

Deriving symptom networks from digital phenotyping data in serious mental illness

Ryan Hays, Matcheri Keshavan, Hannah Wisniewski and John Torous

Background

Symptoms of serious mental illness are multidimensional and often interact in complex ways. Generative models offer value in elucidating the underlying relationships that characterise these networks of symptoms.

Aims

In this paper we use generative models to find unique interactions of schizophrenia symptoms as experienced on a moment-by-moment basis.

Method

Self-reported mood, anxiety and psychosis symptoms, self-reported measurements of sleep quality and social function, cognitive assessment, and smartphone touch screen data from two assessments modelled after the Trail Making A and B tests were collected with a digital phenotyping app for 47 patients in active treatment for schizophrenia over a 90-day period. Patients were retrospectively divided up into various non-exclusive subgroups based on measurements of depression, anxiety, sleep duration, cognition and psychosis symptoms taken in the clinic. Associated transition probabilities for the patient cohort and for the clinical subgroups were calculated using state transitions between adjacent 3-day timesteps of pairwise survey domains.

Results

The three highest probabilities for associated transitions across all patients were anxiety-inducing mood (0.357, $P < 0.001$), psychosis-inducing mood (0.276, $P < 0.001$), and anxiety-inducing poor sleep (0.268, $P < 0.001$). These transition probabilities

were compared against a validation set of 17 patients from a pilot study, and no significant differences were found. Unique symptom networks were found for clinical subgroups.

Conclusions

Using a generative model using digital phenotyping data, we show that certain symptoms of schizophrenia may play a role in elevating other schizophrenia symptoms in future timesteps. Symptom networks show that it is feasible to create clinically interpretable models that reflect the unique symptom interactions of psychosis-spectrum illness. These results offer a framework for researchers capturing temporal dynamics, for clinicians seeking to move towards preventative care, and for patients to better understand their lived experience.

Keywords

mhealth; EMA; networks; psychosis; smartphones.

Copyright and usage

© The Author(s), 2020. Published by Cambridge University Press on behalf of the Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

Serious mental illnesses, such as schizophrenia, which are still often diagnosed using largely static symptom reports, are increasingly viewed as network illnesses¹ existing along a spectrum of symptoms, severity and time. The dynamic and multidimensional nature of psychosis is clear,² but the static labels and current diagnoses are unable to capture this complexity. New tools and approaches, such as smartphone digital phenotyping,³ offer promise in more effectively characterising these dynamic illnesses.

When approaching psychosis spectrum illnesses through a network model, comorbid conditions such as mood and anxiety disorders are not static labels but rather impermanent states that a patient may experience with varying frequency along the course of illness. These comorbidities affect one another across time, as high rates of anxiety and depression in those at risk for developing psychosis suggest these conditions may be related to the underlying psychopathology,⁵ and a wealth of cross-sectional data confirms anxiety and depression are related to psychosis severity.⁵ A network model allows psychosis symptoms to be viewed in the context of other mental health symptoms that are concurrently assessed, providing a conceptual framework for how a patient's prior symptoms may be related to future ones. For example, ecological momentary assessment (EMA) research suggests that psychosis' impact on quality of life is mediated through depression and social functioning⁶ as well as anxiety.⁷ Research also suggests that depressive symptoms may act as a moderator between psychosis and suicidality⁸ and that mood itself may predict psychosis

symptoms.⁹ Teasing apart the causal effects within networks of mental health symptoms is complex, but has strong potential for personalised and preventative psychiatric care.

Digital phenotyping

One useful tool to embrace this complexity is digital phenotyping. This method uses smartphones to capture EMA data, such as self-reported symptoms, as well as functional data from sensors embedded in mobile devices,³ such as step count and survey response time for each EMA question. For example, information from smartphone screen interactions (for example latency) can be used as a proxy for cognitive state¹⁰ and measurements of sleep/physical activity/sedentary activity can be derived from a smartphone's accelerometer.^{11,12} Additionally, this novel smartphone data are captured longitudinally, enabling the observation of temporal dynamics across different physical and social environments.

Current applications of digital phenotyping in mental health research are expanding. In schizophrenia, they have already been used to explore the relationships between anticipation and experience of pleasure¹³ as well as geolocation and mood.¹⁴ However, few studies have explored how the relationships between the variables measured with digital phenotyping tools can be coalesced into a single context – i.e. a symptom network.¹⁵ In prior work, our team has combined different digital phenotyping signals using anomaly detection to predict a specific event (relapse),^{11,16} and

another team's digital phenotyping work on suicidal ideations has suggested unique clusters of phenotypes.¹⁷ Still, there is a need to model and understand symptom interactions in their own right, without having to consider a prediction or clustering task that seeks to inform a single clinical outcome.

Generative models

Discriminative models, such as logistic regression and support vector machines, use data to discern between discrete outcomes – for example whether an exercise regimen will prevent a future stroke. However, they do not provide information regarding the underlying distribution of the data. Generative models, on the other hand, are able to provide information regarding the distribution of two or more variables, and from this distribution new samples can be generated – for example a cancer patient's oncogenomic profile can be used to predict what their profile will look like 1 month into the future. With their ability to learn the distributions of longitudinal, multivariate data, generative models may be able to elucidate the underlying relationships that characterise networks of mental health symptoms, as experienced on a moment-by-moment basis. The neuroscience research community has already realised the potential of generative models¹⁸ and has used such as a foundation for computational-based approaches to neuropsychiatry.^{19,20} However, the potential of this digital phenotyping approach is unknown and challenging to explore because few studies have offered the tools or code to enable others to reproduce experiments or data pipelines.

Study aims

In this paper we explore how generative models utilising digital phenotyping data can be used to capture unique symptom interactions in severe mental illness. Utilising open source digital phenotyping tools and fully sharable methods, we present an example of replicable research in the hope others will expand, challenge and adapt our efforts.

Method

Recruitment and participation

For this study, 47 patients in active treatment for schizophrenia were recruited from a community mental health centre in the metro Boston area, USA and from several satellite programmes that patients in care at this community health centre attend during the day. A total of 43 healthy controls were recruited from Craigslist and local colleges. Four patients were removed from the study because of study drop-out, and one patient was removed because of complete non-engagement – i.e. they completed no digital assessments. Five controls were removed from the data-set because of study drop-out. Demographic information for the remaining participants is listed in Table 1.

Clinical diagnoses of schizophrenia were confirmed with the treating psychiatrists, and healthy controls were screened for current or previous mental illness. Inclusion criteria for those with schizophrenia included: age 18 or older, in active treatment at a community mental health centre, owning a smartphone able to run the study app and the ability to sign informed consent. Comorbid illness was not an exclusion factor. Inclusion criteria for study controls were: age 18 or older, no reported mental illness (both current and prior) and owning a smartphone able to run the study app.

All procedures involving human patients were approved by both the Beth Israel Deaconess Medical Center and State of Massachusetts Department of Mental Health Institutional Review Boards. Written informed consent was obtained from all participants.

