Machine learning and big data in psychiatry: toward clinical applications

Robb B Rutledge[1,2], Adam M Chekroud[3,4] and
Quentin JM Huys[1,5,6]

Psychiatry is a medical field concerned with the treatment of mental illness. Psychiatric disorders broadly relate to higher functions of the brain, and as such are richly intertwined with social, cultural, and experiential factors. This makes them exquisitely complex phenomena that depend on and interact with a large number of variables. Computational psychiatry provides two ways of approaching this complexity. Theory-driven computational approaches employ mechanistic models to make explicit hypotheses at multiple levels of analysis. Data-driven machine-learning approaches can make predictions from high-dimensional data and are generally agnostic as to the underlying mechanisms. Here, we review recent advances in the use of big data and machine-learning approaches toward the aim of alleviating the suffering that arises from psychiatric disorders.

Addresses
[1] Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, England, United Kingdom
[2] Wellcome Centre for Human Neuroimaging, University College London, London, England, United Kingdom
[3] Department of Psychiatry, Yale University, New Haven, CT, United States
[4] Spring Health, New York, NY, United States
[5] Division of Psychiatry, University College London, London, England, United Kingdom
[6] Camden and Islington NHS Foundation Trust, London, England, United Kingdom

Corresponding author: Huys, Quentin JM (qhuys@cantab.net)

Today, medicine is increasingly focused on pragmatic evidence-based practice: the systematic study of what treatments work best for a given problem [1]. Meta-analyses and systematic reviews leverage the large body of data available from double-blind placebo-controlled clinical trials to infer average effects across a population (e.g. for individuals with major depression, selective serotonin reuptake inhibitors are an effective first-line treatment [2]). However, each patient has a unique combination of comorbidities, socioeconomic conditions, medical history, personality, and other factors known to relate to prognosis and treatment response [3]. For example, past cocaine use might downregulate dopamine function [4], which in turn influences learning and future decisions [5] with consequences for future mental health. Since each individual may have a complex and perhaps unique combination of factors, even a very large data set may not provide an opportunity to study similar individuals. Furthermore, the way we traditionally analyze data is by focusing on comparing groups that contain people who are not similar, rather than trying to understand and model how those person-to-person differences impact outcomes. Since both the brain and psychiatric disorders themselves are complex, the full scope of differences between people are in turn complex. In this article, we discuss how big data and machine learning are uniquely placed to address these complexities.

Individual predictions through machine-learning approaches

Other areas of medicine have been successful in developing approaches that more efficiently use existing data. The obvious question is what type of data is useful. Psychiatry's first port of call should be existing data from randomized controlled trials (RCTs), since patients are usually phenotyped carefully and monitored closely over time. However, RCTs are expensive to run, and thus usually only enroll a relatively small number of patients. Statistically speaking, datasets with more questionnaire items than participants can pose problems regarding overfitting: when there are more predictors than data points, enough predictors can always (please emphasize) produce perfect predictions [6]. Even standard cross-validation approaches can overestimate predictive accuracy and reduce the true generalizability of models. Nevertheless, Machine learning provides ways to address this (Box 1), with promising results.

Uher and co-workers first extracted robust factors from questionnaire responses in an unsupervised manner and then tested their power to predict treatment outcomes [7,8]. Other subsequent machine-learning studies including neuroimaging studies (e.g. [9]) focused on recovering diagnostic information from high-dimensional data. Recently, large numbers (>100) of clinical variables have been successfully used to predict treatment response. In first-episode psychosis, response to five different

Box 1 Explicit assessment of validation is critical

It is likely that we are collectively deluding ourselves about the true generalizability of our models. Validation is typically assessed with leave-one-out approaches or at best k-fold cross-validation within a single study. However, both approaches are known to overestimate predictive accuracy [59], and are better interpreted as demonstrating that model performance was not grossly influenced by a small number of individuals in the training data. Leaving individual sites out of a multi-site study is a slight improvement, but is still considered internal validation [60]. For example, Koutsouleris tested a first-epi-sode psychosis algorithm on sites left out of an international multi-site RCT and found significant site-to-site variability [10].

When validation has been explicitly tested on an independent sample, performance invariably degrades. Chekroud *et al.* found that the predictive accuracy of an antidepressant treatment algorithm that was 65% accurate during cross-validation was only 60% accurate in an independent RCT cohort, even though the training data included thousands of patients [34]. A large consortium found no replicable genetic result associated with treatment response in 3756 patients with major depression [62].

