Neuroscience and Biobehavioral Reviews 152 (2023) 105259

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Insights from personalized models of brain and behavior for identifying biomarkers in psychiatry

Brian Kraus [a,*], Richard Zinbarg [a,b], Rodrigo M. Braga [c], Robin Nusslock [a,d], Vijay A. Mittal [a,d,e,f,g], Caterina Gratton [a,c,h,i,j]

[a] Department of Psychology, Northwestern University, Evanston, IL, USA
[b] The Family Institute at Northwestern University, Evanston, IL, USA
[c] Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[d] Institute for Policy Research, Northwestern University, Evanston, IL, USA
[e] Institute for Innovations in Developmental Sciences (DevSci), Northwestern University, Chicago, IL, USA
[f] Northwestern University, Department of Psychiatry, Chicago, IL, USA
[g] Northwestern University, Medical Social Sciences, Chicago, IL, USA
[h] Interdepartmental Neuroscience Program, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[i] Department of Psychology, Florida State University, Tallahassee, FL, USA
[j] Program in Neuroscience, Florida State University, Tallahassee, FL, USA

ARTICLE INFO

Keywords:
Biomarkers
Personalized Models
Precision Imaging
Psychopathology
Functional Connectivity
fMRI
Psychiatry

ABSTRACT

A main goal in translational neuroscience is to identify neural correlates of psychopathology ("biomarkers") that can be used to facilitate diagnosis, prognosis, and treatment. This goal has led to substantial research into how psychopathology symptoms relate to large-scale brain systems. However, these efforts have not yet resulted in practical biomarkers used in clinical practice. One reason for this underwhelming progress may be that many study designs focus on increasing sample size instead of collecting additional data within each individual. This focus limits the reliability and predictive validity of brain and behavioral measures in any one person. As biomarkers exist at the level of individuals, an increased focus on validating them within individuals is warranted. We argue that personalized models, estimated from extensive data collection within individuals, can address these concerns. We review evidence from two, thus far separate, lines of research on personalized models of (1) psychopathology symptoms and (2) fMRI measures of brain networks. We close by proposing approaches uniting personalized models across both domains to improve biomarker research.

1. Background

In translational neuroscience, it has long been a stated goal to identify "biomarkers", or neural correlates of psychiatric disorders that can be used to facilitate diagnosis, determine prognosis, and improve treatment (Charney et al., 2002). This goal is primarily inspired by the shortcomings of currently used diagnostic nosologies in psychiatry (American Psychiatric Association, 2013; World Health Organization, 1992), in which symptoms of disorders often overlap between diagnoses and treatment outcomes and are highly variable even for people who receive the same diagnosis (Gordon and Redish, 2016). One example of this issue is seen in the criteria for diagnosing major depressive disorder and generalized anxiety disorder, which both include the same symptoms of fatigue and sleep disturbance (American Psychiatric Association, 2013). Partially in response to these shared diagnostic features across different disorders, the identification of biomarkers is the central objective of the Research Domain Criteria (RDoC) initiative: a large effort in clinical neuroscience to identify the neural underpinnings of psychopathology symptoms to better define diagnostic categories in psychiatry (Cuthbert and Insel, 2013; Insel, 2014). Through this and similar endeavors, clinical neuroscience has focused on finding individual differences in brain function which reliably covary with psychiatric symptoms.

One method commonly employed in biomarker research is magnetic resonance imaging (MRI), which can be used to non-invasively measure human brain anatomy and function and track their relationship with symptom severity. Evidence from MRI research has shown that symptoms of psychopathology are not only associated with alterations in the

* Corresponding author.
E-mail address: btkraus@u.northwestern.edu (B. Kraus).

https://doi.org/10.1016/j.neubiorev.2023.105259
Received 1 March 2023; Received in revised form 22 May 2023; Accepted 30 May 2023
Available online 1 June 2023

functions of single brain regions, but can also result from the interaction of many distributed areas of the brain (Menon, 2011). Given this evidence, functional connectivity MRI (fcMRI) has become a prominent method to study psychopathology. Research utilizing fcMRI has been used to identify "networks" of brain regions with correlated activity (Power et al., 2011; Yeo et al., 2011), such as the default mode network (Greicius et al., 2003). Using this technique, many studies have found links between alterations of brain networks and symptoms of psychiatric disorders (Baker et al., 2014; Zhang et al., 2021). For example, evidence from many studies has found a relationship between increased connectivity strength among the regions of the default mode network and symptoms of depression (Kaiser et al., 2015; Mulders et al., 2015). Together, these findings highlight the presence of distributed alterations in brain function in psychopathology.

Yet, despite its promise, biomarker research using MRI has so far yielded predominantly small brain-behavior relationships that can only be measured reliably with thousands of participants (Marek et al., 2022). These somewhat underwhelming results from investments in large sample size datasets such as the Human Connectome Project (HCP; Van Essen et al., 2012), UK Biobank (Miller et al., 2016), and Adolescent Brain Cognitive Development (ABCD) study (Casey et al., 2018) suggest that the typical study designs used in biomarker research may not be ideal for identifying actionable targets for diagnosis and intervention. Here, we posit that a major limitation of these previous efforts is that they largely rely on group-level or cross-sectional inferences based on a relatively small amount of data per participant.

In this review, the term "group-level" broadly refers to methods of characterizing data from individuals which explicitly average measures across people (e.g., comparing mean scores on a measure between groups of individuals). Here, we also use this term to include methods of analyzing data which artificially equate measures across individuals to meet certain assumptions of statistical analyses. Common examples of this assumption are that the same sum scores on a self-report measure of depression symptoms represent the same level of psychopathology in two different people (Fried and Nesse, 2015), or that the functions of brain regions defined in groups of individuals generalize to any one individual (Gordon and Nelson, 2021; Salvo et al., 2021). Though these examples (and other similar methods) are not explicitly averaging data across people, their assumption of equivalence likely distorts our understanding of the true degree of inter-individual variability in these measures in the population.

As we will discuss further in this review, with the amount of data typically collected per participant in group-level studies, measures of both brain networks and behavior have poor reliability and fail to capture valid person-specific brain and behavioral profiles. This means that the majority of research on biomarkers cannot, by design, identify reliable relationships between symptoms and neural substrates *within a given individual*. Thus, these group-level study designs can only indirectly infer that the relationships they observe actually hold at the level of individuals – the level of analysis where biomarkers are theorized to exist.

As an example, consider the aforementioned meta-analytic link between depression symptoms and default mode network connectivity. Though group-level approaches can answer research questions about people *on average*, they leave potentially relevant aspects of individual-level functioning unaddressed (see Table 1). This is true for research questions using group-level analyses in both between-subject (e.g., is the average score on a measure between groups significantly different?) and within-subject (e.g., is the average score on a measure at timepoint one significantly different from the average score at timepoint two?) study designs. In order to provide actionable indicators of prognosis and treatment, biomarker research must be able to answer questions such as *who* specifically benefits from a given treatment and after *how long* do they begin to improve? And *which* symptom profiles distinguish these individuals from those who show little or no benefit from a treatment? Without directly answering these questions at the level of individuals, the generalizability of group-level results to any individual person is unknown.

Table 1
A subset of potential research questions relating to the relationship between default network connectivity and depression symptoms are shown. Examples of questions that can be answered using an individual-level approach versus a more typical group-level approach using are provided for within and between subject study designs.

Study Design	Group-Level Analysis	Individual-Level Analysis
Between Subjects	• Does a group of depressed individuals show more default network connectivity than a group of non-depressed individuals *on average*? • Is there a relationship between individual differences in default network connectivity and depression symptoms across people *on average*?	• Are there individuals with depression who show stronger default mode network connectivity relative to non-depressed individuals? • Does the relationship between default network connectivity and depression symptoms vary across individuals? • Who specifically benefits from a given intervention, and what individual characteristics distinguish them from those who do not benefit?
Within Subjects	• Do changes in depression symptoms over time covary with changes in default network connectivity *on average*? • Does a given intervention change the relationship between default network connectivity and depression symptoms *on average*?	• How does a given person's default mode network connectivity vary over time with changes in depression symptoms? • How long does it take for the relationship between default network connectivity and depression symptoms to begin to change in an individual in response to a given intervention?

Given these limitations of group-level analyses in biomarker research, it is fair to question whether the goals of precision diagnosis and treatment outlined in the RDoC framework (Cuthbert and Insel, 2013) can be realized using typical group-level biomarker study designs (i.e., Hariri, 2009). In response to these limitations, we propose that biomarker research should, instead, focus on the individual as the primary unit of analysis. One potential methodological approach for this focus is that of personalized (or "precision") models: defined here as estimates of brain function and behavior obtained from extensively sampled individual participants (Gratton et al., 2020; Wright and Woods, 2020). Previous evidence suggests that personalized models can produce data that is reliable and valid at the level of individual participants without relying on pooling data across individuals. Thus, extensive measurement of individual participants can directly address heterogeneity in the within-person mechanisms that are proposed to underlie psychopathology and brain function.

In this review, we focus on recent efforts to construct personalized models in the domains of psychopathology symptoms and fcMRI. This is a narrative review and thus the selection of included papers is not exhaustive, but is meant to highlight the potential of personalized models in biomarker research. While typically the use of personalized models in these domains has been mutually exclusive, we discuss similar overarching themes across these two research domains and propose strategies for uniting them in future research. Though the focus here is explicitly on measures of psychopathology and fcMRI, the issues covered here are relevant to many other biological measures and behavioral phenotypes. The first section of this review focuses on evidence that

averaging data across people via group-level approaches obscures important information about individual differences in brain networks and psychiatric symptoms. The second section focuses on the insights personalized models can provide into how psychiatric symptoms and brain networks change dynamically change over time. In the final section, we propose future directions and study designs for research that combines personalized approaches in both neuroimaging and psychopathology.

2. Averaging data across participants can obscure individual differences in brain and behavior

Many traditional group-level experimental designs seek to determine whether there are significant differences between particular groups or relationships between certain variables in a population (Whitley and Kite, 2012). This is the most common design adopted in many studies of psychopathology symptoms and fcMRI, including the large consortia mentioned in the introduction. While in many contexts this across-person aggregation has desirable effects, such as reducing measurement error and increasing statistical power (Cohen, 1992), these benefits are dependent on the underlying properties of the data being averaged. In situations where substantial inter-individual differences exist in the underlying data, pooling data across individuals can actually result in decreased validity and/or reliability (Epstein, 1983). Thus, evidence for consistent and substantial individual differences in psychopathology and fcMRI measures would pose a problem for biomarker research which aggregates or artificially equates these measures across individuals.

In this section, we review studies using personalized models of fcMRI and psychopathology to examine the extent of individual differences in these measures across participants. We will discuss issues of inter-individual heterogeneity and validity that prevent traditional group-level approaches from providing accurate representations of these measures in individual people. We argue this evidence indicates that biomarker research would benefit from a greater focus on measuring brain function and psychopathology at the individual-level.

2.1. Personalized models of functional connectivity can better distinguish how brain organization varies across individuals

In the field of neuroscience, non-invasive neuroimaging research into brain structure and function has helped elucidate the neural bases of cognition, emotion, and behavior (Gazzaniga, 2004). As these processes have long been known to be altered in psychopathology (Beck, 1979; Kraepelin, 1921), better understanding their underlying mechanisms can help identify the neural substrates that drive mental illness. More recently, it has become clear that the neural substrates underlying these internal states and behaviors are not solely reliant on individual regions of the brain, but are the product of interactions across distributed large-scale brain networks (Dosenbach et al., 2007; Van Den Heuvel and Pol, 2010). In line with this evidence, many studies of large-scale brain networks in clinical populations have identified significant relationships with psychopathology (Baker et al., 2014; Menon, 2011; Zhang et al., 2021).

Previous work in neuroimaging has established that using fcMRI, brain networks can be non-invasively measured in human participants (Van Dijk et al., 2010). This method can be used to measure networks such as the default mode (Greicius et al., 2003) and frontoparietal (Dosenbach et al., 2006, 2007), which have been shown to be involved in processes such as internally-oriented cognition and cognitive control, respectively. Many studies have replicated the presence of these networks across different groups of participants (e.g., Power et al., 2011; Yeo et al., 2011), establishing that their topography is relatively consistent when (fcMRI) data is averaged across groups of participants. This past work suggests that there are many commonalities in functional brain organization across the population.

Despite these many commonalities, recent evidence from "precision" fMRI has helped elucidate the many ways in which brain network organization varies across individuals (Gordon and Nelson, 2021; Gratton et al., 2020; Michon et al., 2022). Precision fMRI refers to personalized models of brain activity and brain networks obtained via extensive sampling[1] of single individuals (Gordon et al., 2017c; Greene et al., 2019; Laumann et al., 2015; Marek et al., 2018). This extensive sampling is necessary to obtain reliable estimates of individual brain function, as with smaller amounts of data fcMRI estimates can be quite noisy (Gordon et al., 2017c; Kraus et al., 2021; Noble et al., 2017). In many previous studies using fcMRI, this limitation has necessitated group-level approaches for defining brain networks in order to produce robust results.

This limitation is notable as many recent fcMRI studies have shown that, while there are commonalities in their topography, some brain network locations show large idiosyncratic variations in individuals (Bijsterbosch et al., 2018; Braga and Buckner, 2017; Finn et al., 2015; Gordon et al., 2017b; Gordon et al., 2017a; Gordon et al., 2017c; Gratton et al., 2018; Greene et al., 2019; Kong et al., 2019, 2021; Laumann et al., 2015; Marek et al., 2018; Mueller et al., 2013; Seitzman et al., 2019). In fact, *every individual* measured so far has exhibited reliable and substantial deviations in fcMRI that are not reflected in group-defined functional atlases or network topography estimates (Dworetsky et al., 2021; Seitzman et al., 2019). These areas of idiosyncratic variation have been shown to be reliable when large amounts of fcMRI data are collected from each individual, indicating that they are not a result of measurement error (Seitzman et al., 2019). Furthermore, they do not appear to be attributable to gross anatomical differences in brain structure across individuals (Gordon et al., 2017a), motion artifacts (Gordon et al., 2017c), or errors in registering fMRI data to a common brain template (Seitzman et al., 2019). Altogether, this suggests that reliable idiosyncratic differences in brain networks are present across individuals.

The presence of these individual differences in the spatial layout of brain networks has important implications for interpreting measures of fcMRI. This is because in most group-level fcMRI analyses, the same coordinate seed region(s) (or functional parcellation, e.g., the 264 regions identified in Power et al., 2011) are selected for all participants in a dataset. In using these methods to calculate functional connectivity, it is assumed that regions of the brain are functionally analogous in their spatial layout across participants. Under this assumption in group-level analyses, individual differences in fcMRI are then typically interpreted as variations in connectivity strength between a specified set of regions.

