Realizing the Clinical Potential of Computational Psychiatry: Report from the Banbury Center Meeting, February 2019

Michael Browning\(^a\)+, Cameron S. Carter\(^b\), Christopher Chatham\(^c\), Hanneke Den Ouden\(^d\), Claire M Gillan\(^e\), Justin T. Baker\(^f\), Adam M. Chekroud\(^g\), Roshan Cools\(^h\), Peter Dayan\(^i\), James Gold\(^j\), Rita Z. Goldstein\(^k\), Catherine A. Hartley\(^l\), Adam Kepecs\(^m\), Rebecca P. Lawson\(^n\), Janaina Mourao-Miranda\(^o\), Mary L. Phillips\(^p\), Diego A. Pizzagalli\(^q\), Albert Powers\(^r\), David Rindskopf\(^s\), Jonathan P Roi\(^er\)ser\(^t\), Katharina Schmack\(^u\), Daniela Schiller\(^v\), Miriam Sebold\(^w\), Klaas Enno Stephan\(^x\), Michael J Frank\(^y\)+, Quentin Huys\(^z\)+, Martin Paulus\(^zz\)+

\(^*\)Corresponding author
\(^+\)Meeting co-organizers; alphabetical order
\(^a\) Department of Psychiatry, University of Oxford, Oxford, UK: michael.browning@psych.ox.ac.uk
\(^b\) Departments of Psychiatry and Psychology, University of California at Davis
\(^c\) Neuroscience & Rare Diseases (NRD), Roche Pharma Research & Early Development, Roche Innovation Center Basel
\(^d\) Radboud University, Donders Institute for Brain, Cognition and Behaviour.
\(^e\) School of Psychology, Trinity College Dublin, Dublin, Ireland
\(^f\) Institute for Technology in Psychiatry, McLean Hospital and Department of Psychiatry, Harvard Medical School
\(^g\) Spring Health, New York City
\(^h\) Radboud University Medical Center, Department of Psychiatry, Nijmegen, The Netherlands & Donders Institute for Cognition, Brain and Behaviour, Centre for Cognitive Neuroimaging
\(^i\) Max Planck Institute for Biological Cybernetics, Tübingen, Germany
\(^j\) Maryland Psychiatric Research Center, University of Maryland School of Medicine
\(^k\) Dept. of Psychiatry & Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY
\(^l\) Department of Psychology, New York University
\(^m\) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
\(^n\) Department of Psychology, University of Cambridge, Cambridge, UK.
\(^o\) Centre for Medical Image Computing, Department of Computer Science and Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, UK
\(^p\) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA
\(^q\) Department of Psychiatry, Harvard Medical School & McLean Hospital, Belmont, MA
\(^r\) Department of Psychiatry, Yale University School of Medicine, New Haven, CT
\(^s\) Educational Psychology, CUNY Graduate Center, New York, NY
\(^t\) Institute of Cognitive Neuroscience, University College London, London, UK
\(^u\) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
\(^v\) Dept. of Psychiatry & Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY
Keywords: Computational Psychiatry, Reliability, Validity, Collaboration

Word count: 1939

Number Figures: 1

Number Tables: 1
Computational psychiatry is an emerging field that examines phenomena in mental illness using formal techniques from computational neuroscience, mathematical psychology and machine learning (1–6). These techniques can be used in a theory-driven manner to gain algorithmic insight into neural or cognitive processes and in a data-driven way to identify predictive and explanatory relationships in complex datasets. The approaches complement each other: theory-driven models can be used to infer mechanisms, and the resulting measurements can be used in data-driven approaches for prediction. Data-driven algorithms can be used to answer theory-driven questions and pose quantitative problems requiring theoretical analysis. Recent computational studies have successfully described and measured novel mechanisms in a range of disorders (7–11), and have identified novel predictors of treatment response (12, 13). These methods hold the potential to improve identification of relevant clinical variables, and could be superior to classification based on traditional behavioral or neural data alone (14–17). However, these promising results have been slow to influence clinical practice or to improve patient outcomes.

In February 2019 a workshop was convened at the Banbury Centre at Cold Spring Harbor, NY. The purpose of the meeting was to identify key developments required in the practice and infrastructure of computational psychiatry research to accelerate its ability to address real world clinical problems in the near future. This report provides a summary of the conclusions of the meeting. At its core are suggestions to improve the measurement properties of computational assays through a rapid, iterative process that leverages coordinated waves of online and clinical testing, followed by deployment of the assays in innovative study designs to address clinically relevant questions. We particularly focus on theory-driven tasks but, where possible, the potential of data-driven approaches is also highlighted. Finally, the report suggests that for the promise of computational psychiatry to be realized, the research environment must be developed to encourage large-scale, collaborative, interdisciplinary consortia.
We first summarize the need for computational assays with improved measurement properties and describe an iterative optimization and validation procedure by which such assays may be developed and deployed in clinically informative studies. We then consider broader adaptations to the research environment that may accelerate the translation of these techniques.

