Precision clinical trials: a framework for getting to precision medicine for neurobehavioural disorders

Eric J. Lenze, MD; Ginger E. Nicol, MD; Dennis L. Barbour, MD, PhD;
Thomas Kannampallil, PhD; Alex W.K. Wong, PhD; Jay Piccirillo, MD;
Andrew T. Drysdale, MD, PhD; Chad M. Sylvester, MD, PhD; Rita Haddad, MD;
J. Philip Miller, AB; Carissa A. Low, PhD; Shannon N. Lenze, PhD;
Kenneth E. Freedland, PhD; Thomas L. Rodebaugh, PhD

The goal of precision medicine (individually tailored treatments) is not being achieved for neurobehavioural conditions such as psychiatric disorders. Traditional randomized clinical trial methods are insufficient for advancing precision medicine because of the dynamic complexity of these conditions. We present a pragmatic solution: the precision clinical trial framework, encompassing methods for individually tailored treatments. This framework includes the following: (1) treatment-targeted enrichment, which involves measuring patients' response after a brief bout of an intervention, and then randomizing patients to a full course of treatment, using the acute response to predict long-term outcomes; (2) adaptive treatments, which involve adjusting treatment parameters during the trial to individually optimize the treatment; and (3) precise measurement, which involves measuring predictor and outcome variables with high accuracy and reliability using techniques such as ecological momentary assessment. This review summarizes precision clinical trials and provides a research agenda, including new biomarkers such as precision neuroimaging, transcranial magnetic stimulation–electroencephalogram digital phenotyping and advances in statistical and machine-learning models. Validation of these approaches — and then widespread incorporation of the precision clinical trial framework — could help achieve the vision of precision medicine for neurobehavioural conditions.

The burden of neurobehavioural illness and need for precision medicine

Neurobehavioural conditions comprise psychiatric, neurologic and behavioural syndromes such as depressive, anxiety and psychotic disorders; developmental and age-related cognitive impairment or decline (e.g., dementia); substance use disorders; complex neurobiobehavioural conditions such as obesity;[1] and somatic disorders such as bothersome tinnitus and chronic pain. These disorders contribute substantially to disability,[2] mortality[3] and increased health care costs. Treatments for these disorders include medications, neuromodulation devices and behavioural and lifestyle interventions, all of which target the brain and behaviour to improve symptoms and reduce cognitive and functional impairments.[4,5]

The persistently high burden of neurobehavioural illness results from a lack of highly effective treatments, a barrier that can be attributed at least in part to poorly understood pathophysiology. In conditions whose pathophysiology is better understood, such as cardiovascular disease and stroke, comprehensive targeted prevention strategies can be developed and disseminated.[6,7] Prevention approaches that target multiple cardiac risk factors — such as combination statin and antihypertensive therapy[8] and advances in revascularization —have yielded sharp reductions in morbidity and mortality. In contrast, "blockbuster drugs" for neurobehavioural conditions have not yielded blockbuster improvements. For example, placebo-controlled trials in depression and anxiety disorders have shown very small average effect sizes.[9,10,11] Schizophrenia rarely remits completely with antipsychotics, and these medications also produce neurologic and cardiometabolic side effects.[12] Behavioural treatments such as cognitive behaviour therapy may have some advantages over medications (e.g., cost-effectiveness over long-term follow-up[13]), but real-world implementations result in poor fidelity of treatment delivery[14,15] and compromised adherence.[16,17] Mobile health interventions promise improved access and therapeutic reach, but utilization studies report that treatment engagement for most interventions

wanes after approximately 2 weeks.[18,19] Finally, neurobehavioural disorders tend to be chronic and relapse even after a positive response to treatment in the short term.[20,21]

Precision medicine defined

Precision medicine (also referred to as personalized or stratified medicine) is conceptualized as a treatment approach that accounts for individual variation,[22] with the primary goal of providing an individually tailored course of treatment. It is an approach that seeks to provide the right treatment to the right patient, at the right dose, at the right time, with the expectation of better health at a lower cost.[23] Precision medicine would be a compelling solution to reducing the burden of neurobehavioural disorders, because it would involve giving drugs, neuromodulation devices and behavioural treatments to the patients most likely to benefit, as well as optimizing these benefits.

The Precision Medicine Initiative from the US National Institutes of Health (NIH) defines precision medicine as an approach for disease treatment and prevention that accounts for individual variability in genes, environment and lifestyle.[24] It contends that extensive biobehavioural assessment of individual patients will reveal treatment-relevant heterogeneity within diagnostic categories. For example, "depression" is not a single disease, but a heterogeneous syndrome with many underlying neurobiological, behavioural and environmental causes. If subtypes could be distinguished, providers could personalize treatment for depression.[25,26] In this scenario, health care providers can characterize patients based on their biological, environmental and behavioural parameters — using genomic and other biomarker tools along with behavioural phenotyping via patient-reported data and sensors — and choose treatments tailored to the needs of individual patients.

Shortcomings of the current precision medicine approach in randomized controlled trials

Research into individual variability has not translated into more effective treatments for neurobehavioural syndromes, despite heavy investment in the fields of molecular diagnostics, neuroimaging and genomics.[27] Evidence of this is the lack of clinically validated biomarkers in mood,[28] anxiety,[29] psychotic,[30] substance use[31] and eating disorders[32] to guide treatment. A similar lack of progress has been noted for behavioural markers in guiding and individualizing behavioural treatments such as psychotherapy.[33] We argue that the lack of progress is due to a flawed approach for how precision medicine paradigms are conceptualized and tested in neurobehavioural randomized controlled trials (RCTs). This flawed approach fails to take into account the complexity of neurobehavioural disorders in 3 ways.[9,34–37]

First, patient subgroups are defined by baseline pre-randomization markers that are hypothesized to modulate the effect size of the intervention (i.e., modulators), an approach commonly used in cardiovascular disease, for example. This approach assumes that the disorder's pathophysiol-

ogy is well-known, such that a static baseline marker (such as a genotype or other biomarker) can reliably enrich the sample (i.e., distinguish a patient subgroup that is most likely to benefit from a treatment). However, a treatment course in neurobehavioural conditions is a complex, dynamic interplay between a patient and their treatment, including neurobiology that is altered by treatment exposure, expectancy effects, tolerability and engagement with treatment providers. Pretreatment variables are probably not good predictors of these many patient–treatment interactions.

Second, RCT treatment arms typically focus on providing a specific exposure (or dose) of treatment. When flexibility is permitted, it is limited in extent and is typically ad hoc. Clinical medicine has a habit of depicting treatment response as a dose–response curve, a unidimensional pre- to post-treatment change as a function of treatment exposure. However, neurobehavioural treatment response may not obey a dose–response curve; it could evolve in complex, nonlinear ways both acutely and in the longer term, as patients' needs, preferences, psychosocial context and neurobiology change over time.[33]

Third, RCTs for neurobehavioural disorders measure outcomes infrequently with imprecise assessments. Our ability to identify individually relevant brain and behavioural changes depends on a high level of measurement accuracy, combined with appropriate statistical modelling, neither of which is widely used in RCTs.[38] Imprecise assessment can also interfere with predictor measurement, such as behavioural markers with high variability[39] or neuroimaging markers with poor reproducibility.[40]

As a result of these flaws, RCTs do not dynamically or accurately determine the best treatments and tailored regimens, which may change as patient behaviour and neurobiology evolves during the course of treatment.[41]

Precision clinical trials

To overcome these design flaws, we propose a framework that we have termed the precision clinical trial (PCT). This framework comprises of a set of clinical trial design decisions aimed at delivering optimized interventions to individual patients who are most likely to benefit (Fig. 1). The key point is that the PCT is a pragmatic framework that acknowledges that neurobehavioural disorders are complex and their pathophysiology is largely unknown, so that the path to optimized treatments for individual patients requires different clinical trial methods than standard RCTs.