However, as brain organization is variable across individuals (e.g., white circles in Fig. 1A versus Fig. 1B), this assumption is likely violated in group analyses in practice (Fedorenko, 2021). This is because differences in the strength of fcMRI between regions are not the only plausible explanation for any observed differences, as individual variations in brain network topography can also alter functional connectivity estimates (Gordon and Nelson, 2021). For example, while the inferior parietal lobe (i.e., white circles in Fig. 1) is often considered a canonical region of the default mode network, in any one individual this region may not actually be strongly coupled to the default network (Dworetsky et al., 2021). Additionally, the degree to which functional network topography differs between individuals varies by brain network, further complicating the interpretation of group-level studies (Finn et al., 2015; Gordon et al., 2017c; Gratton et al., 2018; Kong et al., 2019; Seitzman et al., 2019). Thus, typical estimates of fcMRI strength are not able to

[1] Though "precision" fMRI is typically defined as > 120 min of resting state data collected over multiple sessions (Gratton et al., 2020), here extensive data collection refers to > 40 min of fMRI data collected over multiple sessions (for examples of both types of datasets, see Gratton et al., 2020). However, the exact amount of data needed to produce reliable estimates of functional brain networks will depend on multiple parameters, including the anatomical location being measured, MRI scan parameters, and the fcMRI measure in question (see footnote 3 in Section 3.1).

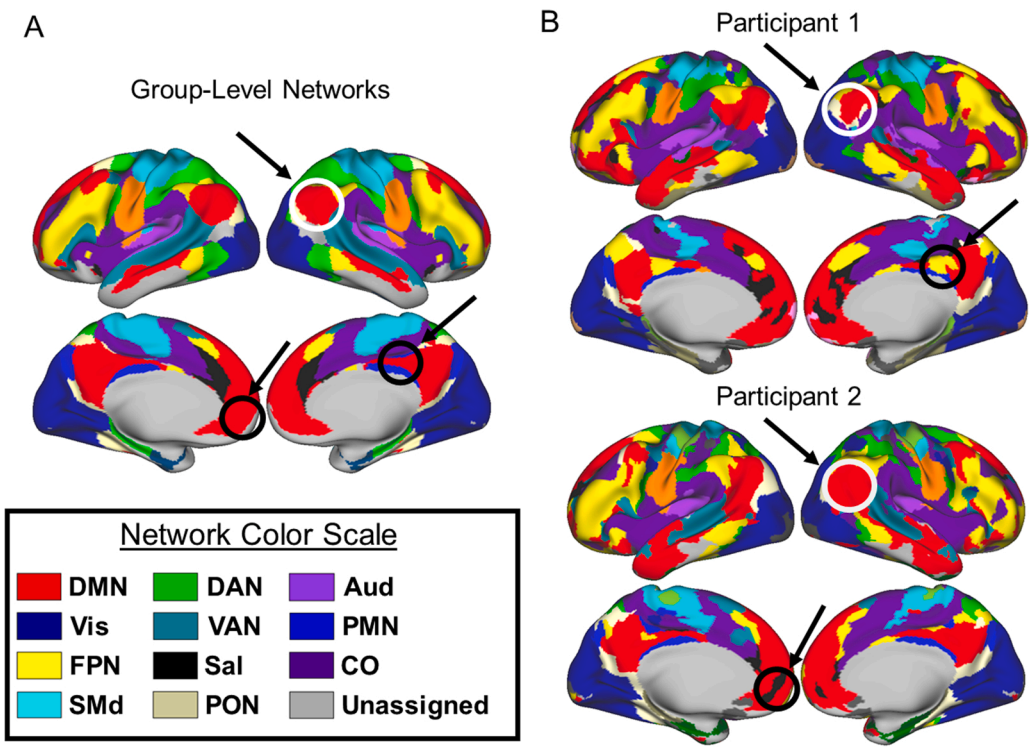

Fig. 1. An example of the differences in individual (Participants 1 and 2; B) brain network organization when compared to a group-level estimate (A; Power et al., 2011) are shown. Examples of regions where brain network topography differs across individuals are shown in white circles, and regions where the topography differs in each individual from the group-level networks are shown in black circles. The colors which represent each network are shown in the bottom left.
The images in this figure are based on those from Gordon et al. (2017c).

dissociate whether observed results are due to individual differences in the spatial extent of brain networks or the degree of functional coupling between analogous regions.

To illustrate this issue with real data, an example is shown from two participants (Participant 1 and Participant 2) from the Midnight Scan Club (Gordon et al., 2017c) dataset (Fig. 1). In this dataset, a large amount of resting-state fMRI data was sampled from ten individuals over ten sessions (Gordon et al., 2017c), yielding highly reliable maps of each individual's brain networks. In this figure, if the white circle on the intraparietal lobule in Fig. 1B was selected as a region of interest for fcMRI analysis, Participant 1 (top) would likely show less coupling with other regions of the default network (red) than the same region for Participant 2 (bottom). In a group-level study design, this finding would be interpreted as reduced connectivity strength within the default network for Participant 1 (relative to Participant 2). However, the fcMRI data obtained from personalized models suggests that a likelier explanation is the difference in size between these two default network regions across individuals: in Participant 1, the (red) default network region encompasses nearly the entire region (white circle) while the same location for Participant 2 contains a much smaller region of the default network. The presence of these individual differences in the spatial layout of brain networks thus complicates the interpretation of measures of fcMRI strength in group-level study designs.

Given this limitation, it is notable that despite the large number of group-level studies examining fcMRI in individuals with psychiatric disorders, there is still no consensus as to how alterations in fcMRI are associated with psychopathology (Etkin, 2019; Saggar and Uddin, 2019). Although increased connectivity within the default network is often observed in studies of major depressive disorder (Kaiser et al., 2015; Mulders et al., 2015), a recent large-sample investigation instead found evidence for decreased connectivity within the default network (Yan et al., 2019). A similar pattern of contradictory results is also observed in studies of schizophrenia, where both increases and decreases in functional connectivity strength are reported between the same networks across different studies (Yu et al., 2012).

A likely reason for these inconsistencies between studies is that they are group-level study designs that lack the resolution into individual participants' brain networks provided by personalized fcMRI. Thus, these studies cannot distinguish individual differences in the spatial topography of brain networks from differences in the connectivity strength of these networks (see Fig. 1). As individual differences in the spatial topography and connectivity strength of brain networks[2] may have different functional correlates, the inability to separately measure them contaminates any attempt to identify associations between brain network connectivity and psychopathology.

Multiple lines of research suggest that idiosyncratic differences in brain network topography are relevant to individual differences in cognition and behavior, suggesting that they are of import to biomarker research. One piece of evidence supporting the relevance of individual differences in network topography to cognition and behavior comes from the spatial locations in which they occur. Individual differences in brain networks occur most frequently in multimodal association networks whose activity correlates with diverse higher-order cognitive processes, and they occur least frequently in unimodal processing networks such as the visual and motor areas (Laumann et al., 2015; Mueller et al., 2013; Seitzman et al., 2019). This localization suggests that they may be a possible substrate of individual differences in these complex functions.

Further evidence demonstrates that the locations of task-related activations (collected in the MRI scanner) largely adhere to the topographic boundaries of personalized brain networks (Braga et al., 2020; DiNicola et al., 2020; Gordon et al., 2017c; Laumann et al., 2015; Seitzman et al., 2019; Tavor et al., 2016). For example, language processing tasks have been shown to preferentially increase fMRI activity within the bounds of personalized language brain networks (Braga et al., 2020), even though the spatial locations of this network differ across people. Results such as these suggest that individual differences in network topography are meaningfully related to changes in brain activity related to ongoing task performance.

[2] These factors can also potentially interact in complex ways given that adjacent network regions can be positioned within the complex geometry of the gyral and sulcal folds of the cortex (Braga et al., 2019; Salvo et al., 2021).

Several studies have also found evidence that accounting for the spatial topography of brain networks improves the prediction of behavioral measures collected outside of the scanner (Bijsterbosch et al., 2018). One illustration of this comes from the work of Kong et al. (2019), (2021) who found that personalized brain network estimates predicted measures of emotion, cognitive ability, and personality significantly better than typical group-level approaches. Another study by Feilong et al. (2021) used "hyperalignment", a method designed to align patterns of fcMRI across participants by controlling for local individual differences in brain network topography. These hyperaligned estimates of functional connectivity were able to explain a much larger proportion of the variance in cognitive ability relative to what is typically found in fcMRI studies (i.e., about 20% vs. about 1% of the variance; Marek et al., 2022), demonstrating the promise of this method in future research. Together, these studies suggest that personalized models of brain networks are better suited to identify relationships with behavioral phenotypes than typically used group-level approaches.

Another approach to personalized models of brain networks has focused on what are termed "network variants", which are extreme deviations in fcMRI in individuals relative to group-level networks (Seitzman et al., 2019). Network variants are defined by patterns of connectivity that are associated with different brain networks than what would canonically be expected at a given location (e.g., the black circles in Fig. 1A which encapsulate default network (red) regions in the group-level network map, but instead contain pieces of other networks in individual participants in Fig. 1B). These regions of the cortex show high reliability ($r > 0.8$) with large amounts of data ($>$ 40 min), demonstrating that they are unlikely to be a byproduct of measurement error. Variants also show high correspondence between task and rest states demonstrating their stability across different task demands (Kraus et al., 2021). Using this technique, a recent study showed that participants with more default mode-like network variants reported increased lifetime substance abuse and lower positive life experiences (Seitzman et al., 2019). As these behaviors are relevant to symptoms of psychiatric disorders, this suggests network variants are plausible candidates as biomarkers for psychopathology.

While the use of personalized measures of fcMRI is relatively new, preliminary evidence suggests this approach has utility specifically for identifying relationships with psychiatric symptoms. Studies using personalized approaches to identify brain networks have been shown to outperform group-level approaches for estimating the relationship between fcMRI and symptoms of obsessive-compulsive disorder (Brennan et al., 2019), major depressive disorder (Zhao et al., 2023), schizophrenia (Fan, Li, Peng et al., 2021; Wang et al., 2020), and bipolar disorder (Wang et al., 2020). Therefore, early evidence suggests that the person-specific information provided by personalized measures of fcMRI is directly relevant for biomarker research.

In addition to self-report measures of psychopathology, similar work has also examined the relationship between measures of fcMRI and passively collected metrics from cell phone data. One recent study found that connectivity in the somatomotor network was related to individual differences in participants' mobility as measured by GPS data (Xia et al., 2022). Additionally, other neuroimaging modalities such as electroencephalogram (EEG) have also been used in conjunction with passively collected data to significantly predict depression symptoms over the course of one month (Shah et al., 2021). Though outside the scope of the current review, passive measures of behavior and other neuroimaging modalities also deserve mention as promising future directions for biomarker research using personalized models.

Altogether, the work reviewed here suggests that reliable individual differences in fcMRI are present across individuals, and that these idiosyncrasies are not well-characterized by group-level fcMRI analyses. Furthermore, accounting for these individual differences in fcMRI has resulted in better predictions of behavior versus group-level approaches, including psychopathology, across multiple studies. Thus, while the use of personalized models in fcMRI is still relatively new, the available evidence suggests that it is better suited to measuring relationships between fcMRI and psychopathology than group-level analyses. The early successes of this approach relative to group-level analyses where network topography is assumed to be invariant across participants demonstrates the promise for personalized models of fcMRI in psychiatric biomarker research.

2.2. *Group-level estimates of the structure of psychopathology symptoms may not generalize to individuals*

While so far this section has focused on the shortcomings of using group-level fcMRI analyses to quantify individual differences in brain networks, evidence suggests that similar issues may exist in the measurement of psychopathology. In research contexts, symptoms of psychopathology are typically assessed either as categorical diagnoses or via self-report measures of symptoms. While these methods of symptom assessment can show good psychometric properties (Watson, 2003), they share the assumption that the constructs they measure are invariant across individuals. For example, in the case of categorical diagnoses, the symptoms of psychiatric disorders are often assumed to belong to a unitary underlying construct which is calculated from patients' responses to questions concerning multiple possible symptoms. As these symptoms are proposed to be fungible indicators of the same underlying unitary construct, even individuals who present with multiple different symptoms may meet the criteria for the same disorder (e.g., Zimmerman et al., 2015).

More recently, dimensional models of psychopathology that advocate measuring symptoms on a continuum have been widely adopted (Cuthbert and Insel, 2013; Kotov et al., 2017; Prenoveau et al., 2010). The assumption of invariance also underlies these dimensional models, as the factor and correlational structures of psychopathology symptoms outlined in these models are assumed to be roughly the same across individuals (Fried, 2015; Schmittmann et al., 2013). Therefore, these methods of quantifying psychopathology share the assumption that, while symptom presentations for a given disorder may vary across individuals, the relationships among each person's symptoms can be approximated by the same underlying structure.

Under the assumption that the structures of psychiatric disorders are assumed to be invariant across individuals, it would be expected that the severity of symptoms from a disorder that a given individual experiences should covary together. In other words, if two symptoms are both associated with a given disorder or meaningfully correlated in a dimensional model, then if one symptom worsens in a person the other symptom would also be expected to worsen. Violations of this assumption – that is, the presence of variable symptom structures across individuals – would complicate the use of group-level study designs for measuring psychopathology. This is especially true for biomarker research, as individual differences in the relationships between symptoms may be associated with distinct biological mechanisms. Thus, it is important to understand how well symptom models estimated from group-level approaches correspond to those of individual participants over time.