**Computational assays for clinical applications – what is missing from current research practice?**

**Measurement**

A key application of computational assays is the estimation of latent behavioral and cognitive variables that underlie clinical observations and measurements. Theory-driven approaches rely extensively on generative models i.e., formal descriptions of the underlying neural and mental processes that are believed to generate empirical observations (see 18 for an example). Fitting generative models to observations has a number of potential advantages. First, it may allow identification and measurement of processes not easily captured by traditional analysis (19), second generative models may improve the validity and reliability by which a process is measured. For instance, generative models can describe, and hence be informed by, latent constructs that tie different features (e.g. reaction time and choice) and modalities (e.g. behavior and physiology) together in a holistic manner (20–23). They also allow artificial data to be simulated and therefore a degree of measurement optimization to occur in silico before the assay is deployed in practice (see Table 1 for details). However, these features do not by themselves guarantee that computational assays provide reliable and valid measures of underlying processes. Rather, the measurement properties of an assay must be assessed and iteratively optimized (see Table 1 for a summary of computation-specific and general metrics of
reliability and validity). Though there are notable exceptions (24–27) the issue of measurement in computational psychiatry has not yet attracted due attention. A principled and efficient process of assay development that optimizes measurement properties from the outset is a key outcome of the proposed framework described below.

**Deployment**

Beyond questions of measurement, a second crucial factor in translating computational assays to clinical application is the deployment of the assays in studies that are able to address clinically useful questions. While cross-sectional designs can be used to assess associations between symptoms and computational processes, they provide relatively limited information on the clinical utility of an assay. Alternative study designs which test the degree to which an assay provides predictive information useful to clinical decision-making, or the causal relationships between computationally measured processes and symptoms are likely to be particularly important here. Data-driven techniques are particularly well suited to deployment in predictive studies.

**Developmental pipeline for clinically useful computational assays**

Here we outline a potential, and necessarily collaborative, framework by which promising computational assays, arising from prior clinical, pre-clinical or theoretical work, may be efficiently developed, validated and deployed to address clinically interesting questions (Figure 1).
Establishing and optimizing the measurement characteristics of novel assays

**Step 1: Assay optimization**

First, the important measurement factors of the assay required to address a specific clinical question are selected and then the structure of the assay is altered to optimize these. Table 1 outlines a non-exhaustive list of important metrics. The selected factors may include both specific computational properties such as parameter identifiability (see Table 1) as well as practical features of an assay (e.g. duration to complete, complexity etc) and clinical validity (e.g. correlation with symptoms or treatment response). An objective function, constructed to reflect the specific priorities of a research project including factors to maximize (e.g. sensitivity to manipulations of key task variables, compliance, parameter identifiability) and minimize (e.g. task duration) may then be formalized in which several factors are combined to produce a single metric of measurement performance. The assay may then be optimized by iterative testing either in silico (Table 1), using high-throughput online data collection (28) or if necessary in more deeply phenotyped clinical populations. Here, optimization occurs by systematically varying aspects of the assay’s configuration (e.g., number of trials per condition, timing of stimulus presentation, reward incentives) in order to maximize the objective function. In some cases, this may also include hand-designed qualitative changes (e.g. to improve the task instructions used). Optimization of data-driven approaches may follow a similar trajectory with, for example, the data features being passed to a classifier that is optimized in terms of the predictive validity of the classifier or the practicality of collecting the data.

**Step 2: Latent structure validation**
Although individual model parameters may underlie specific neurocognitive processes, key constructs of clinical relevance are likely to consist of a latent structure of relations between multiple such parameters within or across tasks (14, 29). A useful step is therefore to describe this structure by collecting data from a range of related assays within a single population of participants. Data-driven techniques such as clustering, latent class analysis or theory-driven techniques such as broader generative modelling approaches can be used to determine the latent structure of the assays. Identified latent structures can be fed back to step 1 to inform the further development of the assays, with the best performing (in terms of the metrics described in Table 1) being further deployed as described below.

**Deployment: Establishing the potential of assays as predictors, targets and mediators**

Next, the potential clinical utility of assays can be tested in proof-of-concept studies examining the predictive ability of the assay and/or the causal relationship between the process measured by the assay and clinically important outcomes such as symptoms.