The PCT framework encompasses many clinical research methods that have emerged over the past several decades. These advances have emerged in isolation and are poorly understood by many neurobehavioural researchers, limiting their wide use. We argue that a "tipping point" has been reached because of the convergence of 3 advances: the feasibility of precision measurements via smartphones; the emergence of highly promising biomarkers and statistical methods to model treatment effects in individuals; and new regulatory guidance that provides a clear pathway for precision medicine in treatment development and testing. As a

result, methods for dealing with individual variation are now ready for broader use in clinical trials to advance precision medicine. Some methods are clinical trial designs that arose from treatment models aimed at personalizing treatment to the individual, such as N-of-1 trials (e.g., a patient is randomized to acute bouts of various treatment options to determine the best long-term option for that patient)[42] and stepped-care models (in which patients who do not respond to an initial treatment receive an alternative treatment or an intensification of the initial treatment, such as an additional level of care or support).[43] Other methods arose from conceptual innovations in trial design, such as doubly randomized preference trials (in which patients are randomly assigned to receive either the treatment of their preference or a randomly chosen intervention)[44] and designs that tailor the treatment to the individual patient (e.g., just-in-time adaptive intervention trials and micro-randomized trials in mobile health research, wherein an intervention is individually optimized according to when and how it is needed;[45,46] and sequential multiphase adaptive randomized trials [SMART],[47] in which patients who do not respond to an initial course of treatment are re-randomized to an augmentation or a switch condition). These novel designs not only contribute to the development of treatments that are more efficient, economical and scalable, but also simultaneously enable treatment parameters to be empirically examined and refined through treatment development.[46] As well, advances in measurement of predictors and outcomes allow researchers to obtain precise, granular and individualized measurements via ambulatory smartphone and sensor-based assessments, as well as from preci-

sion brain-mapping and neurophysiological techniques. New statistical methods allow researchers to model the temporal dynamics of illness at an individual level.[33,48,49]

The PCT framework merges these many methodological advances into 3 linked and complementary features: (1) treatment-targeted enrichment, in which PCTs provide a brief bout of the treatment to probe individual patients' brain and/or behavioural response, followed by randomization to a full course of treatment and examination of how patients' response to the brief bout predicts their outcome to the full course; (2) adaptive treatment, in which PCTs test interventions whose parameters change over time to optimize response, in both the short and long terms; and (3) precise measurement, in which PCTs carry out highly accurate, highly reliable brain and behavioural measurements of individual patients, for predictors and for outcomes. To advance precision medicine, all 3 features should be implemented in concert. Figure 2 provides an example of how a clinical trial of cognitive training would use the PCT framework.

Feature 1: treatment-targeted enrichment

The traditional method of testing for precision medicine in RCTs has been a search for baseline enrichment factors: pretreatment biobehavioural measurements that predict (or moderate) a treatment outcome. These predictors are called enrichment factors following genetics parlance,[50,51] but they can just as easily be neuroimaging, peripheral molecular (e.g., proteomic), symptomatic (e.g., comorbid anxiety) or behavioural. However, this "baseline-only" approach to
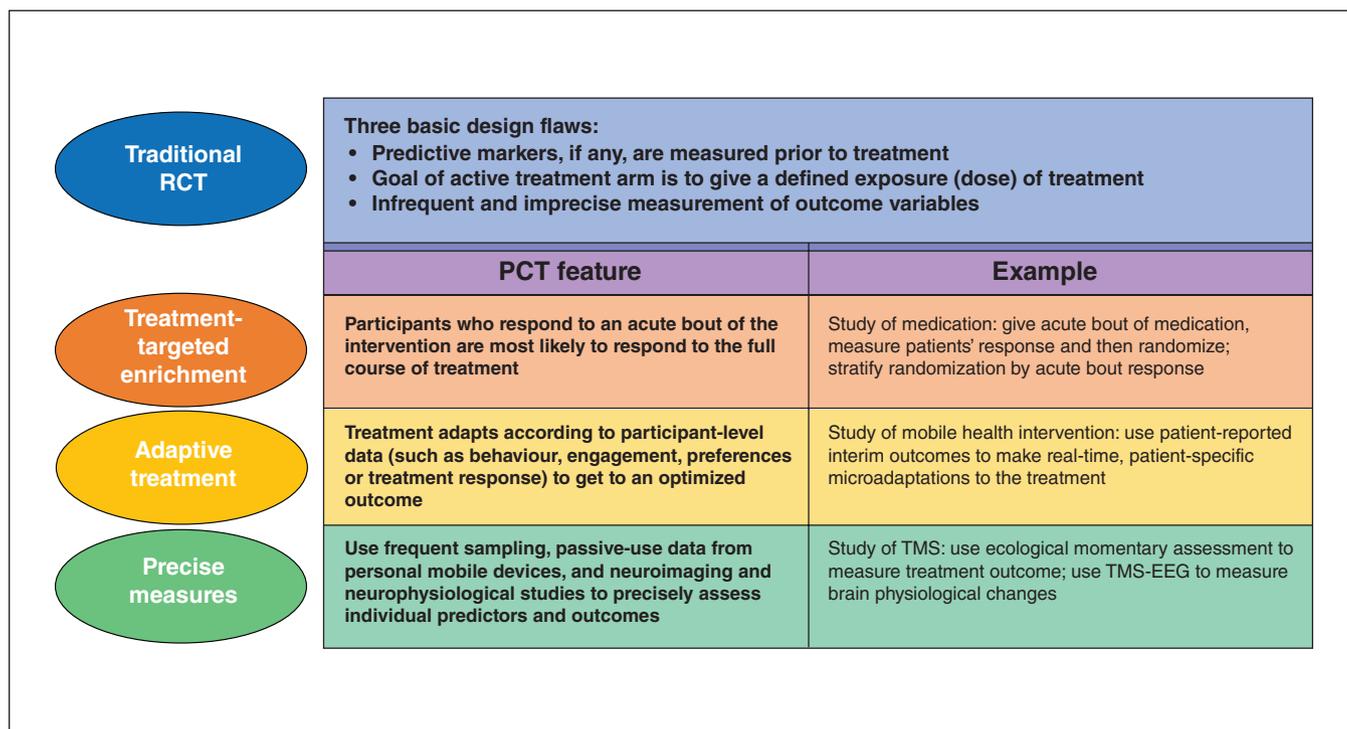

	PCT feature	Example
Traditional RCT	**Three basic design flaws:** • **Predictive markers, if any, are measured prior to treatment** • **Goal of active treatment arm is to give a defined exposure (dose) of treatment** • **Infrequent and imprecise measurement of outcome variables**	
Treatment-targeted enrichment	Participants who respond to an acute bout of the intervention are most likely to respond to the full course of treatment	Study of medication: give acute bout of medication, measure patients' response and then randomize; stratify randomization by acute bout response
Adaptive treatment	Treatment adapts according to participant-level data (such as behaviour, engagement, preferences or treatment response) to get to an optimized outcome	Study of mobile health intervention: use patient-reported interim outcomes to make real-time, patient-specific microadaptations to the treatment
Precise measures	Use frequent sampling, passive-use data from personal mobile devices, and neuroimaging and neurophysiological studies to precisely assess individual predictors and outcomes	Study of TMS: use ecological momentary assessment to measure treatment outcome; use TMS-EEG to measure brain physiological changes

Fig. 1: The precision clinical trial framework. EEG = electroencephalogram; PCT = precision clinical trial; RCT = randomized controlled trial; TMS = transcranial magnetic stimulation.

enrichment factors has been largely unsuccessful in neuro-behavioural disorders, as exemplified by the lack of progress in determining genetic predictors of antidepressant response.[52]

We believe that this lack of success probably reflects the likelihood that neurobehavioural treatment outcomes are a complex interaction between patients and their treatments. Because of this complexity, baseline pretreatment characteristics, whether biological or behavioural, are insufficient to predict patient–treatment interactions, and hence, the treatment outcome. Although it is likely that some interventions are weak, we suspect that sample enrichment based only on pretreatment variables has insufficient sensitivity and specificity to predict treatment response to a high degree.