Within the past decade, evidence has started to accrue that the structure of psychopathology symptoms can be heterogenous across individuals. In one example of how the structure of psychopathology can differ between individuals and groups of individuals, Wright et al. (2015) followed 101 participants who met criteria for any personality disorder with daily surveys of their symptoms for 100 consecutive days (see Fig. 2). When they analyzed the group-level structure of this self-report data (collapsing across time and factor analyzing the data across participants, akin to most typical group-level cross-sectional designs) two factors emerged: internalizing and externalizing (Fig. 2 A). Unsurprisingly, this factor structure mirrored that of prior reports based on group-level cross-sectional self-report data in psychopathology (Forbes et al., 2021; Kotov et al., 2017; Krueger et al., 1998). The authors also modeled the data longitudinally (averaged across

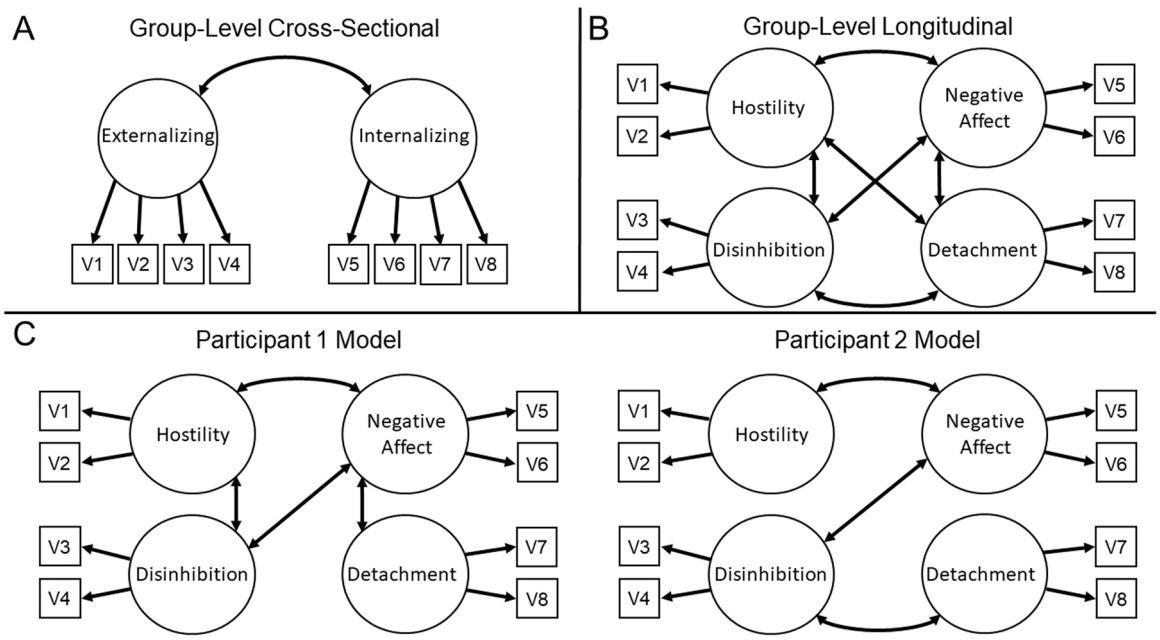

Fig. 2. An example of differences in the structure of psychopathology is shown for models generated using self-report data from either (A) a group-level design averaged across time (cross-sectional), (B) a group-level design averaged between participants (longitudinal), or (C) models fit to individual participants' data across time (personalized). Manifest variables (e.g., V1) are shown in squares and latent factors estimated from these variables (e.g., Externalizing) are shown in circles. Significant correlations between factors are denoted by double-headed arrows. Across the different levels of analysis (i.e., A, B, and C), the factors generated from the manifest variables and the intercorrelations among the factors are noticeably different. For instance, though the cross-sectional data is best characterized two broad factors (A), a four factor solution fits better when the data are longitudinal (B). Importantly, the correlations among the factors in the personalized models (C) are not the same as those reported for either the cross-sectional (A) or longitudinal (B) models. These correlations also vary across individuals (C; Participant 1 versus Participant 2), providing evidence that meaningful individual differences exist in the structure of psychopathology symptoms. This figure is based on the results of Wright et al. (2015).

participants) and found four factors fit the data best: negative affect, detachment, hostility, and disinhibition (Fig. 2B). Thus, the group-level results point to a two factor structure models in the cross-sectional data, and a four factor structure model in the longitudinal data (though see Section 4.5 for some caveats of this interpretation).

However, from these results it is unclear how either of these two group-level models fit individual participants. To address the question of inter-individual differences, Wright et al. (2015) then compared these correlation structures to longitudinal data from individuals using personalized models (Fig. 2 C). Despite forcing the factor solution for individuals to be the same as the group-level longitudinal data (as shown in Fig. 2B), the intercorrelations among these factors differed across individuals. For example, while Participant 1′s personalized model shows that hostility was related to disinhibition and that detachment was related to negative affect; Participant 2′s model, instead, showed that disinhibition was directly related to detachment. This provides evidence that the factor structure within individuals can show inter-individual differences, and thus may not necessarily match a structure of psychopathology generated from group-level data.

Notably, in Wright et al.'s (2015) study the symptom data at the individual-level showed good stability across the 100 days, as measured with both the stability of the mean response (average $r = 0.85$) and variability in responses (average $r = 0.72$). This suggests that individual differences in psychopathology obtained via personalized models likely represent stable trait-like measures of psychiatric symptoms and are not an artifact of measurement error (Wright and Simms, 2016). Further supporting this idea, similar reliability estimates have been obtained for longitudinal measures of psychopathology in other samples (Cranford et al., 2006). Comparable evidence has also been provided using simulations of individual-level longitudinal data (Mansueto et al., 2022). Thus, this study demonstrates that personalized models of psychopathology may capture stable individual differences that are not equivalent to those obtained using group-level approaches.

Other studies further support the notion that group-level designs may not adequately represent the structure of psychopathology in single individuals (Wright and Woods, 2020). In one such study, Fisher (2015) collected data from ten individuals who met criteria for generalized anxiety disorder. These participants completed a daily survey for at least 60 days assessing the symptoms of generalized anxiety disorder as defined by the DSM-5 (American Psychiatric Association, 2013) as well as the related behaviors of avoidance, proactively preparing for possible negative outcomes, procrastination, and reassurance seeking. Using exploratory and confirmatory p-technique factor analysis (Cattell et al., 1947), latent factors were used to estimate personalized models within individuals based on their longitudinal symptom data. Unlike Wright et al. (2015) where the factors were constrained to be the same for each individual's data (Fig. 1C) as for the group-level longitudinal data (Fig. 1B), here the factor solutions were allowed to vary across individuals. Based on the results of this analysis, Fisher (2015) concluded that of the ten participants, only two contained a factor which approximated the symptoms of generalized anxiety disorder in the DSM-5; and no participant had the same latent symptom factors as any other in this group. This suggests that if the factor solution is allowed to vary between participants, the structure of the symptom data may be even more variable across individuals than if this constraint is applied.

In addition to identifying individual differences in the structure of psychopathology, early evidence suggests that personalized models may have utility for selecting treatment interventions. In a recent study, Fisher et al. (2019) used personalized models of psychopathology to guide treatment selection for individuals diagnosed with anxiety and mood disorders. This was done via examining the variance explained by each of the factors obtained from each individual's personalized model and selecting modules from a treatment protocol that best mapped onto these symptoms. The authors found that the estimated effect sizes of their personalized treatments for depression were modestly larger (Hedges $g = 1.86$, 95% CI [1.48 2.24]) than a comparable meta-analytic

estimate (Johnsen and Friborg, 2015) of treatment efficacy (Hedges g = 1.69, 95% CI [1.48 1.89]). Interestingly, the difference between these effect sizes was more pronounced when comparing the effect size of improvement from each session, with Fisher et al. (2019) reporting an average improvement twice as large (Hedges g = 0.24) as was found in Johnsen and Friborg (2015). While preliminary, this suggests that treatments based on personalized models may provide a benefit sooner after beginning treatment than one-size-fits-all approaches. However, in this study no control group was recruited to receive a standardized treatment (or to have their treatment based on symptoms derived from group-level model of psychopathology), so the efficacy of personalized interventions versus a typical treatment cannot be directly inferred. Future studies will be needed to further establish that personalized models capture information about psychopathology that is relevant for treatment.

Together, these results show that heterogeneity in the structure of symptoms can exist across individuals, even for those who meet criteria for the same psychiatric disorder. Thus, if individuals diagnosed with a given disorder are assumed to have a similar symptom correlation structure, this assumption may prove incorrect in many cases. Though we have highlighted only a selection of studies here, others have reported similar results (De Vos et al., 2017; Roefs et al., 2022; Wright et al., 2016, 2019; Wright and Woods, 2020). Altogether, this suggests that the possible lack of generalizability of psychiatric symptom structures from group-level designs to individuals deserves more attention in biomarker research.

3. Quantifying how measures of brain function and psychopathology change over time within individuals is critical for biomarker research

In biomarker research, one of the most important qualities of a measure is its reliability (Gell et al., 2023; Nikolaidis et al., 2022). This is because reliability is directly proportionate to the maximum relationship a measure can have with another variable (Revelle and Condon, 2019; Spearman, 1904): measures with poor reliability have a lowered ceiling on their maximum possible observed correlation value with another measure. This is critical for evaluating the typically small effect sizes observed in biomarker research, as they are determined by the strength of relationships between brain and behavioral measures with varying reliability (Hajcak et al., 2017; Moriarity and Alloy, 2021; Tiego et al., 2023).

The reliability of a measure can be quantified in multiple ways, including internal consistency across participants and the similarity of repeated measures within a participant. One approach to quantifying the degree to which a single measurement shows internal consistency across participants is with split-half reliability, which compares the similarity of results for two halves of the same measure across participants (Lord and Novick, 1968). In contrast, another aspect of reliability focuses on the similarity of results within a participant across repeated measures, such as with test-retest reliability which quantifies the correlation of measures between different timepoints (Revelle and Condon, 2019). A comprehensive understanding of the internal consistency across participants and temporal stability of a measure within participants are critical in biomarker research, as these properties are necessary to properly contextualize the relationship between measures of the brain and psychopathology.

To better understand how these measurement properties contextualize our understanding of biomarkers, one can consider the temporal characteristics of depressed mood in major depressive disorder. In general, while moods are conceptualized to be transient states that fluctuate over relatively brief timescales (Rafaeli et al., 2007), the tendency to persistently experience negative moods is associated with an elevated risk for depression (Burcusa and Iacono, 2007). Thus, while the test-retest reliability of any given mood state is likely not stable over longer timescales, a tendency to experience negative moods can show a longer-term temporal stability. Therefore, to understand the underlying substrates of major depressive disorder, it is necessary to separately identify brain measures that are related to stable traits (i.e., neuroticism, or a trait-like tendency to experience negative moods; Burcusa and Iacono, 2007; Watson and Clark, 1984) from those related to concurrent depressive episodes (i.e., the persistent experience of negative mood due to clinically significant psychopathology) and due to fluctuations from more transient states (i.e., day-to-day or within-day changes in negative moods). Without teasing these separate correlates of brain function apart, our understanding of the mechanisms which instantiate major depressive disorder cannot be considered complete. As these mechanisms exist at the level of individuals, it is also necessary to understand the degree of inter-individual heterogeneity in the neural correlates of these temporally separable aspects of psychopathology.

In this section, we review studies examining transient state-like and stable trait-like aspects of brain networks and psychopathology. First, we review studies which examine the contributions of transient and stable factors to fMRI data and discuss evidence that personalized models improve our measurements of these sources of variance in brain networks. Next, we review past work examining the relationship between transient and stable trait-like aspects of psychiatric symptoms. Again, we discuss evidence for individual differences in these measures and how they hinder our understanding of the mechanisms underlying psychopathology. Overall, we argue that dissociating state-like and trait-like characteristics of both brain and behavior is critical for improving biomarker research in psychiatry.

3.1. Reliable fMRI data within individuals is necessary for dissociating trait and state aspects of functional connectivity

Issues of measurement reliability and stability are especially relevant for fcMRI, as past work has demonstrated that the sampling variability of MRI data is high (Laumann et al., 2016). In other words, using a small amount of fMRI data produces fcMRI measures with high variability (approximated by a multivariate normal distribution; Laumann et al., 2016), leading to estimates that are not representative of the true value of functional connectivity. This results in measures of fcMRI that exhibit poor test-retest reliability with low amounts of fMRI data (Gordon et al., 2017c; Laumann et al., 2015; Noble et al., 2017, 2019). Consistent with this observation (and the low amounts of fMRI collected from individuals in most studies; Naselaris et al., 2021), a recent meta-analysis found the mean estimate of fcMRI reliability across studies to be poor (ICC = 0.29; Noble et al., 2019). This poor reliability means that individual differences in brain networks are difficult to accurately characterize with small amounts of fcMRI data, making this a less than ideal measure for individual differences research.

Important contributions towards understanding the amount of data necessary to obtain reliable fcMRI measurements at the individual level have come from "precision" fMRI experiments (introduced previously in Section 2.1). In these experiments, multiple extended sessions of fMRI data are collected from each individual (Gratton et al., 2020; Michon et al., 2022), in contrast with more typical designs in which 5–10 min of fMRI data are obtained. An early demonstration of the power of precision fMRI came from the MyConnectome dataset (Laumann et al., 2015), in which one participant completed 104 fMRI scans over the course of approximately one and a half years. In this study, it was shown that cortical brain networks defined via fcMRI could achieve high test-retest reliability ($r > 0.95$) within a single participant. Importantly, Laumann

et al. (2015) were able to demonstrate that greater than 30 min of data[3] were needed to obtain these reliable estimates, indicating that personalized fcMRI estimates can become stable with large (or "precision") amounts of data.

A similar result was observed in the Midnight Scan Club dataset (previously mentioned in Section 2.1), where a large amount of resting-state fMRI data was sampled from ten individuals over ten sessions (Gordon et al., 2017c). The same pattern of high reliability ($r > 0.85$) with greater than 30 min of data was replicated in this sample. Other estimates of the reliability of fcMRI are consistent with these findings, showing that using small amounts of data (6 min), test-retest reliability is significantly lower ($\Phi_{UV} = 0.18$) than with large amounts (144 min) of data ($\Phi_{UV} = 0.65$; Noble et al., 2017). It should be noted as well that task-related activations estimated from fMRI follow a similar pattern to those seen with fcMRI, with larger amounts of data collected per participant increasing within-subject reliability (Chen et al., 2021; Nee, 2019). Thus, collecting large amounts of fMRI data within individuals has been shown to produce reliable and personalized estimates of brain function. Given the benefits of using reliable personalized measures in biomarker research, these results suggest that extended fcMRI data collection approaches are crucial for better understanding the neural correlates of behavior.

In addition to producing reliable measures in individuals, other studies suggest that personalized fcMRI models can be used to separate stable trait-like aspects of brain networks from transient state-like factors (Gratton et al., 2020). As discussed previously, an important consideration in clinical applications is the ability to track stable neural characteristics of individuals that predispose them to psychopathology, such as the tendency to experience depressive episodes. This is critical because a plethora of evidence suggests that a stable trait-like vulnerability to experiencing psychopathology is present in many disorders (Burcusa and Iacono, 2007; Janoutová et al., 2016). Personalized brain networks estimated using precision fcMRI have been shown to be stable across sessions (Gordon et al., 2017c; Laumann et al., 2015; Seitzman et al., 2019) and task states (Gratton et al., 2018; Kraus et al., 2021), suggesting that they represent stable trait-like markers of brain organization.

In line with this evidence, Gratton et al. (2018) showed that fcMRI measures are dominated by commonalities across participants and stable individual features that distinguish one participant from another. The authors also found evidence for transient variability in fcMRI measures across days and cognitive states, but these effects were substantially smaller in magnitude than the more stable features. Importantly, while there was significant variance associated with more subtle transient state-like factors, the biggest state effects observed in fcMRI were individual-specific (Gratton et al., 2018). Thus, fcMRI approaches which utilize group-level approaches are not well-suited to measuring these transient changes.

