{11}$$

$$\mathbf{C}_k^{1/2} \sim \mathrm{LKJ}(1) \tag{12}$$

where the vector notation for the inverse-gamma prior of $\boldsymbol{\nu}_k$ is to indicate independent inverse-gamma priors for each of the vector's components. Table 1 presents the hyper-parameter settings for this response model.

B.2 Predictions

To predict trajectories, having the initial score (i.e., the intercept) observed via the inverse transform:

$$\hat{\alpha}_{jk,0} = \frac{2y_{j,0} - y_{\min} - y_{\max}}{y_{\max} - y_{\min})}, \quad k \in \{1, 2, 3\},\, j \in \{1, \ldots, J\}\,. \tag{13}$$

Conditioning on the observed intercept, the untransformed linear coefficients are normally distributed:

$$\boldsymbol{\alpha}_{jk} | \hat{\alpha}_{jk,0} \sim \mathcal{N}(\hat{\boldsymbol{\alpha}}_{jk}, \hat{\boldsymbol{\Sigma}}_{jk})\,, \tag{14}$$

Hyper-parameter	a_ψ	b_ψ	g
Value	2	5	1

Table 2: Hyper-parameter settings for the cluster-assignment model

The predictive posterior parameters can be found by conditioning formulas for multivariate normal distributions:

(15)

$$\hat{\boldsymbol{\alpha}}_{jk} = \mu_{\boldsymbol{\alpha},k} + \frac{1}{\sigma^2_{\boldsymbol{\alpha},k,0}} \begin{bmatrix} \sigma^2_{\boldsymbol{\alpha},k,0} \\ \sigma_{\boldsymbol{\alpha},k,0,1} \end{bmatrix} (\alpha_{jk,0} - \mu_{\boldsymbol{\alpha},k,0})$$
$$\hat{\boldsymbol{\Sigma}}_{jk} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha},k} - \frac{1}{\sigma^2_{\boldsymbol{\alpha},k,0}} \begin{bmatrix} \sigma^2_{\boldsymbol{\alpha},k,0} \\ \sigma_{\boldsymbol{\alpha},k,0,1} \end{bmatrix} \otimes \begin{bmatrix} \sigma^2_{\boldsymbol{\alpha},k,0} \\ \sigma_{\boldsymbol{\alpha},k,0,1} \end{bmatrix}, \tag{16}$$

where $\sigma^2_{\boldsymbol{\alpha},k,0} := [\boldsymbol{\Sigma}_{jk}]_{0,0}$, $\sigma_{\boldsymbol{\alpha},k,0,1} := [\boldsymbol{\Sigma}_{jk}]_{0,1}$ and $\otimes$ denotes the outer product of two vectors.

B.3 Cluster-assignment model

We now specify a prior for the cluster membership assignment by making a simplifying assumption for the derivations. For an individual j, we set a categorical prior:

$$p(c_j = k|\boldsymbol{\Phi}, \mathbf{z}_j) = \frac{\exp(\mathbf{z}_j \cdot \boldsymbol{\phi}_k)}{1 + \sum_{k'=1}^{K-1} \exp(\mathbf{z}_j \cdot \boldsymbol{\phi}_{k'})}, \tag{17}$$

where we use a compact notation with $\boldsymbol{\Phi} := [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K]$ representing the matrix of cluster assignment parameters. We place a Zellner's g-prior (Zellner, 1986) on the cluster assignment parameters $\boldsymbol{\Phi}$:

$$\boldsymbol{\phi}_k|\mathbf{Z}, \psi \sim \mathcal{N}(\mathbf{0}, g\psi^{-1}(\mathbf{Z}^\top \mathbf{Z})^{-1}), \quad k \in \{2, 3\}, \tag{18}$$

where $\psi \sim \Gamma^{-1}(a_\psi, b_\psi)$ is a scaling parameter which is also inferred. We set $\boldsymbol{\phi}_1 := \mathbf{0}$, so that the model is not over-specified. In general, the parameter g can be set according to expert knowledge or estimated by a variety of methods (Liang et al., 2008). However, since we will infer ψ from data, we simply set $g := 1$. Note that, we now have a data-dependent prior for $\boldsymbol{\Phi}$ which allows us to automatically scale its probability distribution according to the range of the covariates $\mathbf{Z}$ in the data. Also note that $\mathbf{Z}$ is not part of the response data. Hyper-parameter settings for this model are summarised in Table 2.

B.4 Inference

We apply HMC within Gibbs for inference in this setting, since we still have discrete variables representing the cluster assignments to infer, i.e. $c_j \in \{1, 2, 3\}$. In particular, we adopt the modified Gibbs sampler proposed by Liu (1996) to sample the discrete variables. For the continuous variables, we adopt the popular no-U-turn sampler, an adaptive version of HMC (Hoffman and Gelman, 2014).

References

Hoffman, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random cor-

relation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9):1989–2001.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Liu, J. S. (1996). Peskun's theorem and a modified discrete-state Gibbs sampler. *Biometrika*, 83(3):681–682.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier Science Publishers, Inc., New York, NY.