In contrast, PCTs predict treatment outcomes by directly measuring the interaction between the patient and the treatment. The most straightforward way to do this is to briefly expose all patients to the intervention before randomization and measure their acute response (brain and/or behavioural); examples include cognitive performance changes in response to an acute bout of cognitive training; rapid symptomatic changes in response to a single dose of a medication; or digitally measured behavioural change during a brief exposure to mobile health intervention and self-reported motivation for continuing the treatment. Some of these would be considered run-in clinical trials,[53] in which the response becomes part of the eligibility criteria. Then, patients are randomized and the acute response is used as a predictor variable for post-randomization outcomes.

This design feature, which we term "treatment-targeted enrichment," is built on the concept that an individual's early response to an intervention is likely to predict their subsequent response to that intervention. Patients tend to respond to treatments that worked before, and they tend to stay well on treatments that got them well.[54] For example, a patient's initial acute response to antidepressants or other psychiatric medications predicts longer-term responsivity, such as prevention of relapse.[55] Similarly, N-of-1 trials have evolved in behaviour-modification studies and clinical medicine[56] to optimize a patient's long-term response by randomizing them to brief courses of different treatments and then providing them with a long-term course of the most favourable treatment.[57]

Treatment-targeted enrichment is possible when an individual patient's initial brain or behavioural response to an intervention has the potential to predict their overall longer-term response. Neuromodulation treatments have this potential; for example, 1 dose of repetitive transcranial magnetic stimulation (rTMS) can produce putatively beneficial cognitive effects.[58] Similarly, behavioural treatments such as exposure therapy guide treatment based on initial patient reactions to acute exposures. Many pharmacological treatments with the ability to effect rapid symptom response (such as ketamine and neurosteroids) could feasibly serve as acute predictors of long-term response.[59] Even when rapid effects cannot be detected using traditional outcome assessments, many treatments under investigation have mechanisms that can be tested in humans using noninvasive neurobiological measures (including electroencepalography [EEG][60] and functional MRI [fMRI]),[61] and cognitive and behavioural phenotyping with mobile technology,[62]

Example: A novel, 12-week course of computerized working memory training, designed to improve cognitive dysfunction in anxiety disorders

Traditional randomized controlled trial
Randomize to either working memory training or sham, fix treatment parameters throughout the 12-week intervention, and measure outcome with a single neuropsychological battery and a single anxiety assessment at baseline and end of treatment

Precision clinical trial
1. Treatment-targeted enrichment: Participants complete a baseline working memory task while undergoing fMRI, followed by an acute bout of in-scanner working memory training. After the scan, randomize participants but stratify by degree of increased inactivity in the dlPFC after training. Hypothesis: those with greater activity increases in the dlPFC will show greater cognitive benefit from training

2. Self-adapting treatment: Participants who are randomized to treatment perform the working memory training in-scanner again 2 weeks after treatment initiation. Participants with evidence of low target engagement (little or no change in dlPFC activity from baseline) have their treatment adjusted, either by increasing the difficulty of the working memory training or switching to an alternative treatment. Participants with declining engagement in the training receive a phone call from a coach with motivational feedback to encourage continuation

3. Precise measurement of outcomes: Ten days of smartphone-based brief cognitive assessments and ecological momentary assessment of anxiety symptoms, 3 times daily, at baseline and end of treatment

Fig. 2: Example of a precision clinical trial of cognitive training. dlPFC = dorsolateral prefrontal cortex; fMRI = functional MRI.

potentially yielding robust predictors of treatment response. Treatment-targeted enrichment could also assess patient adherence and engagement during an acute bout as predictors of longer-term adherence, engagement and response.

Focusing trials on patients who are most likely to respond is conceptually appealing, and it may increase trial success rates and improve power for the analyses that will help to determine why participants respond.[63] In this way, the PCT framework is concordant with the experimental therapeutics agenda of the NIH, which requires target engagement tests in clinical trials.[64] It will be important to first explore and then validate these treatment-based enrichment factors, including how and when to test them and how to use them in a clinical trial or a sequence of trials (e.g., initially exploratory, then as stratification variables and ultimately as inclusion criteria). These strategies to develop, explore and validate enrichment strategies are actually encouraged under current guidance from the US Food and Drug Administration (FDA).[65]

One source of resistance to the concept of treatment-targeted enrichment is that if such predictive factors exist, why have we not discovered them already? One answer is that, in spite of the existence of individual differences in response,[66] little neurobehavioural research has examined treatment-interactive effects as predictors of ultimate response; this stands in contrast to robust and even transformative work from other fields (such as oncology) with treatment-targeted enrichment.[67–69] As well, we have an increasing ability to precisely measure acute brain and behavioural changes using emerging techniques such as precision brain-mapping or digital phenotyping as described later. A second source of resistance is concern about the notion of excluding participants who need treatment but who are unlikely to respond. The silver lining here is that such participants are needed for other studies focusing on treatments thought to work via different mechanisms. In fact, focusing participation on treatment-targeted enrichment factors could lead to less reliance on other inclusion criteria, including diagnostic or symptomatic criteria. This could lead the neurobehavioural field to use basket and umbrella clinical trial designs (randomizing patients based on acute response but with a multitude of allowable diagnoses) akin to those seen in medical oncology, where patients are recruited based on genetic enrichment but with a multitude of affected organ systems allowed.[70] For example, a new behavioural treatment might be developed for use in a variety of presentations of anxiety and depression, but patients entering the trial would be limited to those who show a specified acute response to the treatment. In summary, the concept of treatment-targeted enrichment is a well-supported one that has great promise for identifying patients who are most likely to benefit from neurobehavioural treatments.

Including treatment-targeted enrichment factors need not come at the expense of traditional (i.e., pre-treatment) enrichment factors, where there are recent promising findings in mental health (e.g., depression[26,71]) and addiction (e.g., nicotine dependence[72]). However, given the dynamic complexity of treatment for neurobehavioural disorders, it is probable that treatment-targeted factors will be necessary to predict treatment outcome to an extent that would be clinically valuable.

Feature 2: adaptive treatments

Evidence-based treatments are adapted to different contexts (e.g., different subpopulations or settings), and there is a growing interest (emerging from implementation science) on systematically characterizing the many ways a treatment adaptation improves or reduces intervention effectiveness in a subpopulation.[73] To be optimally effective, however, a treatment must also be individually adaptive: that is, it adjusts to individually dynamic changes, such as new or persistent symptoms; changing needs and preferences; or behaviours or life stressors that lead to chronicity or relapse. In the setting of a PCT, treatment is responsive to iterative measures reflecting individually dynamic symptoms, needs and contexts.

Traditional clinical trials often have minimal flexibility, if any (e.g., fixed-dose drug trials, fixed-protocol behavioural interventions), perhaps because of a conservative interpretation of the guidance of "adequate and well-controlled studies" to support a drug's approval.[74] In contrast, adaptive treatment is commonplace in clinical care: medications and doses are adjusted based on individual patients' tolerability and symptom severity; behavioural treatments progress session by session, with clinicians choosing how to proceed and what to emphasize based on patient response; and stepped-care and collaborative care models have built-in intensification of treatment when symptoms persist or worsen. However, such adaptations are typically ad hoc, and data related to adaptation effectiveness, timing and type are often not systematically collected, limiting the ability to assess whether such adaptations actually improve individual treatment outcomes.