Despite being a relatively new area of study, evidence does exist that personalized fcMRI is well-suited to track changes in brain networks within individuals. In one particularly compelling example, Newbold et al. (2020) applied an arm cast to the dominant arm of three healthy adults. As is typically observed (Biswal et al., 1995), at baseline there was a high degree of functional coupling between the bilateral somatomotor cortices which are primarily responsible for planning and executing motor movements. However, absent any injury to the casted arm, the specific regions of somatomotor cortex of all three individuals corresponding to this arm showed substantial decoupling from other somatomotor regions. Once the casts were removed, fcMRI in the affected regions gradually returned to baseline levels. These significant changes in fcMRI began within 48 hours of the cast being applied, demonstrating the potential for tracking changes in the adult brain over relatively short timescales.

Another study tracking change in fcMRI over time is that of Pritschet et al. (2020), who collected fcMRI in a neurotypical female over the course of the menstrual cycle for 30 days. Consistent fluctuations in whole-brain connectivity between multiple brain networks and levels of estrogen and progesterone were observed across days, indicating that these hormone levels covaried with changes in brain network connectivity throughout the menstrual cycle. Again, this result indicates that precision fcMRI methods can be used to detect subtle but significant variation in brain networks over time.

In addition to tracking fcMRI over the course of days, a previous study by Porter et al. (2022) shows the potential of personalized fMRI approaches to track changes over much shorter timescales. In this study, the authors classified different task states using fcMRI with cross-validated machine learning methods in a precision fMRI dataset. The authors were able to classify differences in these states across people at significantly above chance, but these predictions were much more successful (~30% boost in accuracy) when classifying states within the same person. This pattern was then replicated in an independent dataset with new tasks. These results emphasize that personalized fcMRI models can be powerful approaches to measuring changes in brain states, detecting effects that are not observable in group-level approaches.

A small set of studies have also used personalized fcMRI to track changes in psychopathology over time. While these studies did not collect enough data to be classified as precision fMRI (i.e., they were based on < 10 min of data per participant), they used fcMRI methods that provide personalized estimates of brain networks.[4] Nevertheless, the use of these methods increased the observed effect sizes between measures of brain network connectivity and changes in psychopathology symptoms. In one study, participants diagnosed with first episode schizophrenia were scanned at baseline and after an 8 week treatment with antipsychotic medication (Fan, Li, Guo et al., 2021). The results showed that personalized fcMRI measures were able to predict positive symptoms of schizophrenia significantly better than a group-level approach ($r(28) = 0.57$ versus $r(28) = 0.22$). In another study, changes in fcMRI were measured in participants with obsessive-compulsive disorder both before and after treatment at an inpatient facility (Brennan et al., 2019). Though fcMRI measured using both personalized and group-based parcellations could significantly predict self-reported symptoms, only the personalized parcellations significantly predicted symptom change in response to treatment ($r(39) = 0.374$ versus $r(39) = 0.177$). Changes in personalized estimates of fcMRI have also been shown to be more strongly associated with changes in symptoms in major depressive disorder over the course of treatment than group-level approaches (Zhao et al., 2023). Together these studies provide preliminary evidence that the use of personalized models outperform group-level analyses in tracking changes in

[3] Note that these reliability estimates are for low motion fMRI data (as head motion distorts fcMRI estimates; Power et al., 2012) and for fcMRI estimates of the cerebral cortex. The cerebellum ($r > 0.9$ with > 90 min of data; Marek et al., 2018) and the subcortex ($r > 0.7$ with > 100 min of data in most regions; Greene et al., 2019) require substantially more low motion fMRI data to achieve high reliability. However, new MRI data collection methods have shown promise for improving the reliability of fcMRI in these regions with lesser amounts of data (Lynch et al., 2021).

[4] Lesser amounts of fMRI data have low reliability per participant, but some methods of quantifying personalized measures of fcMRI use template-based or hierarchical models to estimate individual-level measures in non-precision datasets (e.g., Glasser et al., 2016; Guntupalli et al., 2018; Harrison et al., 2015; Kong et al., 2019, 2021; Wang et al., 2015). The priors inherent in these models may help to compensate for the lower reliability seen in these data types, and these approaches can perform better than group-level approaches in many cases. However, they have less flexibility to identify idiosyncratic brain features than precision fcMRI methods. Thus, precision fcMRI is likely to improve upon the benefits of these low-data personalized methods in future work.

psychopathology symptoms across time.

In addition to these studies, there is evidence that precision fMRI data can also increase the effect sizes observed between fcMRI and psychopathology. In one recent study, Gordon et al. (2018) used a precision imaging approach with large amounts of fcMRI data collected per participant in a sample of individuals diagnosed with post-traumatic stress disorder. The authors found significant associations between self-reported symptoms of post-traumatic stress and fcMRI derived from a group-level functional atlas. However, when small amounts (10 min) of fMRI data were used to calculate fcMRI measures from each individual, the relationships between brain networks and symptom severity were no longer significant ($r(24) = -0.47$ versus $r(24) = -0.22$). This suggests extended data collection approaches are likely to improve our ability to find reliable associations between fcMRI and psychopathology, even when a group-level approach is used.

The evidence presented in this section demonstrates that large amounts of fMRI data are necessary to obtain reliable individual-level measures of fcMRI. Furthermore, reliable estimates of fcMRI are necessary to accurately track stable features of brain networks and separate these from more subtle changes in brain network connectivity across time and states. Preliminary evidence also suggests that both personalized models and precision fMRI data improve our ability to detect relationships between fcMRI and psychopathology. Altogether, this suggests that identifying relationships between fcMRI and psychopathology can benefit from personalized models based on extended data collection in fMRI datasets.

3.2. Individual differences in the dynamic processes underlying psychopathology are relevant for biomarker research

While this section has so far discussed the utility of personalized models for measuring changes in brain networks over time, prior work suggests this issue is also relevant for quantifying symptoms of psychopathology. This is because although the severity of symptoms of psychopathology can change over short timescales (DeYoung et al., 2020), there is evidence that dynamic but reproducible patterns exist among symptoms of psychopathology over time (e.g., Shackman et al., 2016). These observations have led to many clinical theories characterized by dysregulated mechanisms of affect, cognition, and behavior within individuals (e.g., Behar et al., 2009). However, in past work these mechanisms have largely been studied using group-level approaches. Given that the proposed mechanisms in these theories are present at the level of individuals, it is important for biomarker research to better understand the extent to which individual differences in these mechanisms exist.

One example of theories positing the existence of dynamic mechanisms in psychopathology comes from studies of neuroticism, which is defined by a pervasive experience of heightened negative affect (McCrae and Costa, 2008; Mineka et al., 2020; Watson and Clark, 1984). Previous evidence suggests that elevated levels of neuroticism (Conway et al., 2016) and experiencing life stressors (Hammen, 2005) are risk factors for developing depressive episodes. However, the degree to which these two risk factors interact is unclear, as multiple theories have been proposed about the dynamic effects of neuroticism on negative affect levels after experiencing life stressors (Shackman et al., 2016). In these theories, negative affect is hypothesized to be elevated either by (a) increased exposure or reactivity to stressors, or (b) a tonic elevation in negative affect. Evidence from one longitudinal study using (approximately) biennial assessments suggests that these risk factors interact with each other, such that life stress has a larger effect on risk for depression in people with elevated neuroticism (Kendler et al., 2004). These results suggest that elevated neuroticism amplifies the impact of stressful life events on depression risk, which is consistent with a theory commonly referred to as the *stress amplification hypothesis* (see Fig. 3B). This hypothesis posits that for people with elevated levels of neuroticism, significantly greater increases in negative affect occur after experiencing life stressors than for people with lower levels of neuroticism.

In contrast to these results, a more recent longitudinal study using annual assessments by Mineka et al. (2020) found there were significant main effects of life stress and neuroticism on future risk for depression, but the interaction between these effects was not significant (and in fact was significantly smaller than their reported main effects). The results of this study are therefore more consistent with the theory of *stably elevated negative affect* (also known as the still-water sea level (neuroticism)-wave crest (stress) model; Zinbarg et al., 2022; see Fig. 3A). This theory states that elevated neuroticism is associated with consistently elevated negative affect (regardless of a clear source of distress), with neuroticism and life stressors exhibiting predominantly independent effects on negative affect. Thus, these two theories make different predictions about how negative affect changes in reaction to life stressors for people high or low in neuroticism.

Together, these studies reached mixed conclusions regarding which of these theories is correct.[5] This may be in part because these hypotheses were tested using group-level study designs rather than examined within individual participants. Therefore, it may instead be the case that some individuals in each sample have a reproducible dynamic pattern consistent with the stably elevated negative affect hypothesis, while others adhere to the stress amplification hypothesis. Under these conditions, group-level study designs would tend to provide evidence for an "average" of both models of vulnerability for depression (Fig. 3C).[6] As this example demonstrates, without using personalized models it is difficult to evaluate whether individual differences exist in the theorized mechanisms underlying psychopathology (Wright and Woods, 2020).

Recent work using personalized models supports the idea that dynamic symptom patterns can vary across individuals (Wright and Woods, 2020). One compelling example comes from the aforementioned study by Fisher (2015), which measured ten individuals diagnosed with generalized anxiety disorder. In addition to the individual-level factor analysis described in Section 2.2, dynamic factor modeling (Molenaar, 1985) was performed on the latent factors estimated for each individual. This method estimates both contemporaneous and time-lagged associations between each factor, allowing for covariance between timepoints to be estimated among the factors. Then, the degree to which a latent factor from a previous day covaried with the other factors on the current day was quantified. This allows for estimates of how relationships between symptoms reproducibly and dynamically change over time.

Using this method, clear individual differences emerged in the

[5] Although these studies reached different conclusions, there were also similarities in their results. Consistent with the stably elevated negative affect hypothesis (see Fig. 3A), Kendler et al. (2004) also found significant main effects of life stress and neuroticism on depression risk in addition to the aforementioned interaction. Conversely, consistent with the stress amplification hypothesis (see Fig. 3B), Mineka et al. (2020) reported that there was a trend towards a significant interaction between life stress and neuroticism, but this trend was not significant. In a re-analysis of Kendler et al. (2004), Mineka et al. (2020) also found that the main effects reported by Kendler et al. (2004) were significantly larger than their reported interaction. Thus, to summarize the results between the two studies, consistent with the stably elevated negative affect hypothesis they both found evidence for main effects of life stress and neuroticism on future depression risk. Additionally, support was also found for the stress amplification hypothesis as the interaction between life stress and neuroticism was in the same direction in both studies, though it was only significant in one of them.

[6] Though the results of any one study may vary through sampling variability alone, this would be the expected pattern observed across many study samples. It may also be the case for example that the majority of individuals in the population show a pattern consistent with the stably elevated negative affect hypothesis. In this case, evidence for this hypothesis would be more frequently observed, even if individuals consistent with both hypotheses are present in meaningful proportions in the population.

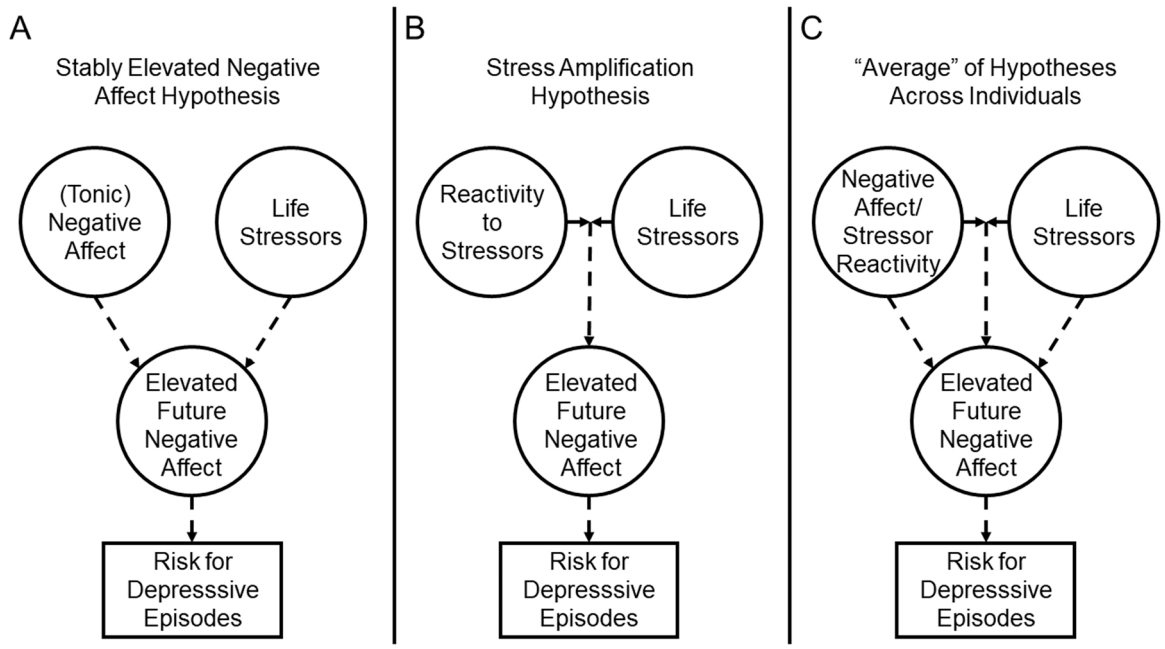

Fig. 3. A hypothetical example is shown outlining how the stably elevated negative affect hypothesis (A) and the stress amplification hypothesis (B) present at the individual level can appear when data pooled across individuals (C). Panel A shows that the effects of negative affect and life stressors on elevated negative affect are additive, consistent with the stably elevated negative affect hypothesis. Panel B illustrates the stress amplification hypothesis, where negative affect and life stressors interact to produce elevated negative affect. In both models, elevations in negative affect increase the subsequent risk for future depressive episodes. If it is the case that both of these hypotheses hold true in different individuals, it is likely that something resembling the "average" model presented in Panel C would be observed if data was pooled across many samples of individuals.

dynamic processes observed among anxiety symptoms. For instance, one of the ten participants demonstrated a dynamic pattern of psychopathology showing their level of avoidance on the previous day was negatively associated with their general distress and fatigue on the subsequent day. However, three other participants demonstrated a different pattern, with their previous levels of worry and avoidance predicting subsequent increases in worry or generalized anxiety disorder symptoms. These results suggest that personalized models of psychopathology may identify reproducible dynamic patterns of symptoms that are heterogenous across individuals, even within the same disorder.

Additional evidence suggests that reproducible dynamic patterns captured by personalized models of psychopathology may be relevant for treatment. In one study examining the longitudinal trajectory of major depressive disorder, symptoms were tracked over the course of inpatient treatment in 255 individuals every two weeks (Hebbrecht et al., 2020). Using a clustering analysis designed for time series data (Sakoe and Chiba, 1978), the authors clustered 17 symptoms of depression from a commonly used depression symptom scale (Hamilton, 1986). This study found that treatment prognosis was significantly better for individuals whose depression symptoms more strongly covaried. In other words, participants improved significantly more depending on the degree to which their depression symptoms either improved or worsened together, as opposed to having different trajectories for different clusters of depression symptoms.