Table 1 Demographic information for patient and controls^a

| | Control group | Patient group | P |
|----------------------------|---------------|---------------|--------|
| <i>n</i> | 38 | 42 | |
| Age, years: mean (s.d.) | 32.30 (16.36) | 38.09 (14.64) | 0.1 |
| Male gender, <i>n</i> (%) | 20 (52.6) | 23 (54.8) | 1 |
| Ethnicity, <i>n</i> (%) | | | <0.001 |
| American Indian or Alaskan | 0 (0.0) | 2 (4.8) | |
| Native | | | |
| Asian | 26 (68.4) | 0 (0) | |
| Black or African American | 4 (10.5) | 15 (35.7) | |
| Multiracial or other | 3 (7.9) | 2 (4.8) | |
| White | 5 (13.1) | 23 (54.8) | |

a. Patients in active treatment for schizophrenia were recruited from several mental health community centres in the metro Boston area, and healthy controls were recruited from Craigslist and local colleges. Participants were removed from the study because of study drop-out or complete non-engagement – i.e. they completed no digital assessments. The participants who were removed are not included in the demographic information.

In the initial visit, participants were asked to take a series of surveys assessing their lifestyle and physical and mental health, including the Patient Health Questionnaire (PHQ-9),²¹ Generalized Anxiety Disorder (GAD-7),²² Social Functioning Scale,²³ Short Form Health Survey,²⁴ Behavior and Symptom Identification Scale 24,²⁵ Warning Signals Scale²⁶ and Pittsburgh Sleep Quality Index (PSQI).²⁷ For the patient group, the Brief Assessment of Cognition in Schizophrenia (BACS)²⁸ battery was performed, and the Positive and Negative Syndrome Scale (PANSS)²⁹ was used to record symptoms.

The mindLAMP smartphone app,³ a digital health platform developed by our group, was downloaded onto participants' phones. The code and instructions are available online at digitalpsych.org/lamp. Each weekday, participants were notified via the app to undertake a batch of surveys. Notifications for mood, sleep and social functioning surveys were sent to users on Mondays, Wednesdays and Fridays and notifications for anxiety, psychosis and social functioning surveys were sent on Tuesdays and Thursdays. Participants were also prompted on weekdays to complete smartphone-based cognitive assessments, modelled after the Trail Making A and B tests used in prior neuropsychiatric research.³⁰ Although symptom expression may change faster than on a day-by-day basis, previous smartphone studies also assessing longitudinal trajectories have used a similar resolution.³¹ Complete assessment information, including question content and the notification schedule, are listed in Supplementary Table 1 available at <https://doi.org/10.1192/bjo.2020.94>.

After 90 days, study participants were asked to return to the clinic for a follow-up visit, including the same surveys and batteries as at the first visit. Their participation in the study concluded at this time.

EMA data from a previous pilot study of 17 patients with schizophrenia were also used as validation data.¹¹ This study assessed mood, anxiety, psychosis and sleep via self-reported surveys similar to those in main study. Sociability and cognition were not assessed in the pilot study. Details for how this data was used to validate results from the main study are listed in the Statistical analysis section below.

Data normalisation and discretisation

Using the LAMP application programming interface,³² participant data was preprocessed so that daily values were found for each participant in every data domain. Because data was collected naturally, meaning users could engage with the app as they saw fit, there were instances where a participant completed a particular survey more than once on a given day. If this was the case, the

daily value for the given survey was set to the average of all completed surveys of that type on that day.

Data in each domain was normalised to zero mean and unit variance across the patient cohort. Using the domain means and variances derived from the patient group, the control and validation cohorts were also normalised. For each patient, normalised daily survey results were grouped into 3-day bins, as patients were expected to report symptoms in each domain once every 2–4 days (Supplementary Table 1). If there existed several results in a given domain and a given bin, the average of the results was used; if there were no events in a given domain and a given bin, imputation was performed by taking the mean of the bin directly preceding and following the bin in question; if both adjacent bins were empty, the bin was ignored.

Scores for each bin were discretised into one of two states: elevated or stable:

- elevated results are those that are equal to or greater than one s.d. above the domain mean;
- stable results are those that are less than one s.d. above the domain mean.

Although we acknowledge that this threshold has its limitations, as it does not necessarily differentiate between acute psychiatric

episodes and innocuous symptom fluctuation, this 1 s.d. threshold has been used in previous papers to dichotomise EMA results into commonly stable and uncommonly elevated states.³³

Transition probabilities

Adapting from Betzel & Bassett,¹⁸ we calculated transition probabilities for each domain (Fig. 1). We define a ‘transition’ as the elevated/stable state of an assessed domain at a specified time unit compared with the state of the domain in the subsequent timestep.

All combinations of pairwise adjacent bins were generated for all participants in the cohort, and the transition events were counted across all of these pairs, for each domain. Conditioning on the initial state in each pair, the transition events were converted into probabilities:

$$P[x_{t_1}^d = I \mid x_{t_0}^d] = \frac{\# \text{ events } (x_{t_0}^d = x_{t_0}^d \text{ AND } x_{t_1}^d = x_{t_1}^d)}{\# \text{ events } (x_{t_0}^d = x_{t_0}^d \text{ AND } x_{t_1}^d = x_{t_1}^d) + \# \text{ events } (x_{t_0}^d = x_{t_0}^d \text{ AND } x_{t_1}^d \neq x_{t_1}^d)}$$

where x_t^d is a binary elevated or stable score in domain d at timestep $t \in \{t_0, t_1\}$; x_t^d is a cohort-wide variable, being one of the two dichotomous elevated/stable states in domain d at timestep $t \in \{t_0, t_1\}$; and I is a dummy variable indicating a certain state $I \in \{\text{elevated}, \text{stable}\}$.