Once a patient has been randomized to a treatment (v. control), many methods can be used to adapt the treatment. Clinical trial designs such as SMARTs rigorously test the adaptation by introducing a second randomization for treatment of nonresponders, explicitly testing a specific, predefined treatment augmentation (e.g., dose escalation or timing of delivery) or treatment substitution. Similarly, stepped-care interventions[75] test a treatment sequence rather than a single unchanging intervention, although usually not with subsequent randomizations for each step. The just-in-time adaptive interventions[46] seen in mobile health intervention research iteratively assess response and automatically shift treatment parameters in cases of apparent failure. These are examples of outcome-based adaptation, where changes to treatments are made based on prespecified measurements of treatment response. Such changes typically include dose adjustments and treatment augmentations. Preference-based adaptation, on the other hand, elicits patient needs and preferences and then adapts the intervention to accommodate them. An example of this would be a mobile health intervention that continuously elicits a patient's preferred contact times throughout the day and then sends notifications only during those times. Engagement-based adaptation, common in digital and Internet behavioural treatment, adjusts the treatment to optimize a patient's participation or interaction with the intervention; these could include adjustments to increase either intrinsic or extrinsic (incentive-based) motivation.[76]

Adaptive interventions are meant to maximize individual response in the short term by overcoming barriers to treatment engagement, and in the long term by early detection of clinical deterioration, leading to intensification or change of the intervention to strengthen response. The scientific motivation behind adaptive treatment is to optimize treatment capacity by providing the right amount of treatment at the right time, while eliminating treatment parameters that are not beneficial to an individual patient.[46,77]

Initially, PCTs to optimally compare and test the effectiveness of adaptation strategies may involve stepwise changes or intensifications in treatment and microrandomizations.[78] More advanced trials will include autoresponsive versions of adaptive strategies: what we term "self-adaptive treatments" based on the concept of a self-adaptive system (such as computer software) from engineering and computer science.[79] Self-adaptive treatments will require rich and continuous data inputs, involving both objective or passively collected data (via wearable devices) and subjective or actively collected data direct from participants (via ecological momentary assessment on a mobile or wearable device), as well as machine-learning algorithms that incorporate data in real time to predict an individualized treatment. This scenario has the highest likelihood of success in mobile health treatments for now, but other treatments could be self-adaptive, potentially using smartphone and sensor data integrated with electronic health records and leading to automatic prompts for treatment adaptations.

Precision clinical trials that aim to develop and test self-adapting treatments will require data analytic approaches to model treatment response. Such approaches could illustrate why we should reject the (false) assumption that treatment change is always linear and nomothetic (group-level) and accept that it is often a multidimensional dynamic and idiographic (person-specific) process. For example, some have argued that a dynamic approach with experimental analyses at the individual (or idiographic) level is needed in treatment research.[33] New inferential methods allowing nomothetic findings to serve as prior beliefs for optimal idiographic search procedures provide a potential path forward.[80] We recommend that at minimum, researchers use intuitive treatment adaptations in PCTs, using the growing knowledge on contextual adaptations from implementation science and prospectively testing the benefits of adaptation strategies (e.g., preference-based v. outcome-based) to advance our fundamental knowledge of the most robust strategies.

One final scenario addresses a challenge with device-based treatments: the large number of potentially relevant parameters at the outset of treatment. For example, rTMS pulse timing options alone include high-frequency stimulation, low-frequency stimulation, θ burst stimulation and brain-oscillation-synchronized stimulation;[81] individual patients may preferentially respond to one stimulation frequency and not others. It makes little sense to deliver the treatment at suboptimal parameters and only later adjust them based on ongoing nonresponse. An option, then, is to briefly test each paradigm in each patient, identify each patient's optimal personalized parameters based on acute response and then pro-

vide a full course at these optimal parameters. This testing, akin to N-of-1 trials, may be necessary to increase the effect size of devices such as rTMS for conditions such as depression and cognitive decline. As noted previously, cognitive and neurophysiological measures (such as EEG) change acutely (e.g., within 1 rTMS session) and could guide choice of parameter optimization.

Feature 3: precision measurement

Behaviour is difficult to measure precisely, and a change in behaviour is even more difficult to measure precisely (i.e., with high resolution and reliability). Both the predictor and outcome variables must have a very high degree of reliability to test factors that influence or mediate response.[38]

Measuring change in an outcome is particularly difficult when measurement is infrequent.[38,82] Traditional psychiatric measures attempt to address the challenge of reliability via intensive assessment in the clinic on relatively few occasions. Unfortunately, the result is often a failure to generalize to functioning in real-world settings.[83] As well, these assessments tend to rely heavily on retrospective reporting and averaging (i.e., "In the past 7 days …"). This reliance introduces multiple cognitive biases, such as recency and availability bias, removing the measure from the phenomena of interest by several steps.[84] Infrequent, retrospective assessments are unable to provide the temporal resolution necessary to assess dynamic changes in neurobehavioural symptoms over time (i.e., within and between days) and across different contexts (e.g., home, community and workplace). Further intensifying assessments on a limited number of occasions has not solved these problems, and it is implausible that it would do so.[36] The remaining promising strategy is to increase the frequency of measurement, focusing on a smaller number of high-quality items.[85]

Accordingly, PCTs use precise, repeated, real-time, contextually valid assessments. Ecological momentary assessment[86] (EMA) is an ambulatory data-collection technique that queries present-moment experiences multiple times per day or week, permitting frequent, real-time sampling of thoughts, feelings and behaviours. This approach has many advantages[87] over "one-time, in-clinic" measures, including the generation of more stable estimates of neurobehavioural symptoms for outcomes research;[88] the assessment of dynamic changes among theoretically linked constructs;[89] the reduction of recall bias for patients with neurobehavioural disorders;[90,91] and the ability to model causal intra-individual relationships.[92] More specifically, EMA gives researchers the ability to monitor treatment response in real time and in real-world settings. Because EMA allows for the precise measurement of moment-to-moment changes in symptoms and behaviours, it can also be used to measure acute changes for treatment-targeted enrichment, and to intensively monitor the course of treatment, facilitating adaptations and treatment optimization based on ongoing EMA data.

Additionally, the use of EMA in clinical trials allows for precise estimates of associations between variables. For example, we might design an intervention to block people's

tendency to transition into a depressed mood after the experience of a stressor. Then, we are interested not only in the level of a variable, but also in the association between 2 or more variables (e.g., depressed mood and a stressor). Well-designed EMA studies can precisely measure both variables and their temporal alignment; therefore, they can generate and test interventions designed to affect these associations, examining whether the association changes over time and whether that change predicts treatment outcome.[93] Doing so is responsive to large-scale initiatives such as the Science of Behaviour Change initiative at the NIH[94] and the experimental therapeutics focus and Research Domain Criteria initiative at the National Institute of Mental Health (NIMH).[95]

The history of science shows that improved measurement drives new discoveries. Use of EMA has a clear conceptual superiority to standard measurement,[96] because we want to find out what helps patients in their daily lives. This approach has been used in clinical trials,[97] but only to a very limited extent, leading some to describe it as a "neglected methodology."[98] For example, a search of clinicaltrials.gov (performed July 20, 2019) for completed phase 2, 3 or 4 clinical trials that included the term "ecological momentary assessment" anywhere in the registration found only 43 studies, compared to 7820 studies found in the same search with "depression" instead of EMA. A review of these studies found that a modest number focused on smoking ($n = 10$) and mood disorders ($n = 9$), and handful focused on anxiety and stress ($n = 6$), eating disorders and obesity ($n = 3$), or HIV ($n = 4$). A vast number of neurobehavioural conditions were not represented at all. This may not be surprising, because EMA has traditionally been viewed as an expensive and burdensome measurement technique without a clearly articulated benefit. However, these barriers have fallen with the explosion of smartphones and multitude of low-cost EMA options;[99] we argue that a tipping point has been reached, and the time to shift toward EMA is now.