Medicine is riddled with observational or retrospective findings that initially seemed promising, but eventually did not improve outcomes (e.g. the idea that vitamins might prevent cardiovascular disease). As a field, we should implement strict replication and validation proce-dures to avoid undermining the integrity of the care we are attempting to guide [25,60]. A bare minimum is requiring explicit assessment of validation, and a framework for setting individual results into the broader context of findings between pure research and clinical practice [57]. We should also be clearer about how any one finding relates to outcomes. Genome-wide association studies still tell us little about the treatments that will help patients get better, despite samples numbering hundreds of thousands. The currency of medicine is patients getting better and not model parameters. This fact bears repeating.

antipsychotic treatments can be predicted from question-naire responses with 71% balanced accuracy across treat-ment sites [10]. Antidepressant response can be predicted using a similar approach [11••]. When other forms of data have been included in predictive models, performance has generally improved modestly beyond clinical and socio-demographic data. Whelan *et al.* predicted development of alcohol misuse in adolescents ($n = 692$) from brain structure and function, individual personality and cognitive differ-ences, environmental factors including gestational ciga-rette and alcohol exposure, life experiences, and candidate genes [12]. While items relating to patient history and personality had the greatest predictive value, neuroimaging and genetic variables still contributed some unique infor-mation. Other studies — although small — also find that cognitive markers and structural and functional MRI can predict clinical outcomes [13–17]. EEG-derived resting-state rostral cingulate theta activity predicts antidepressant response beyond other variables ([18] but see Ref. [19]).

As well as predicting how likely an individual is to recover overall, it is also useful to understand whether certain treatments are more likely to work than others. Some progress has been made, for example, in contrasting antidepressants versus psychotherapy (e.g. [20]). One

other approach developed De Rubeis et al. [21] based on work by Barber and Muenz [22] developed predictors for each RCT arm separately and then applied them to both arms. The difference in predicted outcome can then be interpreted as a reflection of how likely the individual is to respond to one treatment versus the other [23•]. This was also attempted with a broad array of clinical, neuro-imaging, and behavioral variables (c.f [24]). Critics have argued that these tools are not yet useful for clinical practice [25–29] because their performance remains mod-est in terms of explained variance and predictive accuracy [15,30], and the studies have enrolled very small samples relative to the amount of data used for prediction [31]. Nevertheless, one promising study (Figure 1a) suggests that resting-state EEG after medication washout can be used to make antidepressant treatment recommendations to clinicians [32,33•] and merits efforts for replication.

Finally, applying machine learning to existing data might also help identify illness phenotypes that relate more closely to the efficacy of existing treatments. Two recent studies offer notable examples. Applying hierarchical clustering to over 7000 depressed patients from multiple RCTs, Chekroud *et al.* described how symptoms in com-mon depression rating scales hang together, finding three statistically robust and replicable symptom clusters: a mood/emotional cluster, a sleep/insomnia cluster, and an atypical symptom cluster [34]. The three clusters differed in their responsiveness to antidepressants, sug-gesting that treatments could be targeted in individuals based on symptoms. Drysdale *et al.* employed similar hierarchical clustering to resting-state fMRI data from over 300 depressed patients [35], identifying four clusters of patients that differed in their responsiveness to tran-scranial magnetic stimulation therapy. Both studies share commonalities that proved successful: an approach for parsing phenotypic heterogeneity, aggregation of data sets large enough to support the statistical techniques, and demonstration that identified clusters meaningfully relate to treatment outcomes. These large-scale symp-tom-clustering and patient-clustering approaches are beginning to make their way to clinical guidelines on antidepressant selection [36], although the findings are still primarily retrospective rather than prospective.

Collecting bigger data

Given the promise of large samples and the possibility of collecting clinical measures even in large genomic biobanks (e.g. [37]), what big data would be most valuable for psychi-atric research? Online services like Amazon Mechanical Turk (AMT) allow large samples to be rapidly tested on cognitive tasks, allowing robust assessment of novel task characteristics and their relationship to self-reported clinical symptoms [38]. For example, Gillan *et al.* tested 1413 indi-viduals on AMT on both a cognitive task and standard questionnaires covering nine diagnostic categories [39••]. Their analysis revealed three underlying trans-diagnostic