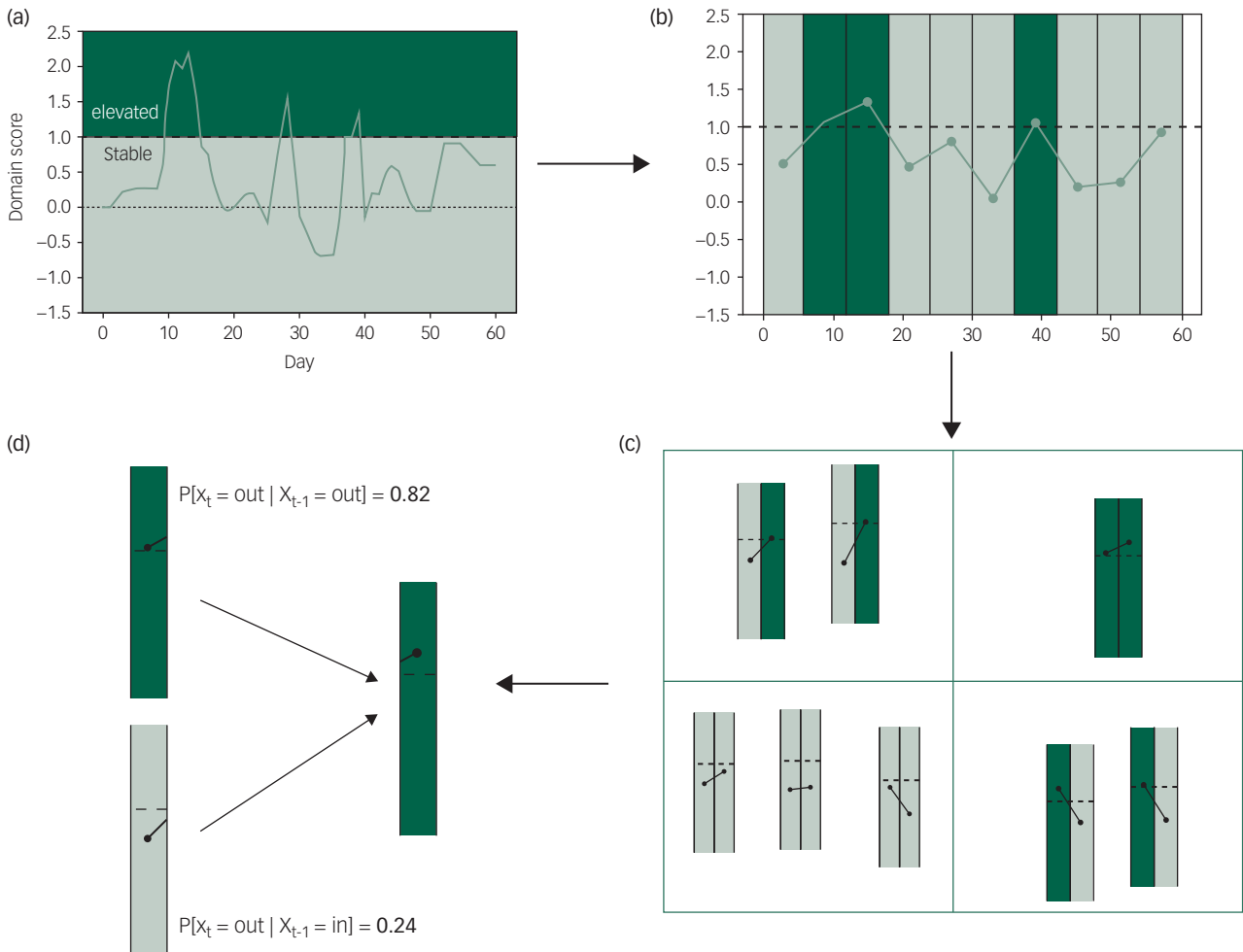


Fig. 1 Generating transition events from semi-continuous ecological momentary assessment (EMA) data.

Self-reported symptom scores were categorised as being elevated (dark green) or stable (light green) based on the predefined threshold of 1 s.d. above the study mean (a); scores were grouped into 3-day windows, and window state categorisations were determined from the mean of all scores in the respective window (b). Pairs of adjacent windows were generated, and similar pairs were grouped together based on the category of both the initial window and the next 3-day window (c). Probabilities were then calculated based on the initial (time t_0) state (d).

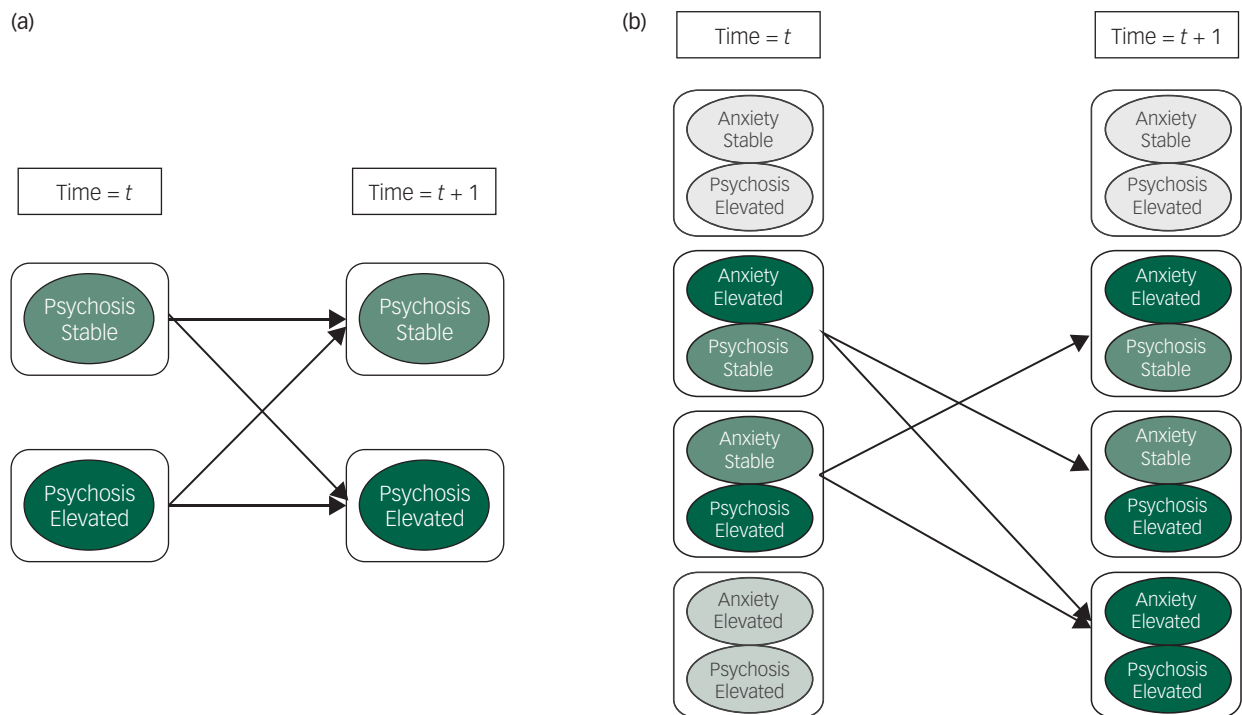


Fig. 2 Transitions between adjacent time states.

When generating transition events for a single domain (a), there are two initial domain states, from which there are each two possible paths; in the two-domain case (b), the number of initial states and possible paths per state each double, increasing the possible transitions by a factor of 4 (the total number of transition is 2^n , where n is the number of domains). By using the subset of associated transitions – those in which there exists only one elevated domain in the initial timestep, followed by an elevated score in the complement domain in the next timestep – we narrow the transition space, focusing on the most clinically relevant transitions.

The mean elevated bout duration, or the number of days in which a participant is expected to remain in an elevated state, was calculated for each domain across the cohort.

After calculating the transition probabilities within each domain (Fig. 2(a)), we calculated the transition events of domain pairs (Fig. 2(b)). By looking at transition events in a pairwise fashion, we double the number of initial states, as well as double the number of states in the $(t + 1)$ timestep. While the number of states grows exponentially with the number of domains in the joint set – 2^n , where n is the number of domains that are being observed concurrently – we reduced this set size by only looking at a subset of transition events – specifically, those in which a stable domain at time t transitions to an elevated state in the $t + 1$ time period, with the complement domain already existing in an elevated state at time t . We call these events associated transitions, as the presence of an elevated state in one domain and the occurrence of a future elevated state in a complement domain suggests that the former domain may be associated with a pathological change in the latter.