**Step 3a: Clinical prediction and covariation**

Longitudinal observational studies may be used to assess whether an assay covaries with mental state changes or traits of interest and whether it has predictive validity, for example by predicting response to treatment (12). The ability of cohort studies to map the development of psychiatric symptoms may be enhanced by innovative study designs such as longitudinal yet brief “natural challenge” studies (30) which make use of patient and healthy cohorts likely to encounter precipitative events expected to result in a change in psychiatric status (for example, patients with an established mental illness starting a new treatment).
Prediction analyses will typically involve a combination of theory-driven and data-driven analysis, with data-driven analyses used to establish the most powerful predictors (12, 31) and to address issues of dimensionality reduction as described for latent structure validation above. Parameters derived from computational assays may be used as any other variable in data-driven analyses.

**Step 3b: Causality and treatment targets**

A second route by which computational assays may impact clinical practice is if the process measured by the assay constitutes a viable treatment target. That is, treatments may be developed specifically to alter the computationally defined process. This question hinges crucially on the causal relationship between the measured process and clinically relevant outcomes such as symptoms or functioning. Causality is most efficiently addressed using experimental medicine designs which manipulate the underlying, computationally measured, process and then assess the consequences of the manipulation on intermediate or clinical outcomes (where this is not possible, quasi-experimental designs may also be useful (32)). Potential manipulations may be selected from one or more pharmacological, brain stimulation, cognitive or psychotherapeutic technique, the key issue being the ability of the intervention to engage with and alter the computationally measured process.

**Step 4: Clinical efficacy**

Regardless of whether the goal of using a computational assay is to predict a clinically relevant outcome or to guide the development of a novel treatment, the efficacy of computationally informed approaches must ultimately be assessed in clinical trials. Such trials may, for example, randomly assign patients to be treated according to a predictive algorithm or standard treatment, or to receive a computationally informed intervention vs. a control.
In summary, these four steps describe a general pipeline of clinical computational assay optimization designed to yield reliable and valid assays that can be deployed in clinically informative study designs.

**Evolution of the research environment**

Computational assays can be applied to pre-existing datasets (33–35), and the sharing of relevant datasets and analytic procedures is clearly of great importance. However, going forward, the process of computational assay development and deployment outlined above requires substantial structural resources well beyond those of individual laboratories. At the very least, this includes shared core infrastructure, particularly in the domain of information technology. It will necessitate common data structures, which include meta-data relevant to measures, models and populations, and common ascertainment procedures across sites that enable individual labs to collect high-quality behavioral and clinical data and, where relevant, physiological or biological data from both public and clinical settings in a universal format (36). Curation of the data will be required to ensure that it is findable, accessible, interoperable and reusable from the outset. Due to rising concerns about data security on the one hand and the need to provide scientists access to data on the other, the secure storage and aggregation of data across sites using a platform which itself may support data analysis, is likely to be essential (37).

Finally, the complexity of the human mind, the diversity of processes of clinical relevance as well as the range of computational theories and interventions also represent a formidable intellectual challenge. It calls for a pooling of multiple strands of expertise and perspectives which can be achieved in appropriately designed multidisciplinary consortia distributed across laboratories that have a common goal and share data and expertise. Although it is beyond the scope of this paper to specify the precise nature and scope of such consortia we suggest that
they are likely to benefit from the inclusion of, at least, expert clinicians, experimentalists and theoreticians.

**Conclusion**

If computational methods are to deliver real advances for patients, we must ensure our approaches are reliable, robust, and address clinically meaningful questions. In this opinion paper we outline processes to improve the measurement properties and deployment of computational assays and highlight the importance of interdisciplinary collaboration.
Acknowledgements

The Banbury Centre Meeting was funded by grants from the Society for Biological Psychiatry; The William K. Warren Foundation; and The Carney Institute for Brain Science at Brown University. We thank Prof John Krystal, Prof Bruce Cuthbert, Prof Anne Churchland and Dr Michele Ferrante for invaluable discussion and advice.