Advances in measurement and analytic methods for precision medicine

Another advance pushing the field toward this tipping point for PCTs is emerging techniques in measurement and statistical analysis that will lead us to the next generation of individualized assessment and treatment.

Emerging measurement methods

Digital phenotyping

Digital phenotyping is the passive gathering of continuous multimodal smartphone sensor data, such as accelerometry, global positioning system data and keyboard use.[100] Data streams automatically generated from smartphones and wearable devices can be combined with collateral data (such as clinical end points or self-report EMA) to generate predictive models of clinical deterioration or treatment response. Because sensor data are passively and continuously collected, often from devices patients already own, this approach is amenable to long-term, real-world monitoring with minimal

burden for patients and capable of capturing dynamic and subtle shifts in behaviour that may serve as early warning signs of relapse or recurrence.

Digital phenotyping is in its infancy; more research is needed to support scalable passive sensor data collection, processing and analyses and to integrate this approach into the clinical assessment and treatment of neurobehavioural disorders.[101] For example, using digital phenotyping to detect behavioural changes that presage worsening depressive symptoms, a PCT could be designed to deliver tailored intervention content when these changes are detected to prevent further clinical deterioration. Depression's presentation tends to be heterogeneous interindividually but stereotypical intraindividually, so patients could be assessed starting at baseline when they are actively depressed and then followed into remission and into the maintenance or relapse prevention stage. Longitudinal data collected as the patient goes out of and then back into depressive states could be used for personalized prediction of depression relapse, and this could be used to trigger intensification of the intervention to forestall relapse.[102] Ultimately, digital phenotyping could provide an individual signature for treatment-based enrichment, aid in response-driven adaptations of treatments and complement (but not likely replace) EMA for precision measurement. For now, PCTs using mobile health technology should also consider collecting digital phenotyping data, using these data to create and validate such digital markers of illness course and response. Employing phenotyping data collection by mobile technology in PCTs that depend less on active patient participation may lead to greater consistency, frequency and accuracy, and may be more amenable to standardization, with incremental potential for the advancement of precision medicine.[103]

Neuroimaging

Neuroimaging refers to a broad class of techniques, but here we focus on fMRI, which indirectly measures neuronal activity across the brain over time at high spatial resolution. Task-based fMRI is a measurement of evoked brain activity in response to an external task, and resting-state fMRI measures correlations in brain activity between different brain areas while a person lies quietly at rest.

Recent advances in neuroimaging data collection and analysis techniques permit highly precise measurement of brain function, making it possible for neuroimaging studies to be incorporated into precision medicine. Precision brain-mapping, which involves the collection of several hours of imaging data from individual patients, has shown individual differences in the organization of functional networks across the cerebral cortex,[104–107] as well as in subcortical structures such as the amygdala.[108] Studies with these long scan times have revealed that both task-based and resting-state brain activity measures can be obtained with high reliability and precision, allowing researchers to focus on individuals rather than on groups. As well, this approach allows for robust and reliable measurement of changes in brain function that occur acutely in response to an intervention.[109]

Neuroimaging-based treatment targets should be defined a priori, because post hoc derivation of treatment targets over a

search space as large as the human brain leads to overfitting to the current data set and limits generalizability to clinical practice. Specific neuroimaging-based treatment targets for depression might include evoked activity in response to a reward (task-based fMRI[110]), connectivity of the default mode network (resting-state fMRI[111]) or a complex multivariate algorithm that uses patterns of activation across the entire brain.[26] Precision imaging techniques naturally lend themselves to the PCT framework by testing whether neural targets are being modulated by treatment, providing potential treatment-targeted enrichment factors. Functional MRI can measure many such targets at once, such as the resting-state connectivity and interconnectivity of multiple brain networks. Specific treatments can then be targeted to individual patients based on their personalized pattern of brain network dysfunction.[112] In one example, a specific package of cognitive training programs could be designed for a patient with depression to correct their pattern of activity in reward and default network circuitry. Patients could be selected for treatment based on their brain response to an acute in-scanner bout of the treatment, consistent with the PCT principle of treatment-targeted enrichment. The treatment could be updated periodically based on a person's neural response to treatment, consistent with the PCT principle of adaptive treatments. Figure 2 depicts these scenarios. An additional possibility would be to use fMRI as the treatment itself, by using real-time measures of brain activity as a source of neurofeedback: patients could be trained to directly modulate their own brain activity, with the training technique adapted to their personalized version of pathology.[113]

TMS-EEG

Concurrent transcranial magnetic stimulation and electroencephalography (TMS-EEG)[60] is another precision technique: the combination of single or paired-pulse TMS with EEG can measure the function and integrity of brain circuits and probe neuroplasticity. In this technique, TMS is used as a probe, rather than as a treatment. Using EEG, we can measure the cortical excitability induced by a single TMS pulse targeting different brain areas, including the motor cortex and the dorsolateral prefrontal cortex. We can also assess how the brain's response to a TMS pulse (equivalent to the individual's motor threshold or a subthreshold pulse) can affect (suppress, for example) its response to a subsequent TMS pulse with various interpulse intervals.[114] This allows inference about N-methyl-D-aspartate receptors, and cortical excitability mediated by γ-aminobutyric acid receptors A and B.[115] Indeed, TMS-EEG is an emerging technique with an explosion of interest because of its potential to provide both mechanistic and highly reliable assessments.[114,116] Therefore, TMS-EEG fits well within the PCT framework: it can be used to identify the brain's acute response to an intervention,[117] potentially allowing for the rapid optimization of treatment parameters or adaptations.

For example, imagine a treatment designed to improve cognitive functioning with a putative mechanism of improving neuroplasticity (including TMS itself, such as θ burst stimulation rTMS). Using TMS-EEG before and after an acute

bout of the intervention will objectively test target engagement of the proposed mechanism and give an early and objective measure of the effect of the intervention on neuroplasticity, predicting end-point response. This early response can be used to adapt the intervention, change its parameters, add a modality or identify nonresponders who will need a completely different intervention.

Advanced analytic techniques

Currently available analysis techniques can already handle the innovations we suggest if they are adapted to the following: intensive longitudinal data with significant missing data despite a high volume of data overall (i.e., measures repeated many times, but some measures given together only a minority of the time); multiple predictors that change value within and between participants; and the ability to specify expected effects ahead of time and determine when data already collected either support or refute the expected effects. One method that includes all of these features is multilevel dynamic structural equation modelling.[118] The key to such methods is likely to be the use of a Bayesian estimator, because they involve vast amounts of data and a high number of parameters. Machine-learning approaches and other Bayesian modelling approaches could be added to this framework so that modelling is conducted automatically throughout the trial instead of at specific times.