Probabilities were calculated for these associated transitions, for patients and controls. Associated transition probabilities for patients were compared with those of the validation cohort, which consisted of 17 patients from a pilot study.

Clinical subgroups

In order to move towards personalisation while still retaining enough data to find associated transitions, models were produced for subgroups of participants experiencing similar symptoms. Participants were divided into these subgroups based on the following in-clinic measures: PHQ-9 (mood/depression), GAD-7 (anxiety), PSQI Sleep Duration (sleep duration), BACS (cognition), composite PANSS (psychosis) and component PANSS (positive and negative symptoms). Participants with more severe symptoms – those

scoring at or above the cohort median for the given measure – were grouped together, and those below the cohort median were also grouped together for further analysis. Subgroups are non-exclusive, so participants can belong to any number of groups (Fig. 3). For each subgroup, mean elevated bout duration and associated transition probabilities were calculated.

Statistical analysis

When comparing mean survey scores and bout lengths, 2-sided t -tests were performed, with a 5% level of significance. Chi-squared was performed to determine the significance of associated transition probabilities, with the null hypothesis stating that the likelihood of a pathological transition in one domain is agnostic to the state of the complement domain. In order to further validate the associated transition probabilities of the patient group, we compared them with the associated transition probabilities of the validation group – i.e. the 17 patients from the pilot study. This comparison was performed with a χ^2 -test, in which a two-way table was used to classify all of the associated and ‘non-associated’ transition events of both the ‘patient’ and ‘validation’ groups. The null hypothesis stated that both the patient set and the validation set would have the same frequencies of transition events, and failure to reject the null hypothesis suggests that the frequencies of transition events does not significantly differ between the patient and validation groups.

All statistical analysis was performed with the SciPy library in Python.

Results

Self-reported survey scores were higher in patients than in controls, in every domain (Fig. 4(a)). The greatest difference occurred in self-

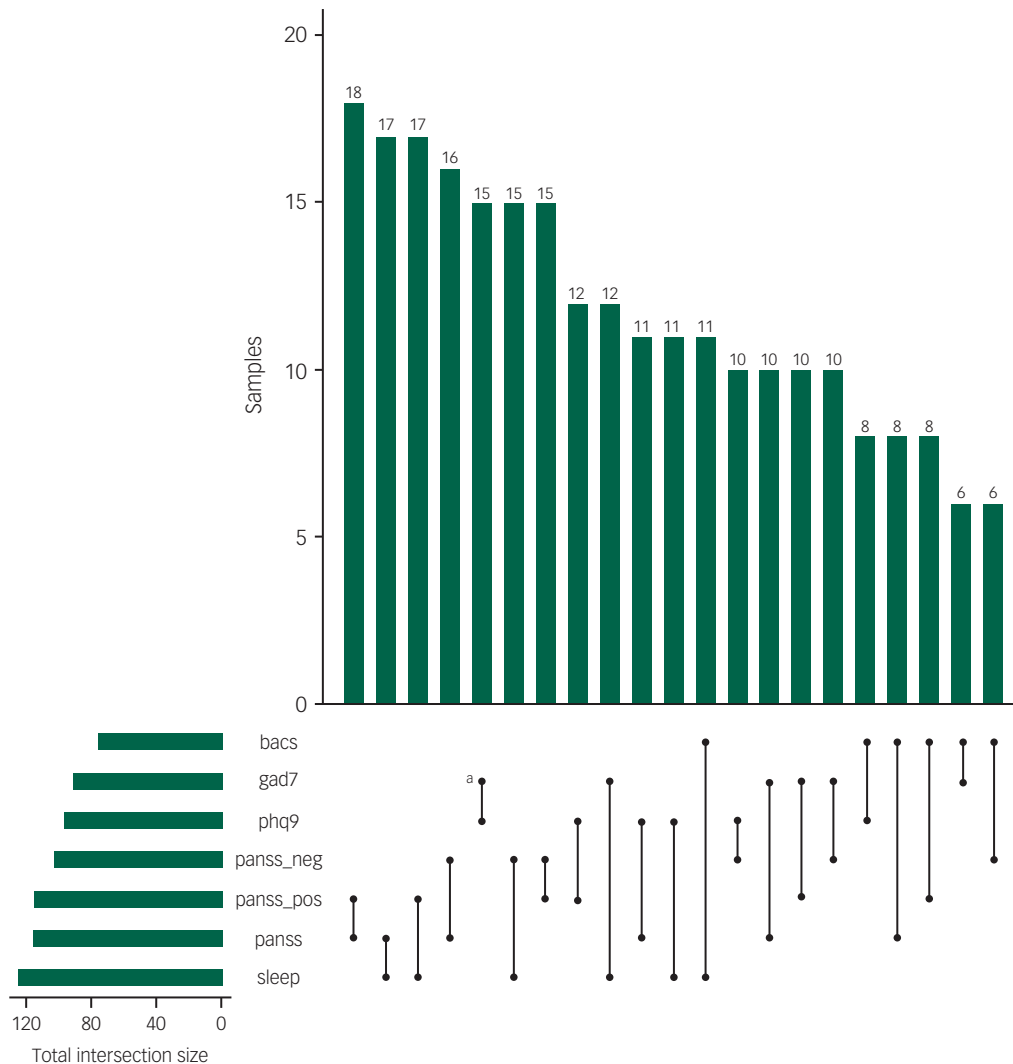


Fig. 3 Each patient could experience multiple pairs of symptoms across the study.

Instead of showing overlap with a series of venn diagrams, this Figure presents a new means to quickly look up the number of pairs and assess their relative frequency. For example, the number of patients in the anxiety–depression pairing subgroup is 15 and this is read on the Figure by looking for the two clinical measures on the horizontal axis (gad7 and phq9) and finding the thin line connecting them (marked with a superscript a). This corresponds to the bar plot that shows there were 15 participants with this pairing. bacs, Brief Assessment of Cognition in Schizophrenia; gad7, 7-item Generalized Anxiety Disorder; phq9, 9-item Patient Health Questionnaire; panss, Positive and Negative Syndrome Scale; neg, negative symptoms; pos, positive symptoms; sleep, Pittsburgh Sleep Quality Index.

reported psychosis and sleep. Self-reported survey scores were also higher in validation participants than in controls. The mean elevated bout duration of patients tended to be higher than that of controls (Fig. 4(c)); however, this difference could not be determined statistically, as there were very few (if any) elevated bouts in controls. There was no significant difference in mean elevated bout length between patients and validation participants. See Supplementary Table 1 for details on participant engagement.

There was no significant difference in attention scores between patients and controls for the easier attention task (the one modelled after Trails A smartphone cognitive assessment). However, patients performed significantly worse than controls at the more complex task, modelled after Trails B smartphone cognitive assessment (Supplementary Table 2). There were no significant associated transitions found between either of the cognitive scores and associated transitions between these cognitive measures and other self-reported symptom domains. Thus, aside from Fig. 5, we do not utilise this data further in figures and analysis.