Figure 1
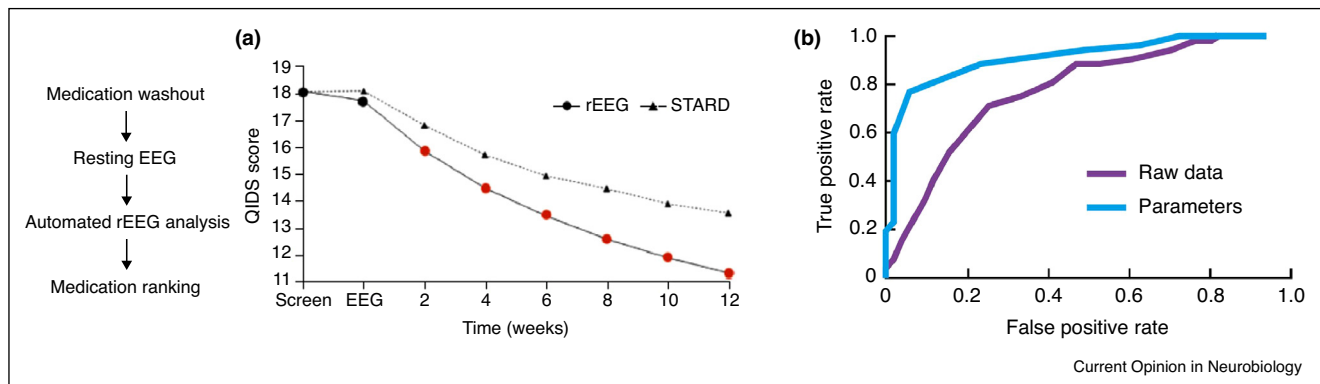
(a) Automated analysis of resting-state EEG (rEEG) after medication washout returned a medication ranking in currently depressed patients. Treatment selection based on rEEG outperformed an optimized clinical protocol based on STAR*D in a 12-site clinical trial [32]. (b) Classifier performance is higher when applied to model parameters estimated from data rather than directly from raw data itself. Behavioral data for 200 agents was simulated with a generative model and split into training and validation sets. A classifier trained on the raw data (purple curve, AUC 0.74) was outperformed by a classifier trained on a single learning rate parameter estimated from the data using a reinforcement learning model (blue curve, 0.87). Panels adapted from Ref. [6].

factors: anxiety/depression, compulsion and intrusive thought, and social anxiety. The compulsion and intrusive thought factor selectively related to task performance, a result replicated in a separate experiment. The same trans-diagnostic factor was associated with high confidence but low self-evaluation ability while the anxiety/depression factor was associated with low confidence but high self-evaluation ability [40]. AMT has also been used to describe the landscape of emotional experience using a factor analysis on emotion labels and ratings elicited by 2185 videos [41]. Concerns have been raised about the reliability of online testing. Reassuringly, depressive symptoms on AMT have high test–retest reliability ($r = 0.87$) after one week [42], suggesting that online data can be sufficiently reliable. More directly relevant to mental health are attempts to leverage online data acquisition and testing for therapeutic means. For example, a smartphone study of digital cognitive behavioral therapy (CBT) in individuals with insomnia successfully reduced insomnia, and mediated reductions in paranoia and hallucinations [43].

Online testing provides a rapid means of prototyping new paradigms in a diverse sample of individuals with a broad range of common clinically relevant symptoms. Depression and anxiety have among the highest costs of all psychiatric disorders and are both as common in online samples as in the general population [42]. Hence, online data collection platforms may allow data to be collected quickly and cheaply, providing an opportunity to acquire enough data to support more sophisticated analyses. There are some downsides to this approach. An online community of workers might also have systematic biases that complicate interpretation of results. Incentivizing individuals to truthfully reveal information

about their private introspective state is also difficult, although some people may be more willing to reveal personal information anonymously than in the presence of a researcher. Promising results obtained in anonymous online studies should be replicated in both more controlled laboratory studies and less controlled smartphone studies.

Mobile devices reveal rich real-time data in the natural environment

Particularly noteworthy are opportunities provided by mobile devices like smartphones. The data that can be collected are extremely rich, spanning 'active' cognitive task and self-report data with consistent results (Figure 2a) for equivalent tasks delivered in the laboratory and by smartphone [44–46]. A smartphone study in 18 420 participants showed that momentary happiness reflected whether things are going better than expected, and not how well things are actually going [47]. Happiness in the computational model could be predicted by the history of rewards and expectations and related to fMRI measurements in the striatum. Depressive symptoms did not affect the strength of this relationship, but instead altered the set point around which happiness varies (Figure 2c) [48••]. Repeated sampling of emotional self-reports has also started to identify dynamic qualities in depression over months and even years. Depressed patients with worse outcomes over several years show stronger inter-connectedness of symptoms [49], and transitions between states of wellness and depression suggest that these states have features that render them inherently stable (Figure 2d) [50]. Features associated with stable healthy and depressed states might be identifiable