Advanced machine-learning algorithms should be developed along with digital phenotyping and other data-rich measurement techniques. Machine-learning models, as opposed to rule-based models, learn from examples and are able to make predictions on previously unseen data.[119,120] Machine learning typically uses hundreds of features (or variables) in a model to make precise predictions about target outcomes. Machine-learning models have been successfully used in a variety of clinical domains, from image-processing in radiology and pathology to predicting clinical outcomes (e.g., early detection of sepsis, readmissions). A recent scoping review identified that more than 300 articles used machine-learning approaches to characterize mental health conditions, including depression, Alzheimer disease, cognitive decline, schizophrenia and suicide.[121] These studies relied on a variety of data sources, including imaging data, surveys, wearable and mobile phone sensors, and data from social media (Twitter, Facebook). However, these studies have mostly been exploratory, and their clinical utility needs further validation. Precision clinical trials using digital phenotyping and other novel measurements as described above (imaging and TMS-EEG) could explore data collected using machine-learning algorithms to provide nuanced and temporal insights into a participant's brain and behavioural response to an intervention; this, in turn, could inform the next PCT to conduct treatment-targeted enrichment and deliver individually tailored interventions. At present, machine-learning approaches are being used for the secondary analysis of data collected from trials — for example, data from wearables or mobile phones — that can provide large amounts of temporal data on patient activities. Such

temporal data from trials have a considerable advantage, because corresponding structured outcomes data are collected as part of the trial, allowing for robust predictive analytics.[122]

With respect to Bayesian estimation, the PCT concept suggests not only that more data should be collected in patients, but also that the importance of particular data will vary across patients. What we need is not knowledge about 1 or 2 key predictors for everyone (the previously practised method), but knowledge about the best predictors for each patient, allowing that this combination may differ person by person. Determining the best predictors for each individual in a group is a complex proposition. Bayesian estimators are particularly efficient for such complex problems, and a novel set of Bayesian estimators has been developed for this purpose.[80,123] These algorithms converge quickly to accurate, predictive behavioural models, telling us what we can expect of individuals rather than what we should expect of the group overall. Further, these novel frameworks can learn from new information, because they use an estimation process that improves automatically as data accumulate in individuals. Because these methods are Bayesian, we can start using them with our best guess (e.g., a "prior") about the individual, which may be based on our knowledge of a group, a theory or data we already have from the individual. The fundamental task is then to evaluate how the person in question differs from our best guess about him or her. This process is termed "differential inference," and it can be performed rapidly when an active learning process optimizes the precise data collected.[124]

One of the advantages of the statistical, machine-learning and Bayesian approaches described above is the ability to apply results directly to individuals. That is, typically when we make statements about one variable predicting another, we make this statement based on group-level designs, so that what we are really saying is that these variables covary across group members. The alternative is an idiographic approach, which examines whether variables covary across time within a single person. Recent years have seen marked increases in calls for such research, along with advances in our ability to conduct such modelling.[33,125] Researchers, as well as research-minded clinicians, should be aware that they need not wait for machine-learning or Bayesian approaches to "come to them"—they should begin thinking about statistical tests and models as applying to individuals instead of only to groups. These approaches are increasingly available and viable for current use.[49]

A research agenda for precision clinical trials

Congruence of PCTs with NIH, FDA and sponsor initiatives

The PCT framework contends that all 3 features — treatment-targeted enrichment, adaptive treatments and precision measurement — are necessary in a trial for advancing precision medicine. In particular, using treatment-targeted enrichment or tailoring without also increasing the precision of measurement is unlikely to succeed in identifying treatment-relevant subgroups or treatment adaptations that increase effect size. Implementing these features may be a wholesale departure from the traditional approach for many trialists. It is important to note, then, that this framework is congruent with several large-scale initiatives within the NIH, particularly in those institutes that deal primarily with neurobehavioural conditions (e.g., NIMH, National Institute on Drug Abuse, National Institute on Alcohol Abuse and Alcoholism, National Institute on Aging, National Center for Complementary and Integrative Health, National Center for Medical Rehabilitation Research). For example, this article has already described how the PCT framework is concordant with the NIH Science of Behaviour Change program and the NIMH experimental therapeutics agenda, which require target engagement tests within clinical trials.[64] As well, the PCT framework is responsive to the NIH Precision Medicine Initiative, although as argued earlier, the framework does call for a focus on in-treatment variables, which is a major departure from the NIH's current focus on baseline, primarily genetic, variables.

Pharmaceutical and device clinical trials for neurobehavioural conditions are largely conducted from companies and are closely tied to the treatment's intellectual property lifecycle. To date there has been little interest in moving from business as usual — namely, conducting small and then large trials within FDA-defined conditions (e.g., major depression), toward the goal of demonstrating a significant (and usually small) group-based treatment effect. However, there are signs of strain in this business-as-usual model.[126] The first sign comes from the high failure rate of new treatments.[127] Precision clinical trials could improve this failure rate; a recent example is the approval of esketamine, the first FDA-approved non-monoaminergic pharmaceutical for major depression. The approval resulted from 1 pivotal trial with PCT features: a randomized withdrawal study in responders with individually optimized dosing.[128] This trial, which demonstrated a large treatment effect size, was conducted along with 3 traditional RCTs, 2 of which failed to show efficacy for their primary end point; esketamine was approved based on the 1 positive traditional trial (of 3) plus the positive PCT. The second sign comes from novel treatments being approved but still underused, probably in part because their expense results in rationing (e.g., through limitations on insurance coverage). Limiting the use of an expensive treatment may be more fruitfully done based on treatment-targeted enrichment (giving the treatment only to those highly likely to benefit) and not arbitrary factors such as having failed numerous less expensive treatments. Third, leaders in industry have made calls for experimental techniques consistent with the PCT framework, stimulating the experimental therapeutics agenda of NIH.[129] Finally, the FDA is encouraging precision medicine. At conferences and in guidance documents, they have described many features of the PCT framework, including lead-in phases for patient enrichment,[65] machine-learning methods to define treatment-relevant patient populations and predict outcomes,[130] and digital technology for more precise assessment ("greater assay sensitivity") in RCTs.[131] As a result, there is now a clear regulatory path for precision medicine using PCT features.

We also note that charitable foundations have innate interests in PCT methodology because of their mission to rapidly translate scientific discoveries into effective neurobehavioural treatments, reducing the morbidity of neurobehavioural disorders. In summary, the PCT framework should be of great interest to funders and regulators of neurobehavioural clinical trials, and trialists can capitalize on this congruence to implement PCT methods into their research.

Implementation of PCTs

Trialists can immediately use the 3 PCT features: an acute phase before randomization with in-depth data collection during that acute phase; the optimization of treatment through step-wise adaptations; and the increased precision of measurement using EMA versions of existing measures, while testing validity and reliability to confirm that the accuracy of measurement is improved. Widespread use of these techniques will advance our application of PCT methodology while creating the conditions to develop newer, more robust, more feasible methods for specific disorders and patient populations. Nevertheless, we do not advocate discarding simple-to-understand heuristics such as "remission" until considerable validation is done for more precise outcome assessment; for now, traditional measures should be carried out alongside PCT features.

Validation of PCT methods and principles

The PCT framework is proposed not as axiomatic but as a wide-open frontier that will need rigorous exploration and testing. Ultimately, wide-scale adoption and acceptability of PCT methods will require demonstration that they advance treatment response at an individual level while improving our understanding of treatment mechanisms, treatment effects and individual-level variability. Paramount for driving implementation of the PCT framework is the validation of its suite of strategies and tools (i.e., determining the reliability and adequate sensitivity and specificity of treatment-based enrichment, self-adapting therapies and precision measurement). Approaches capable of providing high-quality information about individuals at the neurophysiological and behavioural levels, as well as associated data analytic techniques, need both development and testing.[132] Once this outcome is achieved, precision trialists will have a toolbox to draw on rather than only a wide-open frontier to explore. To achieve this goal, each PCT should test not only the effect size of the treatment, but also the validity of the enrichment and parameter optimization strategies.

The PCT framework principles need validation as well. Much of the framework relies on the hypothesis that acute changes are predictive of longer-term changes, which may not be applicable to all treatments. For example, in interpersonal therapy, early gains have not predicted long-term treatment impact, although this scenario may be an artifact of imprecise measurement of change.[133]

The broader deployment of PCT into clinical practice will require advances beyond clinical innovation. Ultimate validation of the PCT approach will lie in its wide-scale use and acceptability to regulators and sponsors. Some PCTs may need large sample sizes to validate these strategies sufficiently to overcome skepticism. Solutions to the large sample size requirement have been proposed, such as decentralized clinical trials.[134] In such studies, where the research investigative site is remote from the participants, some of the same data-collection techniques advocated here — EMA and digital phenotyping — are viewed as the solution to remote outcome assessment. In other words, PCTs and decentralized clinical trials have shared methods and may synergize each other's evolution and implementation.