Elevated PHQ-9, GAD-7, PANSS composite, and PANSS positive subgroups reported higher self-reported survey scores

in all five assessed domains compared with the general patient cohort (Fig. 4(b)). The poor sleep subgroup reported the lowest number of significantly elevated self-reported survey scores relative to the patients, with only anxiety and mood being elevated relative to the patient cohort. There was no significance difference in mean elevated bout length between patients and any of the subgroups (Fig. 4(d)).

For single domain transitions, psychosis had the highest probability of remaining in an elevated state in the following timestep, whereas sleep had the lowest (Supplementary Table 3). However, the psychosis domain had the lowest probability of transitioning into an elevated state.

The three highest probabilities for associated transitions were anxiety-inducing mood (0.357, $P < 0.001$), psychosis-inducing mood (0.276, $P < 0.001$) and anxiety-inducing poor sleep (0.268, $P < 0.001$), respectively (Fig. 5). The three lowest probabilities for associated transitions were psychosis-inducing social (0.09, $P < 0.02$), mood-inducing psychosis (0.162, $P < 0.001$) and sleep-inducing psychosis (0.189, $P < 0.189$). There were even lower transition probabilities for various other

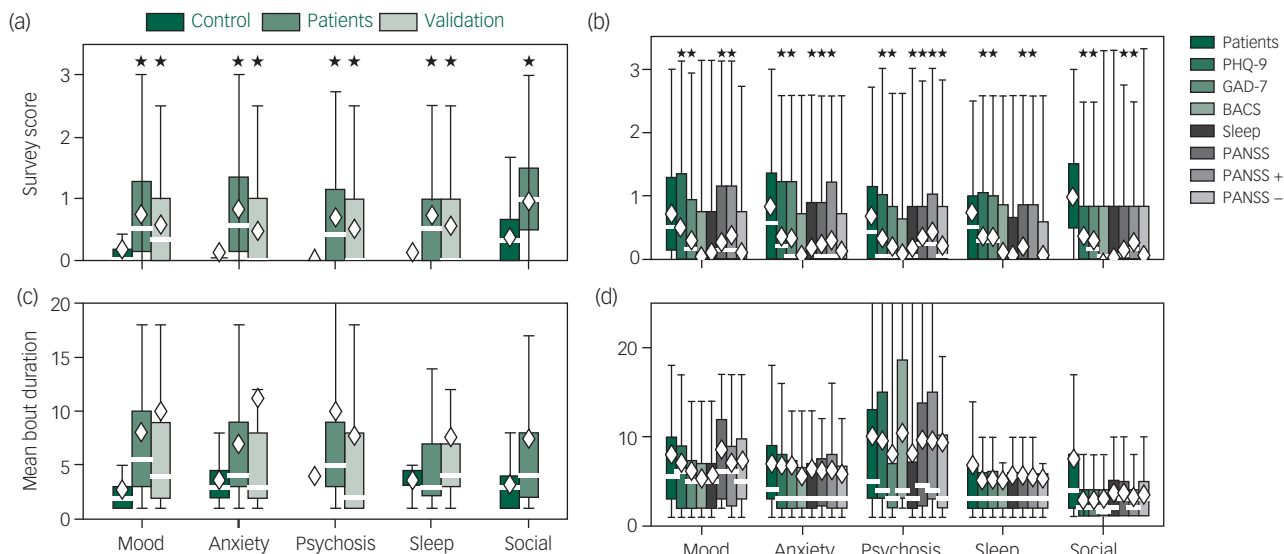


Fig. 4 Summary statistics for survey score and elevated bout duration for patient cohorts (a, c) and clinical subgroups (b, d).

Mean app-reported survey scores (a, b), indicated by the white markers, denote the average response for the specified survey domain across all participants in the cohort, with higher scores indicating more pathological responses; median scores are indicated by white lines. A reported '3' is the maximum survey score, and a '0' is the lowest. Mean elevated bout duration (c, d) indicates the average number of days for a participant to remain in an elevated state once they report an elevated score in a given domain. Asterisks on the patient cohort graphs denote significant values relative to controls; asterisks on the clinical subgroup graphs denote significant values relative to the patient cohort. PHQ-9, 9-item Patient Health Questionnaire; GAD-7, 7-item Generalized Anxiety Disorder; BACS, Brief Assessment of Cognition in Schizophrenia; Sleep, Pittsburgh Sleep Quality Index; PANSS, Positive and Negative Syndrome Scale; PANSS +, PANSS positive; PANSS -, PANSS negative. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

domain pairs – notably those including social functioning – but they were non-significant.

None of the probabilities for associated transitions in the patient set were significantly different than those in the validation set, although the validation probabilities did trend to higher values (Supplementary Table 4).

Discussion

Main findings

In this article we have shown that using digital phenotyping data, it is feasible to create generative models based on transition

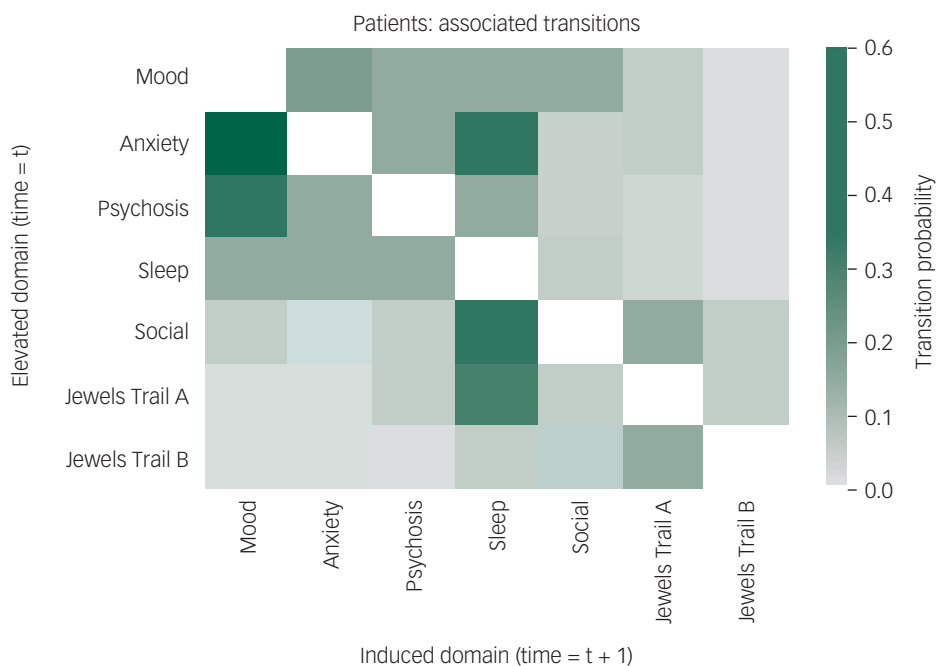


Fig. 5 Transition probabilities of associated-domain events. The left-hand axis is the elevated domain in the initial (time t) timestep, and the bottom axis is the associated domain that transitions from a stable state into an elevated state in the next (time $t + 1$) timestep.

Values along the diagonal were not included, as a domain cannot, by definition, induce itself. Transition probabilities involving cognitive domains (Jewels Trail A and Jewels Trail B) were low and non-significant; thus, they were not included in the node graphs shown in Fig. 4.