Figure 2
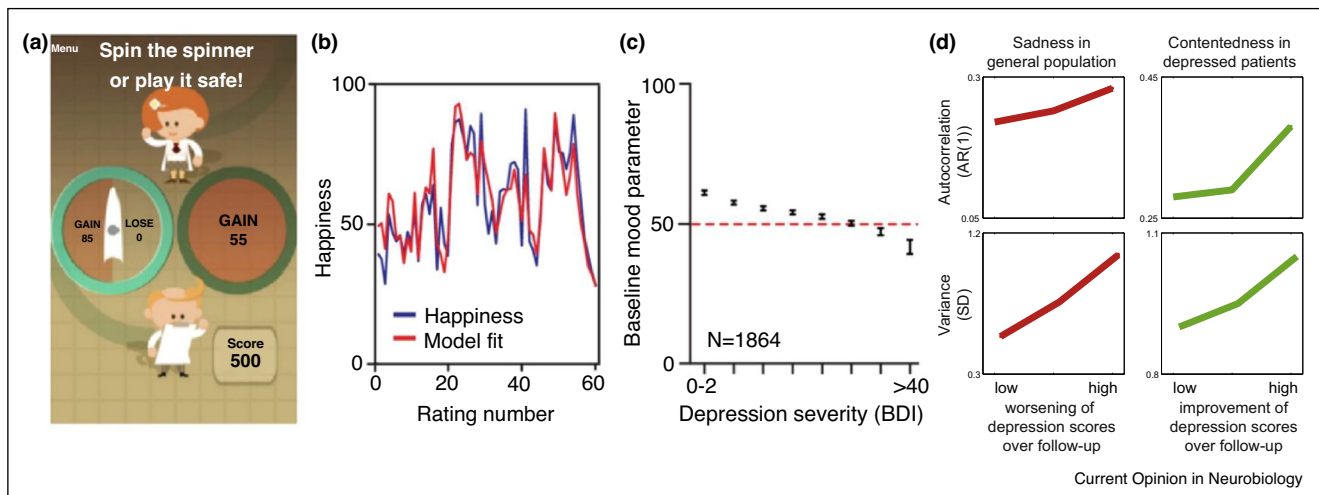
Computational models capture mood fluctuations. **(a)** Risky decision-making task tested both in the lab and using a smartphone platform with over 130 000 participants. **(b)** Lab happiness data and computational model fit. Happiness depends on the history of expectations and prediction errors resulting from those expectations. **(c)** Baseline mood parameters in the model are correlated with depressive symptoms ($n = 1864$; data from Refs. [47,48••]). **(d)** Self-reported symptoms over time show evidence of stable attractor states. Before a transition between healthy and more depressed states, the autocorrelation and variance of symptoms increases, both measures of critical slowing down, a phenomenon arising in dynamical systems transitioning between different stable states. The left panels show that autocorrelation and variance of sadness increase in healthy participants if they experience an increase in depressed symptoms at follow-up, while the right panels show that autocorrelation and variance of contentedness increase in depressed patients that experience an improved state at follow-up. (adapted from Ref. [50]).

in individuals [51], allowing targeted treatments when such features are detected.

Smartphones passively collect a number of streams of data that might eventually contribute toward a richer understanding of psychopathology. Elements including acceleration, movement, location, voice, keyboard use patterns, and social interactions through many sensors might eventually form part of a digital phenotyping of individuals [52,53]. For example, accelerometer data on 91 105 UK Biobank participants were used to derive circadian rhythmicity parameters related to sleep-activity patterns. Circadian disruption was associated with increased lifetime risk of both major depression and bipolar disorder [54].

Theory-driven and data-driven approaches are complementary

Data-driven analyses are agnostic about the nature of the data being analyzed. This can be a weakness because algorithms might neglect what science has already established about the structure of the data. Knowledge of the generative processes can be used to derive so-called sufficient statistics, statistically optimal summaries capturing non-random variation in the data. General-purpose machine-learning algorithms are not obliged to recover sufficient statistics, while theory-driven models are well placed to do so. Hence, theory-driven models can efficiently summarize complex data given things that we know about a generative process, and these summaries can provide inputs

for machine-learning algorithms. Combining theory-driven and data-driven approaches can outperform data-driven approaches alone [6]. Figure 1b provides an example for the case of reinforcement learning. Widely used in the machine learning community, reinforcement-learning algorithms have been extensively examined in neuroscience experiments and the generative processes are increasingly understood at the neurobiological level [55,56].

How do we ensure big data and machine learning improve outcomes?

First, there is a need to generate large data sets that allow the community to parse the complexity inherent in treating psychiatric illness. We have pointed out a number of ways to generate and use large data sets. This will require collaboration, both among academics and with industry. The experience and success of the Psychiatric Genetics Consortium, along with academic–industry partnerships [37] are good examples of how fruitful this approach can be. With sufficiently large data sets, machine-learning techniques can better reach their potential for identifying more expressive functions in data, such as non-linearities or higher-order interaction effects that might not emerge readily using traditional statistical techniques.