Disclosures

MB has received travel expenses from Lundbeck for attending conferences and acted as a consultant for Jansen Research and CHDR. CHC is a full-time employee of Hoffman La Roche. JTB has received consulting fees from Pear Therapeutics, BlackThorn Therapeutics, Niraxx Therapeutics, and AbleTo, Inc. PD is funded by the Max Planck Society. AMC holds equity in Spring Care Inc, Fitbit Inc, and UnitedHealthcare Inc; is lead inventor on three patent submissions relating to treatment for major depressive disorder [a) USPTO docket number Y0087.70116US00, b) USPTO. Provisional Appl. No. 62/491,660, and c) USPTO. Provisional Appl. No. 62/629,041]; and has consulted for Fortress Biotech on antidepressant drug development. JMM is supported by the Wellcome Trust under grant number WT102845/Z/13/Z. Over the past 3 years, DAP has received consulting fees from Akili Interactive Labs, BlackThorn Therapeutics, Boehringer Ingelheim, Compass, Posit Science, and Takeda Pharmaceuticals and an honorarium from Alkermes for activities unrelated to the current review. JPR has acted as a consultant for Cambridge Cognition, Takeda, and GE Healthcare. KES is funded by the René and Susanne Braginsky Foundation. MJF is a consultant for Hoffman La Roche. MLP is an advisor to Spring Care, Inc., a behavioral health startup, he has received royalties for an article about methamphetamine in Uptodate and he has received support from the National Institute of General Medical Sciences (P20GM121312, Paulus). Over the past three years, MLP
received an honorarium from Sunovion Pharmaceuticals. HDO has acted as a consultant for Lundbeck pharmaceuticals.

References:


Table 1: Key Metrics of Reliability and Validity Relevant to Computational Measures (see also 38 for a summary)

<table>
<thead>
<tr>
<th>Specific Computational Measures of Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter recovery, identifiability and sensitive range</td>
</tr>
<tr>
<td>Model recovery</td>
</tr>
</tbody>
</table>
Model selection

Where more than one model may be used to describe subject data, a process of **model selection** is used to select the “best” model. This process typically assesses the balance between the “fit” of the model (the degree to which the model can explain the data) and model complexity (i.e. its representational richness or flexibility to fit data in general). If two models explain the data similarly well, the simpler is preferred (Occam’s razor). Taking into account the fit/complexity trade-off is important since models with higher complexity (e.g. with more parameters) will have higher accuracy than simpler models but may be capturing measurement-specific noise (“overfitting”). Model selection may also concern the qualitative ability of the model to recapitulate some important features of the data. While many computational studies select a single, best model for all participants and compare model parameters between participants, it is also possible to assess whether participants differ in the model which best describes their data. The finding that data from different participants are best described by different models may in itself be interesting and may be described using an hierarchical model in which a higher level selects between separate lower level models (note that, in the absence of a single model used across all participants, between subject comparison of model parameters is not straightforward).

Common Measures of Reliability

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest</td>
<td>The degree to which the measures of individuals within a group maintain a consistent relationship across time is assessed by <strong>test-retest reliability</strong>. Test-retest performance is a critical metric for tasks which are required to measure stable, trait-like, within-subject, processes, and for studies using correlational or longitudinal designs.</td>
</tr>
<tr>
<td>Split-half/interrater reliability</td>
<td>Other forms of reliability such as <strong>split half reliability</strong> or <strong>interrater reliability</strong> estimate measurement variability and may be useful in certain computational tasks.</td>
</tr>
<tr>
<td>Common Measures of Validity</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical validity</strong></td>
<td>Evidence for the <strong>clinical validity</strong> of a measure is provided by associations between it and clinically important outcomes such as symptom scores, treatment response or illness course.</td>
</tr>
<tr>
<td><strong>Convergent/divergent validity</strong></td>
<td>The degree to which a measure of a construct correlates with other measures of the same construct (<strong>convergent validity</strong>) and differs from measures of other constructs (<strong>divergent validity</strong>). These metrics therefore provide an assessment of how certain we can be that we are measuring an underlying construct (<strong>convergent validity</strong>) and the degree to which our measure provides the same/different information to alternative measures (<strong>divergent validity</strong>). Questions of convergent and divergent validity have largely been overlooked in computational psychiatry. As a result, it is not clear, for example, whether learning rates for positive outcomes in the plethora of available reward learning tasks measure the same thing.</td>
</tr>
<tr>
<td><strong>Face and ecological validity</strong></td>
<td>This reflects the degree to which a measurement appears to subjectively measure a process (<strong>face validity</strong>) and the degree to which it captures real life processes (<strong>ecological validity</strong>). Computational approaches are able to decompose the components of complex processes and may therefore facilitate the development of more ecologically valid measures of complex real-life interactions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Practical Characteristics of the Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measure duration, complexity and cost</strong></td>
</tr>
</tbody>
</table>

| Translational Relevance of the Measure |
| Cross species translational potential of the task | Depending on the specific question being addressed, the potential for a behavioral measure to be deployed in non-human species may be relevant for measurement selection. For example, validation of the ability of a computational assay to infer physiological mechanisms may require a degree of experimental control that cannot be achieved in humans. |
Figure 1: A suggested Process by which Computational Measures may be Optimized for Deployment in Clinical Studies