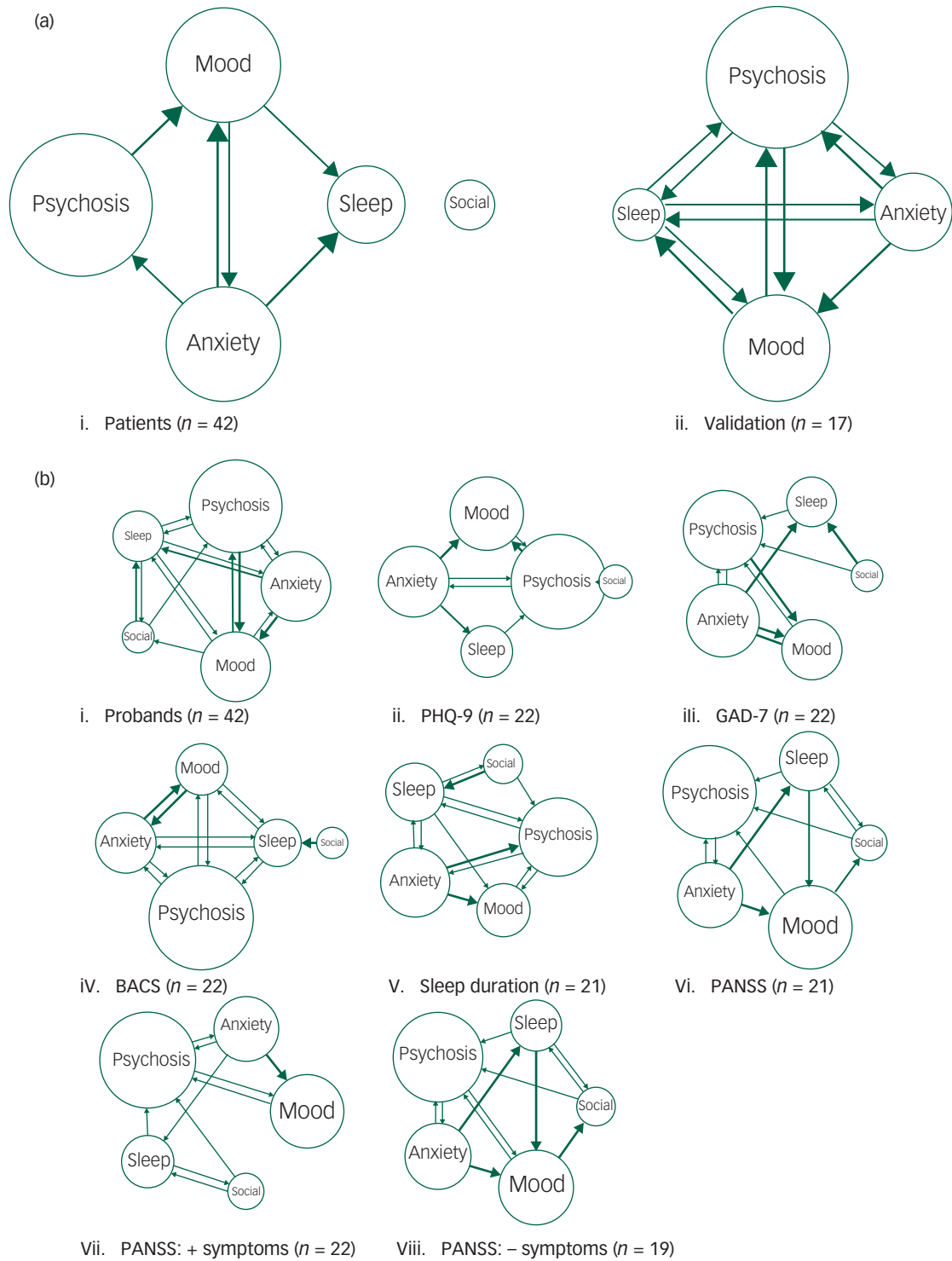


Fig. 6 Node graphs of patient (a-i) and validation cohorts (a-ii), along with clinical subgroups: all patients (b-i), Patient Health Questionnaire-9 (b-ii), Generalized Anxiety Disorder-7 (b-iii), Brief Assessment of Cognition in Schizophrenia (b-iv), Sleep duration (b-v), The Positive and Negative Syndrome Scale (PANSS) (b-vi), PANSS: positive symptoms (b-vii), and PANSS: negative symptoms (b-viii).

The node diameter represents the average bout duration (number of days) once an elevated clinical state is reached in that domain. Edge (arrow) diameter represents the probability of the target node transitioning into an elevated state in the next timestep. Only significant transition probabilities were included as edges. Edges with probabilities less than 0.2 were pruned in patient and validation graphs. There were no significant transitions in the control cohort.

probabilities for patients with psychosis. Our results suggest that patients with elevated symptoms in specific domains may experience downstream effects in other symptom domains, with anxiety-inducing mood as the highest probability of associated transitions (Fig. 5). Our finding suggests that psychosis may be

a moderator of mood symptoms, while itself being induced by anxiety, supporting a network approach towards understanding real-time and dynamic psychopathophysiology. Although the probabilities of associated transitions between either of the smartphone cognitive measures and other self-reported

symptoms were not significant, the ability to utilise multimodal data allows future studies to assess new digital measures in a clinically interpretable context.

Clinical relevance

Our results offer eventual clinical relevance in their ability to offer both patients and clinicians generative models to guide preventative psychiatric care. Clinical experience supports heterogeneity in interactions between and severity of mood, anxiety and psychosis symptoms, which are reflected in our transition probability models (Fig. 6(b)). Although the population-level results presented in this paper reflect broad interactions applicable to all patients (Fig. 6(a)), the numerous symptom-specific models (Fig. 6(b)) offer a tool that could help guide an individual's care.

For example, patients in the GAD-7 subgroup, who reported high anxiety symptoms in clinic, had a 50% higher chance of elevated mood symptoms inducing elevated anxiety compared with the cohort (Fig. 6). This conceptual framework may (a) offer a target for prevention and (b) guide treatment goals towards minimising risk of future transitions.

Although our generative models are not causal, as associated transitions do not account for other variables which may confound – for example the state of other complement symptoms; physical and social activity, measured passively; and demographic information – they offer insights beyond the correlational, cross-sectional relationships. By considering the initial state of the induced symptom domain, they account for autocorrelative effects that may confound results from other correlational models. Prior results on symptoms causality in psychosis supports our finding that both mood and anxiety have an impact on the severity of psychotic symptoms.^{4,5,7,8} While we report a range in the probabilities of associated symptoms (Fig. 5), all symptoms interactions could potentially induce elevated states in one another. This highly connected symptom network highlights the potential for creating more personalised models (Fig. 6(b)).

Although our results are novel, they are also reproducible. This is especially important as network approaches to psychopathological symptom networks are complex, and the ability to re-analyse such data often leads to novel insights and competitive interpretations.^{34,35} Thus, our results can be explored in future studies from a neuroscience perspective, as the generative models presented here may offer targets for generative models of the brain proposed in the Bayesian brain hypothesis³⁶ and structural as well as functional brain networks. These digital phenotyping-derived generative models could thus serve as a bridge between units of analysis in the National Institute of Mental Health's Research Domain Criteria model,³⁷ especially self-reported symptoms/behaviour and circuits.