Though the focus of this review is on self-report measures of psychopathology, the potential for personalized models based on passive sensing data such as GPS, ambient sound, and phone usage also deserves mention (Bentley et al., 2019). For instance, it has long been noted that GPS-tracked patterns of individuals' movement are idiosyncratic, but also contain reproducible patterns (Gonzalez et al., 2008). Individual differences in these metrics have been shown to correlate with personality measures (Wang et al., 2018) and also predict binge drinking in young adults (Bae et al., 2018). Thus, these passive sensing metrics also show promise to measure dynamic changes in behavior related to psychopathology.

In summary, this section discussed evidence for heterogeneity and its import for theories of dynamic processes of psychopathology. Although only a selection of studies are highlighted here, additional work supports the notion that individual differences in these dynamic processes exist (Roefs et al., 2022; Wright and Woods, 2020). The possibility of substantial inter-individual differences in these dynamic processes presents a problem for group-level study designs that seek to test multiple models of psychopathology against each other, as these approaches are unable to account for the possibility that different individuals' symptoms are consistent with different models. Given that different dynamic processes may be associated with distinct biological correlates, grouping data across heterogeneous participants is likely to hinder the search for biomarkers in psychiatry. As personalized models can estimate these distinct patterns within each individual, they are ideal candidates to address these issues.

4. Future directions for biomarker research via combining personalized approaches to measuring brain and behavior

So far, this review has discussed prior research generating personalized models of psychopathology symptoms and brain networks measured with fcMRI. Section two discussed evidence that personalized models of fcMRI and psychopathology contain information about individual-level functioning that likely cannot be observed using group-level study designs. Section three reviewed evidence of how each of these measures changes over time, with evidence that personalized models of psychopathology and brain networks are well-suited to measuring both stable, reproducible patterns and more transient states. Yet, despite these common themes among personalized models of fcMRI and psychopathology, so far there has been a lack of integration from these two separate domains into unified research designs.

Towards this end, the goal in this final section is to provide suggestions for how to unify these two approaches into a research agenda that incorporates personalized models of both psychopathology and fcMRI into biomarker studies. Here, three study designs are proposed which

aim to investigate relationships between the stable trait-like and transient state-like sources of variance in psychopathology and fcMRI (Fig. 4). In the first of these designs, here referred to as the *stable-trait* design (Fig. 4A), the goal is to link stable, trait-like aspects of brain networks to similarly stable models of psychopathology at the individual level. The second design, referred to here as the *transient-state* design (Fig. 4B), is designed to identify transient changes in fcMRI that occur concurrently with transient changes in psychiatric symptoms within an individual. Lastly, the third design, referred to here as the *intervention* design (Fig. 4C), is designed to measure changes in personalized fcMRI and psychopathology in response to an intervention.

While these are discussed with respect to fcMRI (given the large body of literature linking psychopathology to distributed brain networks; Kaiser et al., 2015; Mulders et al., 2015; Xu et al., 2019; Yu et al., 2012), these designs could also naturally be extended to fMRI task activation analyses as well as other neuroimaging measures. Relatedly, it should be noted that the suggested study designs in Fig. 4 are not mutually exclusive, and that different aspects of them can be combined for different purposes. For example, the transient-state design could be used to measure both stable and transient aspects of brain networks and psychopathology, although the stable-trait design likely provides a more time- and cost-effective means to measure stable aspects of functioning. The specifics of these designs are further elaborated on in the sections below, as well as some general methodological considerations for research utilizing personalized models.

4.1. Stable-trait study design

One approach for finding biomarkers using personalized models of fcMRI and psychopathology is to look for relationships between stable trait-like measures of brain networks and psychopathology. This type of study design is referred to here as the stable-trait study design (Fig. 4A). As previous evidence suggests that measures of fcMRI (Gratton et al., 2018; Laumann et al., 2015) and self-reported psychopathology (Wright and Simms, 2016) show high temporal stability over the course of at least several months, this suggests that these measures largely reflect stable individual differences in brain function and psychopathology within individuals. This experimental design assumes that stable individual differences in the mechanisms which promote and sustain idiosyncratic relationships among symptoms of psychopathology have different underlying neural correlates, and that these should be reflected in stable individual differences in fcMRI.

In this design, sufficient precision fcMRI and psychopathology self-report measures would be collected (Fig. 4A) to reliably distinguish stable trait-like variance in these measures from other sources of variance within individuals. In the example from Fig. 4A, precision fMRI measures (collected across multiple sessions to reliably measure fcMRI; see Section 3.1) would be collected over the course of the study. Extensive psychopathology symptom data would also be collected to create personalized models of symptoms for each individual. Given that the focus of this design is on stable trait-like features, fcMRI and psychopathology measures need not be contemporaneous, under the assumption that the trait-like components of these measures should show little change over time. However, it may still be ideal to obtain contemporaneous measures of fcMRI and psychopathology, as even trait-like measures from these domains may not be completely stable over time (Gratton et al., 2018; Pincus and Wright, 2010).

For hypothesis testing in this example, individuals could be grouped by commonalities in their personalized psychopathology models, and fcMRI features which distinguish these groups could be identified such as the size and connectivity strength of personalized brain networks (note that the opposite approach of grouping by fcMRI commonalities could also be used). For instance, individual differences in fcMRI could be compared across participants who better fit the pattern of the stably elevated negative affect hypothesis (Fig. 4A, left) versus individuals who adhere more closely to the stress amplification hypothesis (Fig. 4A, right). Then, significant differences in brain network connectivity could be evaluated between these groups. Alternatively, instead of explicitly grouping individuals, continuous measures could be used for hypothesis testing. This study design is conceptually extremely similar to that of many contemporary biomarker studies using fcMRI, but differs in that it incorporates the added individual-level information provided from personalized models of brain networks and psychopathology. This type of study could potentially help identify neurobiologically supported diagnostic categories encompassing specific subpopulations of individuals who have similar symptom presentations. Another possible application of this study design could be to provide information that could be used in risk calculators (e.g., Fusar-Poli et al., 2019) to determine the likelihood that an individual may eventually present with a given disorder.

4.2. Transient-state study design

Another study design for identifying biomarkers using personalized models of fcMRI and psychopathology is to identify relationships between fluctuations in these measures over time within a person. This type of longitudinal within-subjects design is denoted here as the transient-state design (Fig. 4B). As reviewed in Section 3.1, while fcMRI networks are largely stable, they also change subtly over time within individuals (Gratton et al., 2018; Pritschet et al., 2020). Similarly, personalized models of psychopathology exhibit stable patterns of dynamic processes and mechanisms which sustain the observed symptoms of psychopathology (Wright et al., 2015; Wright and Simms, 2016). However, a key feature of these personalized models is that the reported values for each symptom vary across measurement timepoints (Wright and Woods, 2020).

To capture transient variability, this type of study would utilize repeated longitudinal measurements of precision fMRI paired with psychopathology symptoms (e.g., paired fMRI scans and psychopathology measures at many timepoints as shown in the bottom of Fig. 4B). By collecting enough data for precision fcMRI and thus obtaining a reliable measurement at each timepoint, transient changes in connectivity could be quantified separately from the stable trait-like aspects. Using the transient-state design, it is possible to measure whether personalized estimates of fcMRI vary systematically as a function of dynamic changes in psychopathology over time. Thus, the neural correlates of different "states" within a pattern of psychopathology (e.g., Fig. 3) can be identified.

This experimental design is well-suited to measure changes in personalized fcMRI models related to different processes and mechanisms found in personalized models of psychopathology. Using again the example of neuroticism, life stressors, and depression, personalized models of fcMRI could be generated for time periods in which life stressors do or do not occur (Fig. 4B, left) and in which negative affect is predicted to subsequently increase (Fig. 4B, right). Transient changes in fcMRI that are associated with tonic elevations in negative affect could then be compared to fcMRI measured shortly after experiencing a life stressor. In this way, transient changes in fcMRI that consistently occur within different "states" of a given personalized model of psychopathology could be quantified.

One important assumption of this study design is that these dynamic processes in psychopathology are temporally separable by long enough time periods that it is feasible to obtain separate fcMRI measurements. Some evidence for this assumption exists, as increased worry has been shown to lead to either increases or decreases in negative affect the subsequent day after worry was elevated (Fisher, 2015). The evidence that personalized models of psychopathology show meaningful individual differences over this length of sampling interval (24 hours) suggests that it is feasible to collect enough fcMRI data to covary with transient changes in symptoms. This type of study design has many possible applications, such as identifying brain networks that are associated with dynamic changes in psychiatric symptoms, to help identify

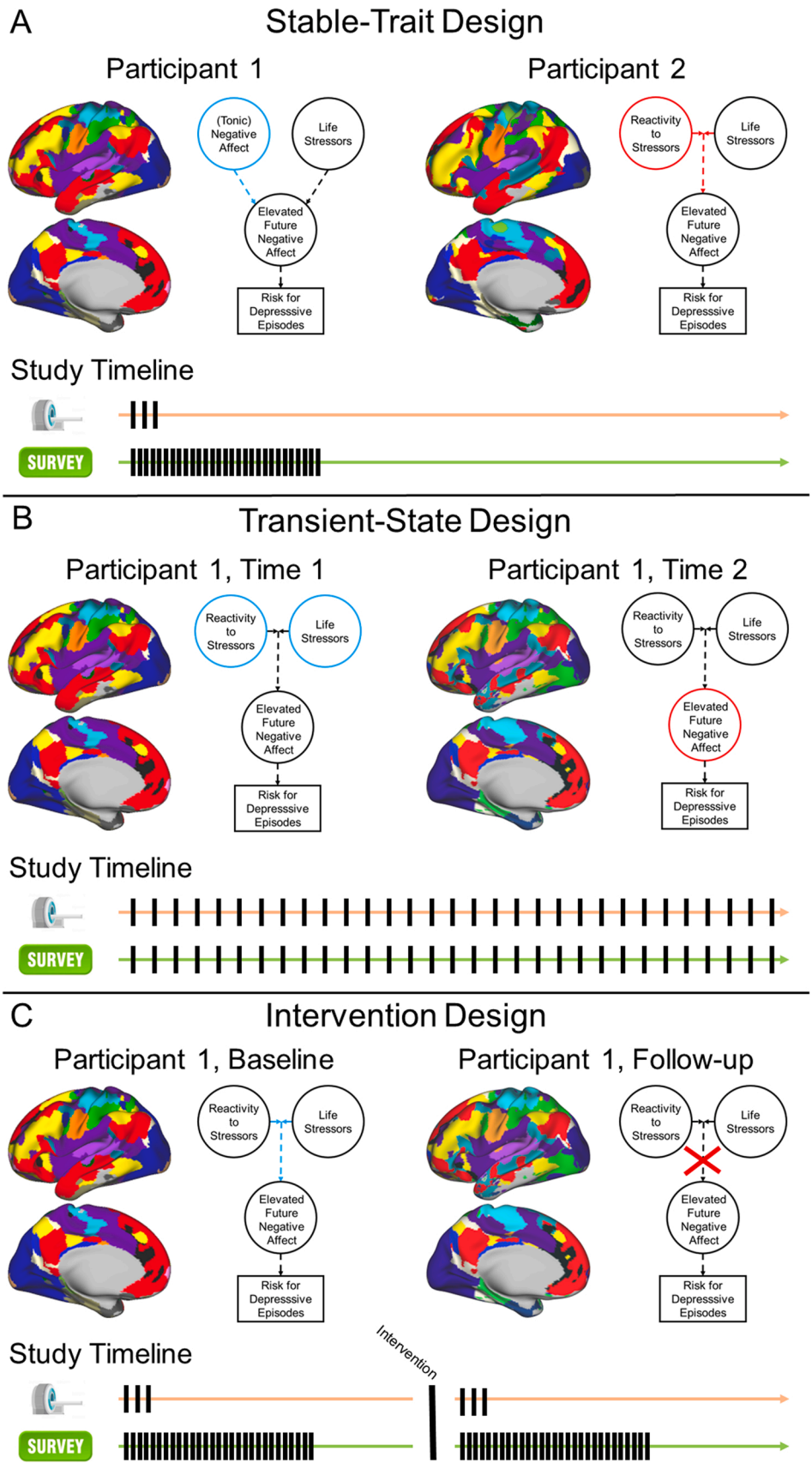

Fig. 4. Visual depictions are shown for examples of all three study designs outlined in this section using the previously discussed examples from Fig. 1 and Fig. 3. For each design, differences in personalized estimates of fcMRI and psychopathology are estimated in (A) stable-trait, (B) transient-state, or (C) intervention study designs. An example timeline for each study design is shown in the lower portion of each section in the figure, showing a hypothetical study timeline of precision fcMRI (orange line) and self-report survey (green line) sampling for each participant. (A) In the stable-trait design, trait-like individual differences in personalized fcMRI are compared between individuals who show a pattern similar to the stably elevated negative affect hypothesis (blue; see Fig. 3A) versus individuals who show a pattern similar to the stress amplification hypothesis (red; see Fig. 3B). (B) In the transient-state design, personalized fcMRI is compared over time within the same individual as symptoms dynamically change from a combination of negative affect and life stressors (blue) to subsequently increased negative affect (red). (C) In the intervention design, personalized measures of fcMRI and psychopathology are compared pre- and post-intervention (red X). In this case, a hypothetical intervention is targeted to disrupt the elevation in negative affect following exposure to a life stressor according to the stress amplification hypothesis (blue). The colors of the brain networks shown correspond to those in the legend of Fig. 1.

clinical cutoffs based on symptom severity, or to identify neural correlates that indicate when an individual may be experiencing clinically significant psychopathology. A better understanding of where in the brain these transient changes in fcMRI occur has the potential to provide insights about the neurobiological underpinnings of psychopathology symptoms.

4.3. Intervention study design

The third study design seeks to quantify how the elements of personalized models of fcMRI and psychopathology change in response to an intervention (e.g., Newbold et al., 2020). In practice, this study design shares many similarities with the stable-trait design, but separately estimates personalized models of fcMRI and psychopathology before and after an intervention. In the pre-treatment interval of this design, personalized models of fcMRI and psychopathology would be estimated as a baseline. Then, after (and/or during) the intervention (Fig. 4C, right), changes in personalized models of fcMRI and psychopathology could again be quantified. Some examples of interventions that could be used with this study design are pharmacological treatments, neuro-stimulation, or psychotherapy.

One method to test for changes in response to the intervention could be to group individuals according to their fcMRI or psychopathology measures at baseline, or to separate individuals based on how they changed in response to the intervention (e.g., responders vs. non-responders). This design could then be used to measure whether, to what degree, or how long a given treatment takes to show efficacy (and at what point its benefits begin to plateau). This study design is not only restricted to interventions which are forms of treatment, but also can be adapted for more transient interventions such as cognitive tasks that have been shown to significantly alter fcMRI within individuals (e.g., Porter et al., 2022).