Second, recovery of imperfect diagnostic labels is of limited value, as is comparison of cognitive and neural processes between healthy and patient populations with a specific diagnosis. The progress over the past thirty years

has shown that progress with these approaches has led to few major breakthroughs. Instead, we suggest that neurobiological research must examine interventions or risk failing to address the core aim of psychiatry, alleviating the burden of illness. Interventional research is more difficult and risky than cross-sectional research . Interventions throw up complex ethical issues and necessarily involve longitudinal designs, which are burdensome for both participants and experimenters. It is particularly risky for early career scientists because failures are costly yet common.

Clearly, these issues span different aspects in the development of novel methods. We recently suggested a roadmap inspired by the drug development pipeline [57] (Figure 3). This distinguishes early phases in which relevant probes are developed and intermediate phases in which the relationship of these probes to clinical issues (particularly treatment) are established. In later phases (akin to phase III or phase IV clinical trials), use of probes is examined with respect to clinical improvement. Adopting such a framework might help structure the field and clarify the importance of examining the relationship between probes and clinical outcomes.

Finally, we hope that big data and machine learning might spearhead a move toward clinically relevant outcomes that nevertheless allow a more direct link to particular neural or cognitive processes. Currently, outcome measures in mental health either reasonably index severity by collapsing across symptoms likely related to neurobiologically distinct domains, or reasonably index individual neurobiological processes but fail to relate to mental health in a clinically relevant manner. For example, an ICD-10 or DSM-5 diagnosis of depression requires a combination of symptoms that can include anhedonia, depressed mood, sleep disturbances, and changes in concentration. The presence of these symptoms indexes suffering, need for care, and important long-term outcomes namely mortality, remission, and return to work, but may not directly relate to any neural or cognitive process. In contrast, detailed measures of motivation can be mapped onto precise neural substrates, but variation in individual processes bears insufficient relationship to the severity of the illness, that is paramount for clinical practice.

Dangers in the application of machine learning to psychiatry

As a medical discipline, psychiatry's first aim is to do no harm. The collection of useful big data in mental health requires deeply personal information. Abuse of this data is possible given the profound stigma and social impairment concomitant with mental illness. At the same time, the specific data available to machine-learning approaches can lead to biases if certain types of data are not well represented. Algorithms can inherit the biases present in the training data sets (e.g. semantics derived from standard internet-based language corpora exhibit common race and

Figure 3
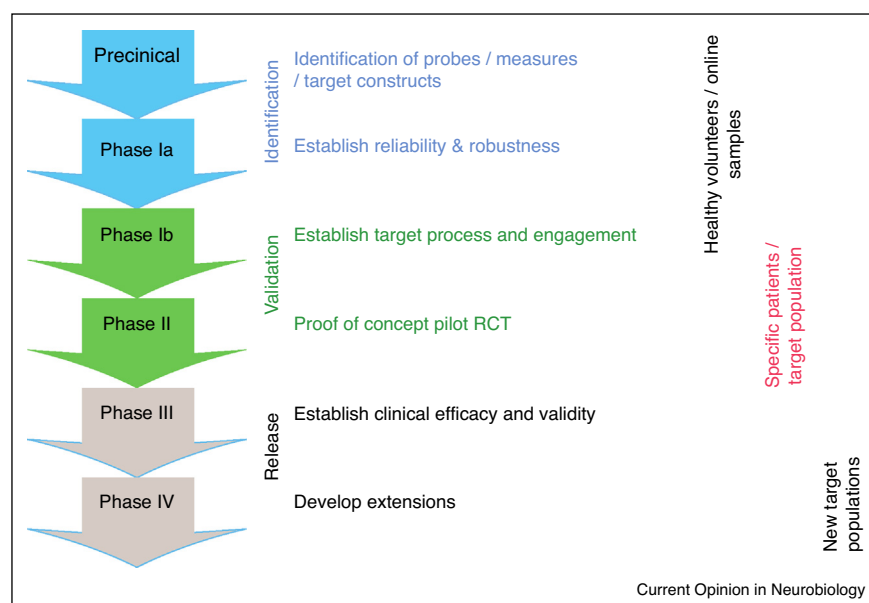
The development of machine learning and computational psychiatry tools for psychiatry can be thought of in parallel to the drug development pipeline. Early preclinical work discovers potential measures, probes and interventions. These then need to be tested for robustness and assessed for their ability to engage the target. Then, the measures need to undergo standard testing in clinical trials to establish their efficacy. Machine learning and big data have roles to play throughout this process, from the identification to the assessment of robustness and to the generalization phase.

gender biases [58]). Given that psychiatric care encompasses compulsory treatment, black box algorithms have the potential to be profoundly discriminatory. This is one issue that the European Union General Data Protection Regulation (GDPR) addresses with the requirement of appropriate procedures to prevent algorithmic biases.