Ethics

As a clinical tool, digital phenotyping offers potential advantages as well as ethical considerations that cannot be ignored. The ability to collect data remotely from a patient's smartphone has become more relevant during the coronavirus disease 2019 (COVID-19) pandemic and the increasingly digital world of mental healthcare. The generative models built on this data add clinical utility by helping clinicians understand symptom patterns and, potentially, guide preventative care. However, before offering any new models or data in a clinical setting, one must take into account workflow considerations so as not to overburden clinicians or patients and impede care. Issues around clinical, legal, and ethical duties to respond to real-time data such as that generated in this study remain in flux. We suggest that use of this data may be best utilised in the context of an entirely new clinical workflow and clinic designed to support the integration of digital data – the digital

clinic. Our team has outlined our experiences and implementation of this new pilot clinic³⁸ and how we build trust and therapeutic alliance when using the mindLAMP app.

As digital phenotyping becomes commonly integrated into remote and in-person care, it is important to consider the perspective and ethical concerns of the primary stakeholder – i.e. the patient – who is meant to benefit from digital phenotyping.

One concern may be that remote observation presents privacy risks, as the data collected remotely via digital phenotyping may be used unethically. This risk is dependent on the type of data being collected. The cognitive monitoring and use of surveys in this study present less privacy risks than other sensors, such as GPS (global positioning system). In the care setting, these ethical considerations require including patient input directly in the development of these tools and the usage of this data. For example, in developing the mindLAMP app, we worked with patients to understand their needs around trust, control and community.³ In our digital clinic, where we use mindLAMP to augment care, we offer it as an adjunctive tool and work with patients to explain why they can trust it and how they control their data, and to co-create the community with whom they want to share their data (often just the clinician, but at times family). Further details on the ethical use of mobile health technology in clinical care are outlined in our prior works.^{39–42}

Limitations

In our model, associated state changes were derived from symptoms pair at 3-day intervals, offering a window large enough to offer a valid transition probability but still small enough to enable early clinical intervention. Our choice of pairwise symptoms enables us to move towards causality while allowing the models to remain clinically interpretable. These pairwise symptom interactions are not independent and may occur concurrently, which could mask more complex symptom interactions. However, a model that utilises more specific combinations of symptoms may not be feasible without a substantially large amount of data. Defining relevant states that are both data-driven and clinically actionable will be necessary for feasible generative models.

Results from validation on a distinct data-set of 17 different patients with schizophrenia show similarity in the duration of elevated symptoms and the interactions connecting them (Figs 5(c) and 6). Findings that transitional probabilities for the training set tend to be lower than those of the validation set may be related to the small size of the latter. Definitions for elevated self-reported clinical states were derived from retrospective data collection, given the lack of prior research using these methods.

Future directions

In this study we focused on active data from surveys as well as passive data from cognitive measures to generate clinically interpretable and actionable models. Future efforts with prospective methods will expand this work to include more multivariate data with sensors that have been used in previous digital phenotyping studies, such as passively derived physical activity,⁴³ GPS⁴⁴ and sleep measurements.⁴⁵

Although the generative models presented here are feasible and clinically relevant, other models may further elucidate symptom interactions in severe mental illness. In generating associated transitions, we consider only a 3-day time lag. However, mental health symptoms may affect future symptoms at varying timescales, and these effects, as a function of the time lag, may be nonlinear and non-stationary. A hidden Markov model, another type of generative model that works on discrete states, may better learn the longitudinal nature of the data in its entirety, as it considers more than

just adjacent pairwise events. An autoregressive model, such as the autoregressive integrated moving average, may characterise the time-lagged effect that a symptom may have on itself – i.e. the ‘memory’ of symptoms. Future efforts will implement these models to provide deeper insight into severe mental illness as quantified through digital phenotyping.

In our clinical subgroups, we clustered patients based on similar symptomology, retaining clinical interpretability. However, there are other methods of clustering that may better characterise patients’ disease states and thus move closer toward personalisation. For example, latent class analysis clusters patients by transforming observed, co-dependent variables – such as mental health symptoms – into underlying classes, which are independent of one another.⁴⁶ These classes may be used as a new method for diagnostic labeling,^{47,48} offering a data-driven approach to psychiatric nosology.

Digital phenotyping and probabilistic transitional models based on associated state change offers a feasible means to approach network interactions of symptoms in a clinically actionable manner. Further research with larger sample sizes using the foundation and sharable tools outlined in this article offers a method towards generating models personalised to each patient, supporting preventative care and improving clinically relevant knowledge.

Ryan Hays, BS, Harvard Medical School, Department of Psychiatry, Beth Israel Deaconess Medical Center, USA; **Matcheri Keshavan**, MD, Harvard Medical School, Department of Psychiatry, Beth Israel Deaconess Medical Center, USA; **Hannah Wisniewski**, BS, Harvard Medical School, Department of Psychiatry, Beth Israel Deaconess Medical Center, USA; **John Torous** , MD, MBI, Harvard Medical School, Department of Psychiatry, Beth Israel Deaconess Medical Center, USA

Correspondence: John Torous. Email: jtorous@bidmc.harvard.edu

First received 22 Apr 2020, final revision 6 Aug 2020, accepted 17 Aug 2020

Supplementary material

Supplementary material is available online at <http://doi.org/10.1192/bjo.2020.94>

Data availability

The data that support the findings of this study are available from J.T. upon reasonable request.

Acknowledgements

The authors would like to thank the many people who have helped test and provide input into the LAMP app. Without their efforts and feedback this work would not have been possible.

Author contributions

R.H. contributed to data analysis, figures, participant recruitment and manuscript writing. M.K. contributed to framework conceptualisation, question formulation and writing. J.T. contributed question formulation, analysis, participant recruitment and manuscript writing.

Funding

This work was supported by a NIMH Mentored Patient-Oriented Research Career Development career training award to J.T. (1K23MH116130-02) and a Young Investigator grant to J.T. from the Brain and Behavior Research Foundation.