Returning again to our example of neuroticism, life stressors, and depressive episodes, in this study design a treatment could be implemented that is hypothesized to diminish the link between life stressors and subsequently increased negative affect (Fig. 4C, right). Alternatively, the intervention could target the connectivity within a given brain network (e.g., the default mode network) and measure changes in connectivity strength versus changes in negative affect after experiencing life stressors. In both cases, the changes in personalized models of fcMRI would be compared to the changes in personalized models of psychopathology. In addition, commonalities could be found between participants who responded strongly a given treatment versus those who showed a weaker response, allowing for new hypotheses to be developed about who specifically might benefit. Thus, this design could inform how treatments influence the relationship between brain networks and psychopathology, and this information could be used to streamline interventions tailored to individuals.

4.4. Methodological issues in integrating personalized models of fcMRI and psychopathology

While the insights gained from these study designs have the potential to improve our ability to identify biomarkers, several methodological issues deserve mention. One is that it is unclear how many participants are necessary to both identify commonalities between participants and examine inter-individual variability in the observed effects within a sample. In initial work, even with samples as small as ten participants, substantial between-subject variability has been found in both measures of fcMRI (Gordon et al., 2017c) and psychopathology (Fisher, 2015). Although these smaller samples are sufficient for establishing the presence and characteristics of individual variation in these domains, larger more diverse samples will be necessary to quantify how representative they are of the population at large.

Another possible methodological concern is that with a small number of participants (necessitated by the larger amount of data collection per person; Naselaris et al., 2021; Nikolaidis et al., 2022), statistical power may be lacking to achieve significant results. However, this is less of a concern for study designs focusing on within individual effects (e.g., the transient and intervention study designs), as in these study designs the statistical power is primarily determined by the number of within-subject data points instead of the number of participants. Even in studies where this is not the case and statistical power is related to overall sample size (e.g., the stable-trait design), statistical power is larger for studies using highly reliable within-subject measures of fcMRI and psychopathology (Gratton et al., 2020; Nikolaidis et al., 2022). Therefore, the inherent benefits of personalized models should ameliorate issues concerning statistical power in these study designs. Additional work in this domain will be needed to better outline the statistical properties of personalized modeling methods.

A related issue with smaller sample sizes is that it is unclear to what degree the results of these studies would generalize outside of a given sample. To mitigate this issue, research based on these study designs can adopt a test-retest design of progressively sampling small groups of participants and attempt to generalize these results to other groups (Nesselroade and Ford, 1985). For instance, participants could be recruited in waves and the results obtained in earlier waves could be confirmed in later waves (e.g., Braga et al., 2020; DiNicola et al., 2020). In these study designs, the degree of reproducibility of a within-subject effect across many samples can be a useful indicator of how well they will generalize out of sample. For example, if numerous waves of samples produce results with a similar effect size and homogenous distributions, it is much likelier that the results will generalize to other participants than if substantial heterogeneity exists between samples. In the latter case, larger-scale replications would be necessary to confirm the generalizability of a given effect. Due to the expense of collecting fMRI data, in certain cases it may be more feasible to first generate personalized models of psychopathology symptoms in a larger group of participants if seeking to test an a priori hypothesis (e.g., to contrast two theoretical models). Subsequently, a subset of individuals could then be selected based on the relevance of their personalized symptom models for fMRI data collection.

The issue of generalizability in biomarker research is also closely related to the issue of obtaining unbiased estimates of predictive validity. As personalized models contain much larger amounts of data per participant than typical studies, it is critical that these studies use best practices for generating accurate predictions from their models (Poldrack et al., 2020). This includes avoiding overfitting models to data (Ying, 2019) and when appropriate using best practices for cross-validation (Bzdok and Meyer-Lindenberg, 2018; Dwyer et al., 2018; Nielsen et al., 2019). Despite these obstacles, the large amounts of data obtained from personalized models of psychopathology will likely be complimentary to the goal of generating models which accurately predict behavior (Yarkoni and Westfall, 2017). For example, person-specific machine learning models have been shown to provide better predictions of task state from functional connectivity than versus cross-subject models (Porter et al., 2022).

While the promise of personalized models in biomarker research is evident, another relevant issue is how this research can provide direct benefits in clinical settings. In most cases, it may not be feasible to require individuals seeking psychiatric treatment to complete extensive data collection over the course of days to weeks (Roefs et al., 2022) or to sit still in an MRI scanner for several hours over multiple sessions (Gratton et al., 2020). One critical insight towards this goal is that although personalized models of psychopathology provide a plethora of information about individual function that cannot be measured using brief, cross-sectional self-report measures, the mean scores of these measures across time in personalized models show good correspondence with these brief self-report measures (DeYoung et al., 2020; Fisher et al., 2018).

For instance, if a meaningfully large correlation is observed between default network connectivity and increased negative affect following a

life stressor, this would suggest that this increased connectivity is a biomarker of increased stressor reactivity. Prospective samples could then be used to replicate the relationship between these two measures and demonstrate the potential for interventions targeting default network connectivity. Based on this replicable relationship, a brief self-report measure could be developed that reliably measures reactivity to life stressors and correlates highly with the values on this same measure obtained via personalized models. Then, in clinical settings scores on this brief self-report measure could be used to determine the likelihood of success for an intervention aimed at reducing reactivity to life stressors by targeting default network connectivity. Thus, the results of the research framework outlined here have the potential to be readily translated to clinical practice with a realistic burden on individuals seeking treatment. However, it should be noted that the reliability and validity of biomarkers are not the only impediments to their clinical utility. Biomarkers face other practical limitations such as the cost of obtaining them and the expertise required to administer and interpret them, which are also obstacles to their widespread use (Kirkpatrick et al., 2020).

Although personalized models have many potential benefits for translational neuroscience, they require a large paradigm shift from the predominant study designs currently used for identifying the neural correlates of psychopathology. Many research groups are not set up to collect and analyze large amounts of data from individual participants, and it is natural to have apprehension about investing a large amount of time and funding into a new endeavor with a seemingly uncertain payoff. However, despite these concerns, the potential benefits of collecting large amounts of data from every individual are readily evident. With modern data sharing capabilities, researchers will potentially be able to examine data from personalized models across samples using different study designs, allowing coherent data collection and analysis strategies to be developed. The personalized models estimated in individual studies can then be aggregated from data collected via many different research groups, informing how prevalent different mechanisms of psychopathology and brain function are in the population at large. This will allow researchers to better understand the generalizability of the results of this research for clinical purposes and estimate the number of people who could potentially benefit from novel treatments targeting specific processes related to psychopathology.

4.5. Limitations and challenges of using personalized models of fcMRI and psychopathology in biomarker research

Though the potential of personalized models of fcMRI and psychopathology are promising, best practices have not yet been established for estimating these models. Several outstanding methodological issues and limitations deserve further discussion here. These issues will have to be resolved for personalized models to reach their full promise in biomarker research.

In fcMRI research, one issue is that in-scanner head motion systematically biases brain network connectivity estimates (Power et al., 2012), and the amount a participant moves within the scanner is significantly correlated with cognitive and behavioral measures (Siegel et al., 2017). Thus, studies which do not use best practices for addressing head motion artifacts are likely obtaining biased estimates of the relationship between fcMRI and behavior (Ciric et al., 2017). In addition to motion, estimates of brain network connectivity can be systematically affected by whether the mean signal from the whole brain (global signal) is removed from the data (Power et al., 2017), choice of processing pipelines (Birn et al., 2014; Ciric et al., 2017), and the optimal resolution for defining brain networks with fcMRI data (Huber et al., 2021; Salvo et al., 2021). While these methodological issues present challenges for generating personalized models of fcMRI, many studies have already demonstrated the potential of this method in individual differences research. Future work will be needed to determine best practices in processing MRI data and accounting for artifacts, an endeavor that is already under way (Ciric et al., 2017; Power et al., 2012).

Like fcMRI research, there are also several outstanding issues in using personalized models of psychopathology for biomarker research. Extensive variation exists in the strategies used by different research groups to define personalized models, and these differences in methodology can significantly affect the interpretation of the results (Bastiaansen et al., 2020). It is also unclear exactly which timescales are appropriate for obtaining these measurements from individuals, and the amount of time between each timepoint during longitudinal sampling can have important effects on the results (Wright and Woods, 2020). Relatedly, a variety of methods for generating personalized models have been used in addition to those noted above. For instance, unified structural equation modeling (Kim et al., 2007), vector autoregressive models (Bringmann et al., 2013), and group iterative multiple model estimation (GIMME; Gates et al., 2014; Gates and Molenaar, 2012) have also shown promise for modeling dynamic relationships among psychiatric symptoms within individuals. However, at this point, it is still an open question as to which of these models perform better under different circumstances.

A related issue is how the idiosyncratic features of these models are interpreted relative to models based on group-level data. Though the evidence presented here suggests that idiosyncratic features are routinely observed in personalized models, the degree to which models of individuals' data differ from those defined at the group-level is not typically systematically evaluated. For instance, it should be possible to test how well a model defined using a group of individuals fits the data obtained from any one individual by evaluating the degree of invariance between the two models (Meredith, 1993). This is important for empirically testing that a personalized model provides information not obtainable via group-level analyses, and thus verifying that the structure of psychopathology symptoms in individuals is significantly different from the group-defined structure.

Furthermore, it is also important to test whether broader factors of psychopathology (e.g., the internalizing and externalizing factors reported in many group-defined models; Kotov et al., 2017; Krueger et al., 1998) can be found in individuals, as this would also provide information about how similar individuals' psychopathology symptoms are to those defined at the group-level. This would require examining the correlations among the factors obtained from personalized models to see if evidence for these broader factors is found within individuals. In addition to providing a more stringent test of how an individual's psychopathology symptoms relate to those obtained in the group, measures of how much a person varies from the group-level model may also have relevance for clinical utility. For instance, it may be that the amount an individual varies from the group-defined model can predict the degree to which typical first-line treatments are effective. Thus, a better understanding of how exactly an individual compares to the group would be valuable information for evaluating the utility of personalized models of psychopathology.

It is also important to note that in many of the studies discussed here the reliability and stability of the reported psychopathology symptom measures were not reported. The reliability of these measures is crucial for their interpretation, and as discussed above the degree to which they are stable across time has important implications for their relevance to psychopathology. Thus, information about how temporally stable these measurements are is critical for understanding the degree to which they represent stable trait-like or transient state-like aspects of psychopathology (Calamia, 2019; Wright and Zimmermann, 2019). Despite these outstanding methodological issues, personalized models of psychopathology provide a unique opportunity for identifying within-person mechanisms which sustain psychopathology, and the information provided by these models likely has relevance for biomarker research.

Though the methodological issues of personalized models discussed here are an obstacle to continued progress in biomarker research, it should be noted that progress towards resolving these issues is already underway. Researchers using personalized models in both fcMRI

(Gratton et al., 2020; Lynch et al., 2021; Salvo et al., 2021) and psychopathology (Bastiaansen et al., 2020; Roefs et al., 2022; Stone et al., 2023; Wright and Woods, 2020) are actively working towards better understanding these issues and improving the methodology in these areas. Thus, although there are still outstanding methodological issues as in any relatively young field of research, the potential of personalized models to improve our understanding of biomarkers remains promising.

5. Conclusion

This review has outlined the rationale for generating personalized models of psychopathology and fcMRI for the purpose of identifying biomarkers in psychiatry. The advantage of this strategy is that personalized models can capture reliable individual differences in brain networks and psychopathology at the individual-level that cannot be obtained from group-level estimates of these measures. Evidence from these models has demonstrated that while commonalities are present across individuals, individuals reliably differ from the group-level pattern. Furthermore, these differences are likely important for better understanding the relationship between brain and behavior. The evidence discussed here for a lack of generalizability in group-level measures of brain and behavior across people should encourage a re-evaluation of the appropriate study designs for biomarker research. Specifically, it calls into question the status quo of sampling more participants at the cost of sampling more data from each individual. In line with the aims of the RDoC initiative (Insel, 2014), a greater focus on individual-level processes presents an opportunity to identify biomarkers that are associated with specific mechanisms of psychopathology.

To bolster this goal, we closed this review by proposing several viable study designs that have the potential to identify biomarkers of the processes and mechanisms which sustain psychopathology within individuals. We also discussed outstanding methodological questions and limitations that will need to be addressed for this new approach to reach its potential. While these uncertainties are likely to result in growing pains as in any other nascent field of research, the evidence presented here suggests that the potential benefits of an increased focus on personalized models in biomarker research greatly outweigh the pitfalls.

While only currently a hypothetical framework, we believe that the combination of personalized models of fcMRI and psychopathology presented in this review are critical to the search for biomarkers in psychiatry. Though there are outstanding methodological issues that must be resolved, we believe that the benefits of this approach greatly outweigh contemporary approaches for identifying biomarkers in psychiatry. We hope that the ideas presented here help facilitate a reconsideration of the optimal study designs for biomarker research, and that research using these personalized approaches can inform our understanding of the neural correlates of psychopathology in the years to come.

Data Availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by NIH grant T32NS047987 (to B.K.) and NIMH grants R01MH118370 (to C.G.) and R00MH117226 (to R.M.B.).

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Bae, S., Chung, T., Ferreira, D., Dey, A.K., Suffoletto, B., 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict. Behav. 83, 42–47.

Baker, J.T., Holmes, A.J., Masters, G.A., Yeo, B.T., Krienen, F., Buckner, R.L., Öngür, D., 2014. Disruption of cortical association networks in schizophrenia and psychotic bipolar disorder. JAMA Psychiatry 71 (2), 109–118.

Bastiaansen, J.A., Kunkels, Y.K., Blaauw, F.J., Boker, S.M., Ceulemans, E., Chen, M., Chow, S.-M., de Jonge, P., Emerencia, A.C., Epskamp, S., 2020. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. J. Psychosom. Res. 137, 110211.

Beck, A.T., 1979. Cognitive Therapy of Depression. Guilford press.

Behar, E., DiMarco, I.D., Hekler, E.B., Mohlman, J., Staples, A.M., 2009. Current theoretical models of generalized anxiety disorder (GAD): Conceptual review and treatment implications. J. Anxiety Disord. 23 (8), 1011–1023.

Bentley, K.H., Kleiman, E.M., Elliott, G., Huffman, J.C., Nock, M.K., 2019. Real-time monitoring technology in single-case experimental design research: opportunities and challenges. Behav. Res. Ther. 117, 87–96.

Bijsterbosch, J.D., Woolrich, M.W., Glasser, M.F., Robinson, E.C., Beckmann, C.F., Van Essen, D.C., Harrison, S.J., Smith, S.M., 2018. The relationship between spatial configuration and functional connectivity of brain regions. Elife 7, e32992.

Birn, R.M., Cornejo, M.D., Molloy, E.K., Patriat, R., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., 2014. The influence of physiological noise correction on test–retest reliability of resting-state functional connectivity. Brain Connect. 4 (7), 511–522.

Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. 34 (4), 537–541.

Braga, R.M., Buckner, R.L., 2017. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. Neuron 95 (2), 457–471.

Braga, R.M., Van Dijk, K.R., Polimeni, J.R., Eldaief, M.C., Buckner, R.L., 2019. Parallel distributed networks resolved at high resolution reveal close juxtaposition of distinct regions. J. Neurophysiol. 121 (4), 1513–1534.

Braga, R.M., DiNicola, L.M., Becker, H.C., Buckner, R.L., 2020. Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. J. Neurophysiol.

Brennan, B.P., Wang, D., Li, M., Perriello, C., Ren, J., Elias, J.A., Van Kirk, N.P., Krompinger, J.W., Pope Jr, H.G., Haber, S.N., 2019. Use of an individual-level approach to identify cortical connectivity biomarkers in obsessive-compulsive disorder. Biol. Psychiatry.: Cogn. Neurosci. Neuroimaging 4 (1), 27–38.

Bringmann, L.F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., Tuerlinckx, F., 2013. A network approach to psychopathology: new insights into clinical longitudinal data. PloS One 8 (4), e60188.

Burcusa, S.L., Iacono, W.G., 2007. Risk for recurrence in depression. Clin. Psychol. Rev. 27 (8), 959–985.

Bzdok, D., Meyer-Lindenberg, A., 2018. Machine learning for precision psychiatry: opportunities and challenges. Biol. Psychiatry.: Cogn. Neurosci. Neuroimaging 3 (3), 223–230.

Calamia, M., 2019. Practical considerations for evaluating reliability in ambulatory assessment studies. Psychol. Assess. 31 (3), 285.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.

Cattell, R.B., Cattell, A.K.S., Rhymer, R.M., 1947. P-technique demonstrated in determining psychophysiological source traits in a normal individual. Psychometrika 12 (4), 267–288.

Charney, D.S., Barlow, D.H., Botteron, K., Cohen, J.D., Goldman, D., Gur, R.E., Lin, K.-M., López, J.F., Meador-Woodruff, J.H., Moldin, S.O., 2002. Neuroscience research agenda to guide development of a pathophysiologically based classification system. In: Kupfer, D.J., First, M.B., Regier, D.A. (Eds.), A Research Agenda for DSM-V, first ed. American Psychiatric Association, pp. 31–84.

Chen, G., Pine, D.S., Brotman, M.A., Smith, A.R., Cox, R.W., Taylor, P.A., Haller, S.P., 2021. Hyperbolic trade-off: the importance of balancing trial and subject sample sizes in neuroimaging. NeuroImage, 118786.

Ciric, R., Wolf, D.H., Power, J.D., Roalf, D.R., Baum, G.L., Ruparel, K., Shinohara, R.T., Elliott, M.A., Eickhoff, S.B., Davatzikos, C., 2017. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154, 174–187.

Cohen, J., 1992. Statistical power analysis. Curr. Dir. Psychol. Sci. 1 (3), 98–101.

Conway, C.C., Craske, M.G., Zinbarg, R.E., Mineka, S., 2016. Pathological personality traits and the naturalistic course of internalizing disorders among high-risk young adults. Depress Anxiety 33 (1), 84–93.

Cranford, J.A., Shrout, P.E., Iida, M., Rafaeli, E., Yip, T., Bolger, N., 2006. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? Personal. Soc. Psychol. Bull. 32 (7), 917–929.

Cuthbert, B.N., Insel, T.R., 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 11 (1), 1–8.

De Vos, S., Wardenaar, K.J., Bos, E.H., Wit, E.C., Bouwmans, M.E., De Jonge, P., 2017. An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. PLoS One 12 (6), e0178586.

DeYoung, C.G., Chmielewski, M., Clark, L.A., Condon, D.M., Kotov, R., Krueger, R.F., Lynam, D.R., Markon, K.E., Miller, J.D., Mullins-Sweatt, S.N., 2020. The distinction between symptoms and traits in the hierarchical taxonomy of psychopathology (HiTOP). J. Personal.

DiNicola, L.M., Braga, R.M., Buckner, R.L., 2020. Parallel distributed networks dissociate episodic and social functions within the individual. J. Neurophysiol. 123 (3), 1144–1179.

Dosenbach, N.U., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., Petersen, S.E., 2006. A core system for the implementation of task sets. Neuron 50 (5), 799–812.
Dosenbach, N.U., Fair, D.A., Miezin, F.M., Cohen, A.L., Wenger, K.K., Dosenbach, R.A., Fox, M.D., Snyder, A.Z., Vincent, J.L., Raichle, M.E., 2007. Distinct brain networks for adaptive and stable task control in humans. Proc. Natl. Acad. Sci. 104 (26), 11073–11078.
Dworetsky, A., Seitzman, B.A., Adeyemo, B., Neta, M., Coalson, R.S., Petersen, S.E., Gratton, C., 2021. Probabilistic mapping of human functional brain networks identifies regions of high group consensus. NeuroImage 237, 118164.
Dwyer, D.B., Falkai, P., Koutsouleris, N., 2018. Machine learning approaches for clinical psychology and psychiatry. Annu. Rev. Clin. Psychol. 14, 91–118.
Epstein, S., 1983. Aggregation and beyond: some basic issues on the prediction of behavior. J. Personal. 51 (3), 360–392.
Etkin, A., 2019. A reckoning and research agenda for neuroimaging in psychiatry. Am. J. Psychiatry 176 (7), 507–511.
Fan, Y.-S., Li, H., Guo, J., Pang, Y., Li, L., Hu, M., Li, M., Wang, C., Sheng, W., Liu, H., 2021. Tracking positive and negative symptom improvement in first-episode schizophrenia treated with risperidone using individual-level functional connectivity. Brain Connect.
Fan, Y.-S., Li, L., Peng, Y., Li, H., Guo, J., Li, M., Yang, S., Yao, M., Zhao, J., Liu, H., 2021. Individual-specific functional connectome biomarkers predict schizophrenia positive symptoms during adolescent brain maturation. Hum. Brain Mapp. 42 (5), 1475–1484.
Fedorenko, E., 2021. The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. Curr. Opin. Behav. Sci. 40, 105–112.
Feilong, M., Guntupalli, J.S., Haxby, J.V., 2021. The neural basis of intelligence in fine-grained cortical topographies. Elife 10, e64058.
Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18 (11), 1664.
Fisher, A.J., 2015. Toward a dynamic model of psychological assessment: implications for personalized care. J. Consult. Clin. Psychol. 83 (4), 825.
Fisher, A.J., Medaglia, J.D., Jeronimus, B.F., 2018. Lack of group-to-individual generalizability is a threat to human subjects research. Proc. Natl. Acad. Sci. 115 (27), E6106–E6115.
Fisher, A.J., Bosley, H.G., Fernandez, K.C., Reeves, J.W., Soyster, P.D., Diamond, A.E., Barkin, J., 2019. Open trial of a personalized modular treatment for mood and anxiety. Behav. Res. Ther. 116, 69–79.
Forbes, M.K., Sunderland, M., Rapee, R.M., Batterham, P.J., Calear, A.L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A.J., Lynch, S.J., 2021. A detailed hierarchical model of psychopathology: from individual symptoms up to the general factor of psychopathology. Clin. Psychol. Sci. 9 (2), 139–168.
Fried, E.I., 2015. Problematic assumptions have slowed down depression research: why symptoms, not syndromes are the way forward. Front. Psychol. 6, 309.
Fried, E.I., Nesse, R.M., 2015. Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. BMC Med. 13 (1), 1–11.
Fusar-Poli, P., Werbeloff, N., Rutigliano, G., Oliver, D., Davies, C., Stahl, D., McGuire, P., Osborn, D., 2019. Transdiagnostic risk calculator for the automatic detection of individuals at risk and the prediction of psychosis: second replication in an independent national health service trust. Schizophr. Bull. 45 (3), 562–570.
Gates, K.M., Molenaar, P.C., 2012. Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. NeuroImage 63 (1), 310–319.
Gates, K.M., Molenaar, P.C., Iyer, S.P., Nigg, J.T., Fair, D.A., 2014. Organizing heterogeneous samples using community detection of GIMME-derived resting state functional networks. PloS One 9 (3), e91322.
Gazzaniga, M.S., 2004. The Cognitive Neurosciences. MIT press.
Gell, M., Eickhoff, S.B., Omidvarnia, A., Kueppers, V., Patil, K.R., Satterthwaite, T.D., Mueller, V.I., Langner, R., 2023. The burden of reliability: how measurement noise limits brain-behaviour predictions. BioRxiv, 2023–02.
Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536 (7615), 171–178.
Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782.
Gordon, E.M., Nelson, S.M., 2021. Three types of individual variation in brain networks revealed by single-subject functional connectivity analyses. Curr. Opin. Behav. Sci. 40, 79–86.
Gordon, E.M., Laumann, T.O., Adeyemo, B., Petersen, S.E., 2017a. Individual variability of the system-level organization of the human brain. Cereb. Cortex 27 (1), 386–399.
Gordon, E.M., Laumann, T.O., Adeyemo, B., Gilmore, A.W., Nelson, S.M., Dosenbach, N. U., Petersen, S.E., 2017b. Individual-specific features of brain systems identified with resting state functional correlations. NeuroImage 146, 918–939.
Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Jacqueline, H.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Snyder, N.U., 2017c. Precision functional mapping of individual human brains. Neuron 95 (4), 791–807.
Gordon, E.M., Scheibel, R.S., Zambrano-Vazquez, L., Jia-Richards, M., May, G.J., Meyer, E.C., Nelson, S.M., 2018. High-fidelity measures of whole-brain functional connectivity and white matter integrity mediate relationships between traumatic brain injury and post-traumatic stress disorder symptoms. J. Neurotrauma 35 (5), 767–779.
Gordon, J.A., Redish, A.D., 2016. On the cusp. Current challenges and promises in psychiatry. In: Gordon, J.A., Redish, A.D. (Eds.), Computational Psychiatry: New Perspectives on Mental Illness. MIT Press, pp. 3–14.
Gratton, C., Laumann, T.O., Nielsen, A.N., Greene, D.J., Gordon, E.M., Gilmore, A.W., Nelson, S.M., Coalson, R.S., Snyder, A.Z., Schlaggar, B.L., 2018. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. Neuron 98 (2), 439–452.
Gratton, C., Kraus, B.T., Greene, D.J., Gordon, E.M., Laumann, T.O., Nelson, S.M., Dosenbach, N.U., Petersen, S.E., 2020. Defining individual-specific functional neuroanatomy for precision psychiatry. Biol. Psychiatry 88 (1), 28–39.
Greene, D.J., Marek, S., Gordon, E.M., Siegel, J.S., Gratton, C., Laumann, T.O., Gilmore, A.W., Berg, J.J., Nguyen, A.L., Dierker, D., 2019. Integrative and network-specific connectivity of the basal ganglia and thalamus defined in individuals. Neuron.
Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. 100 (1), 253–258.
Guntupalli, J.S., Feilong, M., Haxby, J.V., 2018. A computational model of shared fine-scale structure in the human connectome. PLoS Comput. Biol. 14 (4), e1006120.
Hajcak, G., Meyer, A., Kotov, R., 2017. Psychometrics and the neuroscience of individual differences: internal consistency limits between-subjects effects. J. Abnorm. Psychol. 126 (6), 823.
Hamilton, M., 1986. The Hamilton rating scale for depression. Assessment of Depression. Springer, pp. 143–152.
Hammen, C., 2005. Stress and depression. Annu. Rev. Clin. Psychol. 1 (1), 293–319.
Hariri, A.R., 2009. The neurobiology of individual differences in complex behavioral traits. Annu. Rev. Neurosci. 32, 225.
Harrison, S.J., Woolrich, M.W., Robinson, E.C., Glasser, M.F., Beckmann, C.F., Jenkinson, M., Smith, S.M., 2015. Large-scale probabilistic functional modes from resting state fMRI. NeuroImage 109, 217–231.
Hebbrecht, K., Stuivenga, M., Birkenhäger, T., Morrens, M., Fried, E.I., Sabbe, B., Giltay, E.J., 2020. Understanding personalized dynamics to inform precision medicine: a dynamic time warp analysis of 255 depressed inpatients. BMC Med. 18 (1), 1–15.
Huber, L., Finn, E.S., Chai, Y., Goebel, R., Stirnberg, R., Stöcker, T., Marrett, S., Uludag, K., Kim, S.-G., Han, S., 2021. Layer-dependent functional connectivity methods. Prog. Neurobiol. 207, 101835.
Insel, T.R., 2014. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. Am. J. Psychiatry 171 (4), 395–397.
Janoutová, J., Janackova, P., Sery, O., Zeman, T., Ambroz, P., Kovalová, M., Varechova, K., Hosák, L., Jirik, V., Janout, V., 2016. Epidemiology and risk factors of schizophrenia. Neuroendocrinol. Lett. 37 (1), 1–8.
Johnsen, T.J., Friborg, O., 2015. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. Psychol. Bull. 141 (4), 747.
Kaiser, R.H., Andrews-Hanna, J.R., Wager, T.D., Pizzagalli, D.A., 2015. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA Psychiatry 72 (6), 603–611.
Kendler, K.S., Kuhn, J., Prescott, C.A., 2004. The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. Am. J. Psychiatry 161 (4), 631–636.
Kim, J., Zhu, W., Chang, L., Bentler, P.M., Ernst, T., 2007. Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. Hum. Brain Mapp. 28 (2), 85–93.
Kirkpatrick, R.H., Munoz, D.P., Khalid-Khan, S., Booij, L., 2020. Methodological and clinical challenges associated with biomarkers for psychiatric disease: a scoping review. J. Psychiatr. Res.
Kong, R., Li, J., Orban, C., Sabuncu, M.R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., 2019. Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. Cereb. Cortex 29 (6), 2533–2551.
Kong, R., Yang, Q., Gordon, E.M., Xue, A., Yan, X., Orban, C., Zuo, X.-N., Spreng, N., Ge, T., Holmes, A., 2021. Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. Cereb. Cortex 31 (10), 4477–4500.
Kotov, R., Krueger, R.F., Watson, D., Achenbach, T.M., Althoff, R.R., Bagby, R.M., Brown, T.A., Carpenter, W.T., Caspi, A., Clark, L.A., 2017. The hierarchical taxonomy of psychopathology (HiTOP): a dimensional alternative to traditional nosologies. J. Abnorm. Psychol. 126 (4), 454.
Kraepelin, E., 1921. Manic-Depressive Insanity and Paranoia. E. & S. Livingstone.
Kraus, B.T., Perez, D., Ladwig, Z., Seitzman, B.A., Dworetsky, A., Petersen, S.E., Gratton, C., 2021. Network variants are similar between task and rest states. NeuroImage 229, 117743.
Krueger, R.F., Caspi, A., Moffitt, T.E., Silva, P.A., 1998. The structure and stability of common mental disorders (DSM-III-R): a longitudinal-epidemiological study. J. Abnorm. Psychol. 107 (2), 216.
Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.-Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U., 2015. Functional system and areal organization of a highly sampled individual human brain. Neuron 87 (3), 657–670.
Laumann, T.O., Snyder, A.Z., Mitra, A., Gordon, E.M., Gratton, C., Adeyemo, B., Gilmore, A.W., Nelson, S.M., Berg, J.J., Greene, D.J., 2016. On the stability of BOLD fMRI correlations. Cereb. Cortex 27 (10), 4719–4732.
Lord, F.M., Novick, M.R., 1968. Statistical Theories of Mental Test Scores. Addison-Wesley.
Lynch, C.J., Elbau, I., Liston, C., 2021. Improving precision functional mapping routines with multi-echo fMRI. Curr. Opin. Behav. Sci. 40, 113–119.