Pragmatism and feasibility as guides for PCTs

No less important in furthering the goal of PCT implementation in the testing of treatments for neurobehavioural disorders is pragmatism: currently, the amount and quality of data required for precision research can be obtained only with intensive measurements, from hours-long fMRI sessions to the hundreds of EMA measurements in an individual that may be needed for some dynamic analytic models.[135] Even the paradigm of managing the data files produced when using EMA and digital phenotyping may seem overwhelming. These measurement burdens stand in contrast to the "large simple trial" philosophy of making data collection as lean and low-burden as possible in RCTs.

Yet there are reasons to believe that the measures that seem daunting today will be pragmatic in the future. This is because PCTs will highlight the data that are most useful for drawing conclusions about a patient. Although PCTs require working with vast amounts of data at first, they should help us isolate the most informative data and prioritize obtaining and analyzing it in subsequent projects. In other words, the more data we collect initially, the less we will need to collect in the future.

This strategy could easily evolve within the developmental lifecycle of a single treatment. Imagine a new drug developed for depression. Preliminary small patient trials (akin to phase 2a) would invest in measurement with EMA and other techniques relevant to the drug's mechanism, such as TMS-EEG. Intensive measurement would occur within an acute open-label phase to discern how and when treatment-targeted enrichment is most robust, and then in the randomized phase to discern how each patient's course varies and how and when to adapt treatment (e.g., dose changes, augmentation or as-needed counselling). An ensuing trial (phase 2b) would validate these strategies, confirming by replication their value in increasing the effect size of the treatment (as well as other important outcomes such as patient satisfaction). Then, large phase 3 studies could use more parsimonious data collection, assessing each patient only when necessary to inform the likelihood of benefit or treatment adaptation and produce evidence acceptable to regulatory agencies. Finally, by the time the treatment reaches the clinic, these targeting, adaptation and measurement strategies would be made user-friendly for both patient and clinician, with much of the data management and algorithms occurring in the background.

This call for pragmatism and feasibility means that technological changes to improve ease of use are as important as advances in concepts or methods. An example of this is the widespread use of smartphones in the last decade, lowering the barriers to implementing EMA and paving the way for digital phenotyping. Another emerging example is the development of electronic health record–integrated infrastructure for the delivery and collection of patient-reported outcomes), which has opened new avenues for measurement.[136,137] Relying on NIH's Patient-Reported Outcomes Measurement Information System (PROMIS) infrastructure, these patient-reported outcomes can be administered via patient portals on most Internet-enabled devices, and patient responses are incorporated into the electronic health record, allowing it to be used for routine clinical use.[138] Biomarkers such as fMRI and TMS-EEG require continued development so that they could be feasibly conducted in the real world.

Conclusion

The strongest argument for PCTs is ultimately a patient-centred one: we can change what clinical medicine will look like if we adopt PCTs now. An optimist's view is that in 1 generation, the methodological tools in PCTs will be well validated, streamlined and feasible to implement in clinical settings. Clinicians could rapidly assign patients to treatments that they are likely to benefit from and use adaptive algorithms to optimize treatment parameters to individuals.

In this scenario, patients would no longer need to embark on a lengthy trial-and-error process that often leads to suboptimal or ineffective treatment. Instead, treatment at first would be more like an optometry appointment, where a thoughtful process of "which is better, A or B?" determines an optimal treatment choice. Then, the course of treatment would be individually tailored, so that the intervention assesses the patient's needs and outcomes and automatically self-adjusts to optimize positive outcomes, both acutely and over the long term. Finally, patients would benefit from a wider range of treatment options, because treatments with small overall population effects could be discovered to have a large effect in a certain subgroup, particularly when optimized to individual patients.

If this scenario is correct — and individual tailoring of neurobehavioural treatments is possible — then the public health benefits from PCTs would be enormous, given the high burden of these disorders and the suboptimal results of treatment at present. The RCT is long due for an overhaul for neurobehavioural disorders, and the adoption of PCTs will get us, and our patients, to precision medicine.

Affiliations: From the Washington University School of Medicine, St. Louis, Missouri (Lenze, Nicol, Kannampallil Wong, Piccirillo, Drysdale, Sylvester, Haddad, Miller, Lenze, Freedland); the Washington University McKelvey School of Engineering, St. Louis, MO (Barbour); the University of Pittsburgh, Pittsburgh, PA (Low); and the Washington University School of Arts & Sciences, St. Louis, MO (Rodebaugh).