Declaration of interest

J.T. reports unrelated research support from Otsuka. ICMJE forms are in the supplementary material, available online at <https://doi.org/10.1192/bjo.2020.94>

References

- Silbersweig D, Loscalzo J. Precision psychiatry meets network medicine: network psychiatry. *JAMA Psychiatry* 2017; **74**: 665–6.
- van Rooijen G, van Rooijen M, Maat A, Vermeulen JM, Meijer CJ, Ruhe HG, et al. The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. *Schizophr Res* 2018; **208**: 229–44.
- Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *J Technol Behav Sci* 2019; **4**: 73–85.
- Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK. Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. *Schizophr Bull* 2014; **40**: 120–31.
- Hartley S, Barrowclough C, Haddock G. Anxiety and depression in psychosis: A systematic review of associations with positive psychotic symptoms. *Acta Psychiatr Scand* 2013; **128**: 327–46.
- van Rooijen G, van Rooijen M, Maat A, Vermeulen JM, Meijer CJ, Ruhe HG, et al. Longitudinal evidence for a relation between depressive symptoms and quality of life in schizophrenia using structural equation modeling. *Schizophr Res* 2019; **208**: 82–9.
- Huppert JD, Smith TE. Anxiety and schizophrenia: the interaction of subtypes of anxiety and psychotic symptoms. *CNS Spectr* 2005; **10**: 721–31.
- van Rooijen G, Isvoranu A-M, Kruijt OH, van Borkulo CD, Meijer CJ, Wigman JTW, et al. A state-independent network of depressive, negative and positive symptoms in male patients with schizophrenia spectrum disorders. *Schizophr Res* 2018; **193**: 232–9.
- Krabbendam L, Myin-Germeys I, Hanssen M, de Graaf R, Vollebergh W, Bak M, et al. Development of depressed mood predicts onset of psychotic disorder in individuals who report hallucinatory experiences. *Br J Clin Psychol* 2005; **44**: 113–25.
- Liu G, Henson P, Keshavan M, Pekka-Onnela J, Torous J. Assessing the potential of longitudinal smartphone based cognitive assessment in schizophrenia: a naturalistic pilot study. *Schizophr Res Cogn* 2019; **17**: 100144.
- Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J-P. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology* 2018; **43**: 1660–6.
- Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. *NPJ Schizophr* 2017; **3**: 1–6.
- Edwards CJ, Cella M, Emsley R, Tarrier N, Wykes THM. Exploring the relationship between the anticipation and experience of pleasure in people with schizophrenia: an experience sampling study. *Schizophr Res* 2018; **202**: 72–9.
- Fraccaro P, Beukenhorst A, Sperrin M, Harper S, Palmier-Claus J, Lewis S, et al. Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: a systematic review. *J Am Med Inform Assoc* 2019; **26**: 1412–20.
- Lydon-Staley DM, Barnett I, Satterthwaite TD, Bassett DS. Digital phenotyping for psychiatry: accommodating data and theory with network science methodologies. *Curr Opin Biomed Eng* 2019; **9**: 8–13.
- Barnett I, Torous J, Staples P, Keshavan M, Onnela JP. Beyond smartphones and sensors: choosing appropriate statistical methods for the analysis of longitudinal data. *J Am Med Inform Assoc* 2018; **25**: 1669–74.
- Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. *Depress Anxiety* 2018; **35**: 601–8.
- Frässle S, Yao Y, Schöbi D, Aponte EA, Heinzle J, Stephan KE. Generative models for clinical applications in computational psychiatry. *Wiley Interdiscip Rev Cogn Sci* 2018; **9**: e1460.
- Betzel RF, Bassett DS. Generative models for network neuroscience: prospects and promise. *J R Soc Interface* 2017; **14**: 20170623.
- Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* 2016; **19**: 404.
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–13.
- Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; **166**: 1092–7.
- Birchwood M, Smith JO, Cochrane R, Wetton S, Copestake SO. The social functioning scale: the development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. *Br J Psychiatry* 1990; **157**: 853–9.
- Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992; **30**: 473–83.
- Cameron IM, Cunningham L, Crawford JR, Eagles JM, Eisen SV, Lawton K, Naji SA, Hamilton RJ. Psychometric properties of the BASIS-24© (Behavior and Symptom Identification Scale–Revised) mental health outcome measure. *Int J Psychiatry Clin Pract* 2007; **11**(1): 36–43.
- Jørgensen P. Schizophrenic delusions: the detection of warning signals. *Schizophr Res* 1998; **32**: 17–22.
- Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989; **28**(2):193–213.

- 28 Keefe RS, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res* 2004; **68**:283–97.
- 29 Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 1987; **13**: 261–76.
- 30 Hays R, Henson P, Wisniewski H, Hendel V, Vaidyam A, Torous J. Assessing cognition outside of the clinic: smartphones and sensors for cognitive assessment across diverse psychiatric disorders. *Psychiatr Clin* 2019; **42**: 611–25.
- 31 Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MS, et al. CrossCheck: integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatr Rehabil J* 2017; **40**: 266.
- 32 Vaidyam A, Halamka J, Torous J. Actionable digital phenotyping: a framework for the delivery of just-in-time and longitudinal interventions in clinical health-care. *mHealth* 2019; **5**: 25.
- 33 Johns JT, Di J, Merikangas K, Cui L, Swendsen J, Zupunnikov V. Fragmentation as a novel measure of stability in normalized trajectories of mood and attention measured by ecological momentary assessment. *Psychol Assess* 2019 Mar; **31**: 329–39.
- 34 Forbes MK, Wright AG, Markon KE, Krueger RF. Evidence that psychopathology symptom networks have limited replicability. *J Abnorm Psychol* 2017; **126**: 969.
- 35 Borsboom D, Fried EI, Epskamp S, Waldorp LJ, van Borkulo CD, van der Maas HL, Cramer AO. False alarm? A comprehensive reanalysis of “Evidence that psychopathology symptom networks have limited replicability” by Forbes, Wright, Markon, and Krueger. *J Abnorm Psychol* 2017; **126**: 989–99.
- 36 Friston K. The history of the future of the Bayesian brain. *NeuroImage* 2012; **62**: 1230–3.
- 37 Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Transl Psychiatry* 2017; **7**: e1053.
- 38 Rodriguez-Villa E, Rauseo-Ricupero N, Camacho E, Wisniewski H, Keshavan M, Torous J. The Digital Clinic: Implementing Technology and Augmenting Care for Mental Health. *Gen Hosp Psychiatry* 2020; **66**: 59–66.
- 39 Torous J, Roberts LW. The ethical use of mobile health technology in clinical psychiatry. *J Nerv Ment Dis* 2017; **205**: 4–8.
- 40 Torous J, Nebeker C. Navigating ethics in the digital age: introducing connected and open research ethics (CORE), a tool for researchers and institutional review boards. *J Med Internet Res* 2017; **19**: e38.
- 41 Torous J, Ungar L, Barnett I. Expanding, augmenting and operationalizing ethical and regulatory considerations for using social media platforms in research and health care. *Am J Bioeth* 2019; **19**(6): 4–6.
- 42 Nebeker C, Torous J, Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med* 2019; **17**: 137.
- 43 Panda N, Solsky I, Huang EJ, et al. Using smartphones to capture novel recovery metrics after cancer surgery. *JAMA Surg* 2019; **155**: 1–7.
- 44 Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *NPJ Digit Med* 2018; **1**: 15.
- 45 Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *J Med Internet Res* 2019; **21**: e11029.
- 46 Pignon B, Peyre H, Szöke A, Geoffroy PA, Rolland B, Jardri R, et al. A latent class analysis of psychotic symptoms in the general population. *Aust N Z J Psychiatry* 2018; **52**: 573–584.
- 47 Essau CA, de la Torre-Luque A. Comorbidity profile of mental disorders among adolescents: A latent class analysis. *Psychiatry Res* 2019; **278**: 228–34.
- 48 Kongsted A, Nielsen AM. Latent class analysis in health research. *J Physiother* 2017; **63**: 55–8.

EXTRA
CONTENT
ONLINEOPEN
ACCESS