Mansueto, A.C., Wiers, R.W., van Weert, J.C.M., Schouten, B.C., Epskamp, S., 2022. Investigating the feasibility of idiographic network models. Psychol. Methods. https://doi.org/10.1037/met0000466.
Marek, S., Siegel, J.S., Gordon, E.M., Raut, R.V., Gratton, C., Newbold, D.J., Ortega, M., Laumann, T.O., Adeyemo, B., Miller, D.B., 2018. Spatial and temporal organization of the individual human cerebellum. Neuron 100 (4), 977–993.
Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Hendrickson, T.J., 2022. Reproducible brain-wide association studies require thousands of individuals. Nature 603 (7902), 654–660.
McCrae, R.R., Costa Jr., P.T., 2008. The five-factor theory of personality. In: John, O.P., Robins, R.W., Pervin, L.A. (Eds.), Handbook of personality: Theory and research. The Guilford Press, pp. 159–181.
Menon, V., 2011. Large-scale brain networks and psychopathology: a unifying triple network model. Trends Cogn. Sci. 15 (10), 483–506.
Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58 (4), 525–543.
Michon, K.J., Khammash, D., Simmonite, M., Hamlin, A.M., Polk, T.A., 2022. Person-specific and precision neuroimaging: current methods and future directions. NeuroImage, 119589. https://doi.org/10.1016/j.neuroimage.2022.119589.
Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. 19 (11), 1523–1536.
Mineka, S., Williams, A.L., Wolitzky-Taylor, K., Vrshek-Schallhorn, S., Craske, M.G., Hammen, C., Zinbarg, R.E., 2020. Five-year prospective neuroticism–stress effects on major depressive episodes: primarily additive effects of the general neuroticism factor and stress. J. Abnorm. Psychol. 129 (6), 646.
Molenaar, P.C., 1985. A dynamic factor model for the analysis of multivariate time series. Psychometrika 50 (2), 181–202.
Moriarity, D.P., Alloy, L.B., 2021. Back to basics: the importance of measurement properties in biological psychiatry. Neurosci. Biobehav. Rev.
Mueller, S., Wang, D., Fox, M.D., Yeo, B.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., Liu, H., 2013. Individual variability in functional connectivity architecture of the human brain. Neuron 77 (3), 586–595.
Mulders, P.C., van Eijndhoven, P.F., Schene, A.H., Beckmann, C.F., Tendolkar, I., 2015. Resting-state functional connectivity in major depressive disorder: a review. Neurosci. Biobehav. Rev. 56, 330–344.
Naselaris, T., Allen, E., Kay, K., 2021. Extensive sampling for complete models of individual brains. Curr. Opin. Behav. Sci. 40, 45–51.
Nee, D.E., 2019. FMRI replicability depends upon sufficient individual-level data. Commun. Biol. 2 (1), 1–4.
Nesselroade, J.R., Ford, D.H., 1985. P-technique comes of age: multivariate, replicated, single-subject designs for research on older adults. Res. Aging 7 (1), 46–80.
Newbold, D.J., Laumann, T.O., Hoyt, C.R., Hampton, J.M., Montez, D.F., Raut, R.V., Ortega, M., Mitra, A., Nielsen, A.N., Miller, D.B., 2020. Plasticity and spontaneous activity pulses in disused human brain circuits. Neuron 107 (3), 580–589.
Nielsen, A.N., Barch, D.M., Petersen, S.E., Schlaggar, B.L., Greene, D.J., 2019. Machine learning with neuroimaging: evaluating its applications in psychiatry. Biol. Psychiatry.: Cogn. Neurosci. Neuroimaging.
Nikolaidis, A., Chen, A.A., He, X., Shinohara, R., Vogelstein, J., Milham, M., & Shou, H. (2022). Suboptimal phenotypic reliability impedes reproducible human neuroscience. *BioRxiv*.
Noble, S., Spann, M.N., Tokoglu, F., Shen, X., Constable, R.T., Scheinost, D., 2017. Influences on the test–retest reliability of functional connectivity MRI and its relationship with behavioral utility. Cereb. Cortex 27 (11), 5415–5429.
Noble, S., Scheinost, D., Constable, R.T., 2019. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. Neuroimage 203, 116157.
Pincus, A.L., Wright, A.G., 2010. Interpersonal diagnosis of psychopathology. In: Horowitz, L.M., Strack, S. (Eds.), Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions, first ed. Wiley Online Library, pp. 359–381.
Poldrack, R.A., Huckins, G., Varoquaux, G., 2020. Establishment of best practices for evidence for prediction: a review. JAMA Psychiatry 77 (5), 534–540.
Porter, A., Nielsen, A., Dorn, M., Dworetsky, A., Edmonds, D., & Gratton, C. (2022). Masked features of task states found in individual brain networks. *Cerebral Cortex*.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional network organization of the human brain. Neuron 72 (4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59 (3), 2142–2154.
Power, J.D., Plitt, M., Laumann, T.O., Martin, A., 2017. Sources and implications of whole-brain fMRI signals in humans. Neuroimage 146, 609–625.
Prenoveau, J.M., Zinbarg, R.E., Craske, M.G., Mineka, S., Griffith, J.W., Epstein, A.M., 2010. Testing a hierarchical model of anxiety and depression in adolescents: a tri-level model. J. Anxiety Disord. 24 (3), 334–344.
Pritschet, L., Santander, T., Taylor, C.M., Layher, E., Yu, S., Miller, M.B., Grafton, S.T., Jacobs, E.G., 2020. Functional reorganization of brain networks across the human menstrual cycle. NeuroImage 220, 117091.
Rafaeli, E., Rogers, G.M., Revelle, W., 2007. Affective synchrony: Individual differences in mixed emotions. Personal. Soc. Psychol. Bull. 33 (7), 915–932.
Revelle, W., Condon, D.M., 2019. Reliability from α to ω: A tutorial. Psychol. Assess. 31 (12), 1395.
Roefs, A., Fried, E.I., Kindt, M., Martijn, C., Elzinga, B., Evers, A.W., Wiers, R.W., Borsboom, D., Jansen, A., 2022. A new science of mental disorders: Using personalised, transdiagnostic, dynamical systems to understand, model, diagnose and treat psychopathology. Behav. Res. Ther. 153, 104096.
Saggar, M., Uddin, L.Q., 2019. Pushing the boundaries of psychiatric neuroimaging to ground diagnosis in biology. ENeuro 6, 6.
Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust., Speech, Signal Process. 26 (1), 43–49.
Salvo, J.J., Holubecki, A.M., Braga, R.M., 2021. Correspondence between functional connectivity and task-related activity patterns within the individual. Curr. Opin. Behav. Sci. 40, 178–188.
Schmittmann, V.D., Cramer, A.O., Waldorp, L.J., Epskamp, S., Kievit, R.A., Borsboom, D., 2013. Deconstructing the construct: a network perspective on psychological phenomena. N. Ideas Psychol. 31 (1), 43–53.
Seitzman, B.A., Gratton, C., Laumann, T.O., Gordon, E.M., Adeyemo, B., Dworetsky, A., Kraus, B.T., Gilmore, A.W., Berg, J.J., Ortega, M., 2019. Trait-like variants in human functional brain networks. Proc. Natl. Acad. Sci. 116 (45), 22851–22861.
Shackman, A.J., Tromp, D.P., Stockbridge, M.D., Kaplan, C.M., Tillman, R.M., Fox, A.S., 2016. Dispositional negativity: an integrative psychological and neurobiological perspective. Psychol. Bull. 142 (12), 1275.
Shah, R.V., Grennan, G., Zafar-Khan, M., Alim, F., Dey, S., Ramanathan, D., Mishra, J., 2021. Personalized machine learning of depressed mood using wearables. Transl. Psychiatry 11 (1), 1–18.
Siegel, J.S., Mitra, A., Laumann, T.O., Seitzman, B.A., Raichle, M., Corbetta, M., Snyder, A.Z., 2017. Data quality influences observed links between functional connectivity and behavior. Cereb. Cortex 27 (9), 4492–4502.
Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15 (1), 72–101.
Stone, A.A., Schneider, S., Smyth, J.M., 2023. Evaluation of pressing issues in ecological momentary assessment. Annu. Rev. Clin. Psychol. *19*.
Tavor, I., Jones, O.P., Mars, R.B., Smith, S.M., Behrens, T.E., Jbabdi, S., 2016. Task-free MRI predicts individual differences in brain activity during task performance. Science 352 (6282), 216–220.
Tiego, J., Martin, E.A., DeYoung, C.G., Hagan, K., Cooper, S.E., Pasion, R., Satchell, L., Shackman, A.J., Bellgrove, M.A., Fornito, A., 2023. Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. Nat. Ment. Health 1 (5), 5. https://doi.org/10.1038/s44220-023-00057-5.
Van Den Heuvel, M.P., Pol, H.E.H., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur. Neuropsychopharmacol. 20 (8), 519–534.
Van Dijk, K.R., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. 103 (1), 297–321.
Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., 2012. The human connectome project: a data acquisition perspective. Neuroimage 62 (4), 2222–2231.
Wang, D., Buckner, R.L., Fox, M.D., Holt, D.J., Holmes, A.J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., 2015. Parcellating cortical functional networks in individuals. Nat. Neurosci. 18 (12), 1853–1860.
Wang, D., Li, M., Wang, M., Schoeppe, F., Ren, J., Chen, H., Öngür, D., Brady, R.O., Baker, J.T., Liu, H., 2020. Individual-specific functional connectivity markers track dimensional and categorical features of psychotic illness. Mol. Psychiatry 25 (9), 2119–2129.
Wang, W., Harari, G.M., Wang, R., Müller, S.R., Mirjafari, S., Masaba, K., Campbell, A.T., 2018. Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (3), 1–21.
Watson, D., 2003. Investigating the construct validity of the dissociative taxon: Stability analyses of normal and pathological dissociation. J. Abnorm. Psychol. 112 (2), 298.
Watson, D., Clark, L.A., 1984. Negative affectivity: The disposition to experience aversive emotional states. Psychol. Bull. 96 (3), 465.
Whitley Jr, B.E., Kite, M.E., 2012. Principles of Research in Behavioral Science, third ed. Routledge.
World Health Organization, 1992. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. World Health Organization,.
Wright, A.G., Simms, L.J., 2016. Stability and fluctuation of personality disorder features in daily life. J. Abnorm. Psychol. 125 (5), 641.
Wright, A.G., Woods, W.C., 2020. Personalized models of psychopathology. Annu. Rev. Clin. Psychol. 16, 49–74.
Wright, A.G., Zimmermann, J., 2019. Applied ambulatory assessment: integrating idiographic and nomothetic principles of measurement. Psychol. Assess. 31 (12), 1467.
Wright, A.G., Beltz, A.M., Gates, K.M., Molenaar, P., Simms, L.J., 2015. Examining the dynamic structure of daily internalizing and externalizing behavior at multiple levels of analysis. Front. Psychol. 6, 1914.
Wright, A.G., Hallquist, M.N., Stepp, S.D., Scott, L.N., Beeney, J.E., Lazarus, S.A., Pilkonis, P.A., 2016. Modeling heterogeneity in momentary interpersonal and affective dynamic processes in borderline personality disorder. Assessment 23 (4), 484–495.
Wright, A.G., Gates, K.M., Arizmendi, C., Lane, S.T., Woods, W.C., Edershile, E.A., 2019. Focusing personality assessment on the person: modeling general, shared, and person specific processes in personality and psychopathology. Psychol. Assess. 31 (4), 502.

Xia, C.H., Barnett, I., Tapera, T.M., Adebimpe, A., Baker, J.T., Bassett, D.S., Brotman, M. A., Calkins, M.E., Cui, Z., Leibenluft, E., 2022. Mobile footprinting: linking individual distinctiveness in mobility patterns to mood, sleep, and brain functional connectivity. Neuropsychopharmacology 47 (9), 1662–1671.

Xu, J., Van Dam, N.T., Feng, C., Luo, Y., Ai, H., Gu, R., Xu, P., 2019. Anxious brain networks: a coordinate-based activation likelihood estimation meta-analysis of resting-state functional connectivity studies in anxiety. Neurosci. Biobehav. Rev. 96, 21–30.

Yan, C.-G., Chen, X., Li, L., Castellanos, F.X., Bai, T.-J., Bo, Q.-J., Cao, J., Chen, G.-M., Chen, N.-X., Chen, W., 2019. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. Proc. Natl. Acad. Sci. 116 (18), 9078–9083.

Yarkoni, T., Westfall, J., 2017. Choosing prediction over explanation in psychology: Lessons from machine learning. Perspect. Psychol. Sci. 12 (6), 1100–1122.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165.

Ying, X., 2019. An overview of overfitting and its solutions. J. Phys.: Conf. Ser. 1168, 022022.

Yu, Q., A Allen, E., Sui, J., Arbabshirani, R., Pearlson, M., G, D Calhoun, V., 2012. Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. Curr. Top. Med. Chem. 12 (21), 2415–2425.

Zhang, J., Kucyi, A., Raya, J., Nielsen, A.N., Nomi, J.S., Damoiseaux, J.S., Greene, D.J., Horovitz, S.G., Uddin, L.Q., Whitfield-Gabrieli, S., 2021. What have we really learned from functional connectivity in clinical populations? NeuroImage, 118466.

Zhao, Y., Dahmani, L., Li, M., Hu, Y., Ren, J., Lui, S., Wang, D., Kuang, W., Gong, Q., Liu, H., 2023. Individualized functional connectome identified replicable biomarkers for dysphoric symptoms in first-episode medication-naïve patients with major depressive disorder. Biol. Psychiatry.: Cogn. Neurosci. Neuroimaging.

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., Dalrymple, K., 2015. How many different ways do patients meet the diagnostic criteria for major depressive disorder? Compr. Psychiatry 56, 29–34.

Zinbarg, R.E., Williams, A.L., Mineka, S., 2022. A current learning theory approach to the etiology and course of anxiety and related disorders. Annu. Rev. Clin. Psychol. 18, 233–258.