Another danger is the growing perception that machine-learning algorithms are a viable solution in low-N high-p settings, that is, settings where the number of predictors exceeds the number of individuals in the sample. It is correct that machine-learning algorithms will converge in these settings, whereas traditional statistical approaches (e. g. logistic regression) will fail due to unique separation and fitted probabilities of zero or one occur. Many researchers interpret this as an excuse to apply machine-learning algorithms to small samples. However, there is no silver bullet that can replace collecting enough data to generate stable and generalizable predictions [31,34,59]. Although algorithms like the elastic net (a form of penalized regression) and tree-based ensembles (e.g. XGBoost, gradient boosting machines, or random forests) are often used and will fit in low-N high-p settings, caution should be taken to validate the performance of these models on an independent sample in line with established procedures [60,61]. If these caveats are addressed, we believe that these new approaches combining big data, machine learning and theory-driven development of probes have enormous potential to improve treatment of mental illnesses.

Conflict of interest statement

Dr. Chekroud reports holding equity in Spring Care, Inc., Fitbit Inc and UnitedHealthcare Inc. He is lead inventor on three patent submissions related to treatment for major depressive disorder (USPTO docket no. Y0087.70116US00, provisional application no. 62/491,660, and provisional application no. 62/629,041). He has consulted for Fortress Biotech on matters relating to antidepressant drug development and also holds equity in Fitbit Inc and UnitedHealthcare Inc. Dr. Huys and Dr. Rutledge report no financial or other conflicts of interest.

Acknowledgments

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M: **The PRECIS-2 tool: designing trials that are fit for purpose**. *BMJ* 2015, **350**:h2147.

2. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, Leucht S, Ruhe HG, Turner EH, Higgins JPT *et al.*: **Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis**. *Lancet* 2018, **391**:1357-1366.

3. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ *et al.*: **Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice**. *Am J Psychiatry* 2006, **163**:28-40.

4. Ashok AH, Mizuno Y, Volkow ND, Howes OD: **Association of stimulant use with dopaminergic alterations in users of cocaine, amphetamine, or methamphetamine: a systematic review and meta-analysis**. *JAMA Psychiatry* 2017, **74**:511-519.

5. Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PEM, Akil H: **A selective role for dopamine in stimulus-reward learning**. *Nature* 2011, **469**:53-57.

6. Huys QJ, Maia TV, Frank MJ: **Computational psychiatry as a bridge from neuroscience to clinical applications**. *Nat Neurosci* 2016, **19**:404-413.

7. Uher R, Perlis RH, Henigsberg N, Zobel A, Rietschel M, Mors O, Hauser J, Dernovsek MZ, Souery D, Bajs M *et al.*: **Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms**. *Psychol Med* 2012, **42**:967-980.

8. Uher R, Farmer A, Maier W, Rietschel M, Hauser J, Marusic A, Mors O, Elkin A, Williamson RJ, Schmael C *et al.*: **Measuring depression: comparison and integration of three scales in the GENDEP study**. *Psychol Med* 2008, **38**:289-300.

9. Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF: **From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics**. *Neurosci Biobehav Rev* 2015, **57**:328-349.

10. Koutsouleris N, Kahn RS, Chekroud AM, Leucht S, Falkai P, Wobrock T, Derks EM, Fleischhacker WW, Hasan A: **Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach**. *Lancet Psychiatry* 2016, **3**:935-946.

11. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, •• Trivedi MH, Cannon TD, Krystal JH, Corlett PR: **Cross-trial prediction of treatment outcome in depression: a machine learning approach**. *Lancet Psychiatry* 2016, **3**:243-250.
This was the first study to externally validate a treatment prediction model with a psychiatric sample. The study used machine learning to develop a predictive model in one large clinical trial, and then validated the model on an independent clinical trial sample that was collected by other investigators. The model predicted outcomes above chance in the independent trial in arms where patients took similar medications, but did not generalize to medications of a different therapeutic class.

12. Whelan R, Watts R, Orr CA, Althoff RR, Artiges E, Banaschewski T, Barker GJ, Bokde ALW, Büchel C, Carvalho FM *et al.*: **Neuropsychosocial profiles of current and future adolescent alcohol misusers**. *Nature* 2014, **512**:185-189.

13. Phillips ML, Chase HW, Sheline YI, Etkin A, Almeida JRC, Deckersbach T, Trivedi MH: **Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches**. *Am J Psychiatry* 2015, **172**:124-138.

14. Schmaal L, Marquand AF, Rhebergen D, van Tol M-J, Ruhé HG, van der Wee NJA, Veltman DJ, Penninx BWJH: **Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study**. *Biol Psychiatry* 2015, **78**:278-286.

15. Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, Oquendo MA, Bruder G, Pizzagalli D, Toups M *et al.*: **Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design**. *J Psychiatr Res* 2016, **78**:11-23.

16. Korgaonkar MS, Rekshan W, Gordon E, Rush AJ, Williams LM, Blasey C, Grieve SM: **Magnetic resonance imaging measures of**

brain structure to predict antidepressant treatment outcome in major depressive disorder. *EBioMedicine* 2015, **2**:37-45.

17. Etkin A, Patenaude B, Song YJC, Usherwood T, Rekshan W, Schatzberg AF, Rush AJ, Williams LM: **A cognitive–emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial**. *Neuropsychopharmacology* 2015, **40**:1332-1342.

18. Pizzagalli DA, Webb CA, Dillon DG, Tenke CE, Kayser J, Goer F, Fava M, McGrath P, Weissman M, Parsey R *et al.*: **Pretreatment rostral anterior cingulate cortex theta activity in relation to symptom improvement in depression: a randomized clinical trial**. *JAMA Psychiatry* 2018, **75**:547-554.

19. Arns M, Etkin A, Hegerl U, Williams LM, DeBattista C, Palmer DM, Fitzgerald PB, Harris A, deBeuss R, Gordon E: **Frontal and rostral anterior cingulate (rACC) theta EEG in depression: implications for treatment outcome?** *Eur Neuropsychopharmacol* 2015, **25**:1190-1200.

20. Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, Kinkead B, Nemeroff CB, Mayberg HS: **Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder**. *Am J Psychiatry* 2017, **174**:533-545.

21. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L: **The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration**. *PLoS One* 2014, **9**:e83875.

22. Barber JP, Muenz LR: **The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the treatment for depression collaborative research program**. *J Consult Clin Psychol* 1996, **64**:951-958.

23. Webb CA, Trivedi MH, Cohen ZD, Dillon DG, Fournier JC, Goer F,
• Fava M, McGrath PJ, Weissman M, Parsey R *et al.*: **Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study**. *Psychol Med* 2018, **2018**:1-10 http://dx.doi.org/10.1017/S0033291718001708.
Developed and validated a personalized advantage index by deriving separate predictors from each arm of a randomized clinical trial.

24. Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P, DeRubeis RJ: **Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach**. *PLoS One* 2015, **10**:e0140771.

25. Chekroud AM: **Bigger data, harder questions—opportunities throughout mental health care**. *JAMA Psychiatry* 2017, **74**:1183-1184.

26. Paulus MP: **Pragmatism instead of mechanism: a call for impactful biological psychiatry**. *JAMA Psychiatry* 2015, **72**:631-632.

27. Paulus MP: **Evidence-based pragmatic psychiatry—a call to action**. *JAMA Psychiatry* 2017, **74**:1185-1186.

28. Savitz JB, Rauch SL, Drevets WC: **Clinical application of brain imaging for the diagnosis of mood disorders: the current state of play**. *Mol Psychiatry* 2013, **18**:528-539.

29. Jollans L, Whelan R: **The clinical added value of imaging: a perspective from outcome prediction**. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2016, **1**:423-432.

30. Trivedi MH: **Modeling predictors, moderators and mediators of treatment outcome and resistance in depression**. *Biol Psychiatry* 2013, **74**:2-4.

31. Chekroud AM, Koutsouleris N: **The perilous path from publication to practice**. *Mol Psychiatry* 2018, **23**:24-25.

32. DeBattista C, Kinrys G, Hoffman D, Goldstein C, Zajecka J, Kocsis J, Teicher M, Potkin S, Preda A, Multani G *et al.*: **The use of referenced-EEG (rEEG) in assisting medication selection for the treatment of depression**. *J Psychiatr Res* 2011, **45**:64-75.

33. Iosifescu DV, Neborsky RJ, Valuck RJ: **The use of the Psychiatric**
• **Electroencephalography Evaluation Registry (PEER) to personalize pharmacotherapy**. *Neuropsychiatr Dis Treat* 2016, **12**:2131-2142.
Replicated a finding whereby antidepressant treatment allocation through analysis of quantitative EEG outperformed clinical decision-making.

34. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G: **Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach**. *JAMA Psychiatry* 2017, **74**:370-378.

35. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, Fetcho RN, Zebley B, Oathes DJ, Etkin A *et al.*: **Resting-state connectivity biomarkers define neurophysiological subtypes of depression**. *Nat Med* 2017, **23**:28-38.

36. Bayes A, Parker G: **How to choose an antidepressant medication**. *Acta Psychiatr Scand* 2019, **139**:280-291.

37. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF *et al.*: **Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression**. *Nat Genet* 2018, **50**:668-681.

38. Gillan CM, Daw ND: **Taking psychiatry research online**. *Neuron* 2016, **91**:19-23.

39. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND:
•• **Characterizing a psychiatric symptom dimension related to deficits in goal-directed control**. *eLife* 2016, **5**:e11305.
This study used an online platform to collect data in a reinforcement-learning task from a large sample. Task performance related to a single trans-diagnostic factor estimated from questionnaires covering nine diagnostic categories.

40. Rouault M, Seow T, Gillan CM, Fleming SM: **Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance**. *Biol Psychiatry* 2018, **84**:443-451.

41. Cowen AS, Keltner D: **Self-report captures 27 distinct categories of emotion bridged by continuous gradients**. *Proc Natl Acad Sci U S A* 2017, **114**:E7900-E7909.

42. Shapiro DN, Chandler J, Mueller PA: **Using mechanical turk to study clinical populations**. *Clin Psychol Sci* 2013, **1**:213-220.

43. Freeman D, Sheaves B, Goodwin GM, Yu L-M, Nickless A, Harrison PJ, Emsley R, Luik AI, Foster RG, Wadekar V *et al.*: **The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis**. *Lancet Psychiatry* 2017, **4**:749-758.

44. Brown HR, Zeidman P, Smittenaar P, Adams RA, McNab F, Rutledge RB, Dolan RJ: **Crowdsourcing for cognitive science – the utility of smartphones**. *PLoS One* 2014, **9**:e100662.

45. McNab F, Zeidman P, Rutledge RB, Smittenaar P, Brown HR, Adams RA, Dolan RJ: **Age-related changes in working memory and the ability to ignore distraction**. *Proc Natl Acad Sci U S A* 2015, **112**:6515-6518.

46. Rutledge RB, Smittenaar P, Zeidman P, Brown HR, Adams RA, Lindenberger U, Dayan P, Dolan RJ: **Risk taking for potential reward decreases across the lifespan**. *Curr Biol* 2016, **26**:1634-1639.

47. Rutledge RB, Skandali N, Dayan P, Dolan RJ: **A computational and neural model of momentary subjective well-being**. *Proc Natl Acad Sci U S A* 2014, **111**:12252-12257.

48. Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T,
•• Hrynkiewicz L, Lam J, Skandali N, Siegel JZ, Ousdal OT *et al.*: **Association of neural and emotional impacts of reward prediction errors with major depression**. *JAMA Psychiatry* 2017, **74**:790-797.
Computational modelling captured rapid fluctuations in mood during a decision-making task on a smartphone platform. Baseline mood parameters were lower in depressed participants, linking emotional states during smartphone tasks to major depression.

49. van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, Schoevers RA: **Association of symptom network**

structure with the course of depression. *JAMA Psychiatry* 2015, **72**:1219-1226.

50. van de Leemput IA, Wichers M, Cramer AOJ, Borsboom D, Tuerlinckx F, Kuppens P, van Nes EH, Viechtbauer W, Giltay EJ, Aggen SH *et al.*: **Critical slowing down as early warning for the onset and termination of depression**. *Proc Natl Acad Sci U S A* 2014, **111**:87-92.

51. Wichers M, Groot PC, Psychosystems EG: **Critical slowing down as a personalized early warning signal for depression**. *Psychother Psychosom* 2016, **85**:114-116.

52. Torous J, Onnela J-P, Keshavan M: **New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices**. *Transl Psychiatry* 2017, **7**: e1053.

53. Insel TR: **Digital phenotyping: technology for a new science of behavior**. *JAMA* 2017, **318**:1215-1216.

54. Lyall LM, Wyse CA, Graham N, Ferguson A, Lyall DM, Cullen B, Morales CAC, Biello SM, Mackay D, Ward J *et al.*: **Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank**. *Lancet Psychiatry* 2018, **5**:507-514.

55. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH: **A causal link between prediction errors, dopamine neurons and learning**. *Nat Neurosci* 2013, **16**:966-973.

56. Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N: **Arithmetic and local circuitry underlying dopamine prediction errors**. *Nature* 2015, **525**:243-246.

57. Paulus MP, Huys QJM, Maia TV: **A roadmap for the development of applied computational psychiatry**. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2016, **1**:386-392.
Described a roadmap for the development of computational psychiatry tools for clinical practice.

58. Caliskan A, Bryson JJ, Narayanan A: **Semantics derived automatically from language corpora contain human-like biases**. *Science* 2017, **356**:183-186.

59. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. edn 2. Springer-Verlag; 2009.

60. Steyerberg EW, Harrell FE: **Prediction models need appropriate internal, internal-external, and external validation**. *J Clin Epidemiol* 2016, **69**:245-247.

61. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, Collins GS: **External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges**. *BMJ* 2016, **353**:i3140.

62. García-González J, Tansey KE, Hauser J, Henigsberg N, Maier W, Mors O, Placentino A, Rietschel M, Souery D, agar T *et al.*: **Pharmacogenetics of antidepressant response: a polygenic approach**. *Prog Neuropsychopharmacol Biol Psychiatry* 2017, **75**:128-134.