References

1. Vainik U, Baker TE, Dadar M, et al. Neurobehavioral correlates of obesity are largely heritable. *Proc Natl Acad Sci U S A* 2018;115: 9312-7.
2. Murray CJ, Atkinson C, Bhalla K, et al. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA* 2013;310:591-608.
3. Medicine Io, Council NR. *Measuring the risks and causes of premature death: summary of workshops.* Washington (DC): National Academies Press; 2015.
4. Castren E, Kojima M. Brain-derived neurotrophic factor in mood disorders and antidepressant treatments. *Neurobiol Dis* 2017;97: 119-26.
5. Autry AE, Monteggia LM. Brain-derived neurotrophic factor and neuropsychiatric disorders. *Pharmacol Rev* 2012;64:238-58.
6. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014;63(25 Pt B):2889-934.
7. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts). Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J* 2016;37:2315-81.
8. Gupta A, Mackay J, Whitehouse A, et al. Long-term mortality after blood pressure-lowering and lipid-lowering treatment in patients with hypertension in the Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT) legacy study: 16-year follow-up results of a randomised factorial trial. *Lancet* 2018;392:1127-37.
9. Khan A, Mar KF, Brown WA. The conundrum of depression clinical trials: one size does not fit all. *Int Clin Psychopharmacol* 2018;33:239-48.
10. Bokma WA, Wetzer G, Gehrels JB, et al. Aligning the many definitions of treatment resistance in anxiety disorders: a systematic review. *Depress Anxiety* 2019;36:801-12.
11. de Sousa RT, Zanetti MV, Brunoni AR, et al. Challenging treatment-resistant major depressive disorder: a roadmap for improved therapeutics. *Curr Neuropharmacol* 2015;13:616-35.
12. Nicol GE, Yingling MD, Flavin KS, et al. Metabolic effects of antipsychotics on adiposity and insulin sensitivity in youths: a randomized clinical trial. *JAMA Psychiatry* 2018;75:788-96.
13. Wiles NJ, Thomas L, Turner N, et al. Long-term effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial. *Lancet Psychiatry* 2016;3:137-44.
14. Lenze EJ. Solving the geriatric mental health crisis in the 21st century. *JAMA Psychiatry* 2015;72:967-8.
15. Arean PA, Raue PJ, Sirey JA, et al. Implementing evidence-based psychotherapies in settings serving older adults: challenges and solutions. *Psychiatr Serv* 2012;63:605-7.
16. National Academies of Sciences. *Psychosocial interventions for mental and substance use disorders: a framework for establishing evidence-based standards.* Washington (DC): Institute of Medicine; 2015. Available: http://nationalacademies.org/hmd/~/media/Files/Report%20Files/2015/Psychosocial-Report-in-Brief.pdf (accessed 2020 Oct. 29).
17. Lincoln TM, Rief W, Hahlweg K, et al. Who comes, who stays, who profits? predicting refusal, dropout, success, and relapse in a short intervention for social phobia. *Psychother Res* 2005;15:210-25.
18. Metias EF, Cunningham DJ, Howson MG, et al. Reflex effects on human breathing of breath-by-breath changes of the time profile of alveolar PCO2 during steady hypoxia. *Pflugers Arch* 1981;389:243-50.
19. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. *Rheum Dis Clin North Am* 2019;45:159-72.
20. Taylor M, Jauhar S. Are we getting any better at staying better? The long view on relapse and recovery in first episode nonaffective psychosis and schizophrenia. *Ther Adv Psychopharmacol* 2019;9: 2045125319870033.
21. Gueorguieva R, Chekroud AM, Krystal JH. Trajectories of relapse in randomised, placebo-controlled trials of treatment discontinuation in major depressive disorder: an individual patient-level data meta-analysis. *Lancet Psychiatry* 2017;4:230-7.
22. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793-5.
23. Whitcomb DC. Primer on precision medicine for complex chronic disorders. *Clin Transl Gastroenterol* 2019;10:e00067.
24. Stekhoven DJ. *missForest: Nonparametric missing value imputation using Random Forest. R package version 1.4.* 2013. Available: www.r-project.org, https://github.com/stekhoven/missForest (accessed 2020 Oct. 29).
25. Chekroud AM, Gueorguieva R, Krumholz HM, et al. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry* 2017;74:370-8.
26. Drysdale AT, Grosenick L, Downar J, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 2017;23:28-38.
27. Joyner MJ, Paneth N. Promises, promises, and precision medicine. *J Clin Invest* 2019;129:946-8.
28. Hack LM, Fries GR, Eyre HA, et al. Moving pharmacoepigenetics tools for depression toward clinical use. *J Affect Disord* 2019;249:336-46.
29. Boeke EA, Holmes AJ, Phelps EA. Toward robust anxiety biomarkers: a machine learning approach in a large-scale sample. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2020;5:799-807.
30. Brand SJ, Moller M, Harvey BH. A review of biomarkers in mood and psychotic disorders: a dissection of clinical vs. preclinical correlates. *Curr Neuropharmacol* 2015;13:324-68.
31. Volkow ND, Koob G, Baler R. Biomarkers in substance use disorders. *ACS Chem Neurosci* 2015;6:522-5.
32. Galmiche M, Déchelotte P, Lambert G, et al. Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. *Am J Clin Nutr* 2019;109:1402-13.
33. Hofmann SG, Curtiss JE, Hayes SC. Beyond linear mediation: toward a dynamic network approach to study treatment processes. *Clin Psychol Rev* 2020;76:101824.
34. Senn S. Statistical pitfalls of personalized medicine. *Nature* 2018;563:619-21.
35. Stiles WB, Leach C, Barkham M, et al. Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *J Consult Clin Psychol* 2003;71:14-21.
36. Mofsen AM, Rodebaugh TL, Nicol GE, et al. When all else fails, listen to the patient: a viewpoint on the use of ecological momentary assessment in clinical trials. *JMIR Ment Health* 2019;6:e11845.
37. Haberer JE, Trabin T, Klinkman M. Furthering the reliable and valid measurement of mental health screening, diagnoses, treatment and outcomes through health information technology. *Gen Hosp Psychiatry* 2013;35:349-53.
38. Rodebaugh TL, Scullin RB, Langer JK, et al. Unreliability as a threat to understanding psychopathology: the cautionary tale of attentional bias. *J Abnorm Psychol* 2016;125:840-51.
39. Conway CC, Forbes MK, Forbush KT, et al. A hierarchical taxonomy of psychopathology can transform mental health research. *Perspect Psychol Sci* 2019;14:419-36.
40. Poldrack RA, Baker CI, Durnez J, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci* 2017;18:115-26.
41. Ho D, Quake SR, McCabe ERB, et al. Enabling technologies for personalized and precision medicine. *Trends Biotechnol* 2020;38: 497-518.
42. Offord C. N-of-1 studies tackle limitations of randomized controlled trials. *Scientist* 2019 Jul/Aug. Available: www.the-scientist.com/infographics/infographic--n-of-1-studies-tackle-limitations-of-randomized-controlled-trials-66138 (accessed 2020 Nov. 3).
43. Richards DA, Bower P, Pagel C, et al. Delivering stepped care: an analysis of implementation in routine practice. *Implement Sci* 2012;7:3.
44. Zoellner LA, Roy-Byrne PP, Mavissakalian M, et al. Doubly randomized preference trial of prolonged exposure versus sertraline for treatment of PTSD. *Am J Psychiatry* 2019;176:287-96.
45. Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol* 2015;34S:1220-8.
46. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and

design principles for ongoing health behavior support. *Ann Behav Med* 2018;52:446-62.

47. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007;32(Suppl):S112-8.

48. Cao B, Zheng L, Zhang C, et al. DeepMood: modeling mobile phone typing dynamics for mood detection [presentation]. Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining, 23rd International Conference; 2017 Aug. 13–17; Halifax (NS).

49. Piccirillo ML, Beck ED, Rodebaugh TL. A clinician's primer for idiographic research: considerations and recommendations. *Behav Ther* 2019;50:938-51.

50. Reay WR, Atkins JR, Carr VJ, et al. Pharmacological enrichment of polygenic risk for precision medicine in complex disorders. *Sci Rep* 2020;10:879.

51. Antypa N, Drago A, Serretti A. Genomewide interaction and enrichment analysis on antidepressant response. *Psychol Med* 2014;44:753-65.

52. Fabbri C, Serretti A. Clinical application of antidepressant pharmacogenetics: considerations for the design of future studies. *Neurosci Lett* 2018;726:133651.

53. Laursen DRT, Paludan-Muller AS, Hrobjartsson A. Randomized clinical trials with run-in periods: frequency, characteristics and reporting. *Clin Epidemiol* 2019;11:169-84.

54. Kennedy S, McIntyre R, Fallu A, et al. Pharmacotherapy to sustain the fully remitted state. *J Psychiatry Neurosci* 2002;27:269-80.

55. Borges S, Chen YF, Laughren TP, et al. Review of maintenance trials for major depressive disorder: a 25-year perspective from the US Food and Drug Administration. *J Clin Psychiatry* 2014;75:205-14.

56. Kazdin AE. *Single-case research designs: methods for clinical and applied settings.* 2nd ed. New York: Oxford University Press; 1982.

57. McDonald S, Quinn F, Vieira R, et al. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychol Rev* 2017;11:307-23.

58. Barr MS, Farzan F, Rajji TK, et al. Can repetitive magnetic stimulation improve cognition in schizophrenia? Pilot data from a randomized controlled trial. *Biol Psychiatry* 2013;73:510-7.

59. Wilkinson ST, Sanacora G. A new generation of antidepressants: an update on the pharmaceutical pipeline for novel and rapid-acting therapeutics in mood disorders based on glutamate/GABA neurotransmitter systems. *Drug Discov Today* 2019;24:606-15.

60. Tremblay S, Rogasch NC, Premoli I, et al. Clinical utility and prospective of TMS-EEG. *Clin Neurophysiol* 2019;130:802-44.

61. Dichter GS, Gibbs D, Smoski MJ. A systematic review of relations between resting-state functional-MRI and treatment response in major depressive disorder. *J Affect Disord* 2015;172:8-17.

62. Jacobson NC, Weingarden H, Wilhelm S. Using digital phenotyping to accurately detect depression severity. *J Nerv Ment Dis* 2019; 207:893-6.

63. Fairchild AJ, MacKinnon DP. A general model for testing mediation and moderation effects. *Prev Sci* 2009;10:87-99.

64. Lewandowski KE, Ongur D, Keshavan MS. Development of novel behavioral interventions in an experimental therapeutics world: challenges, and directions for the future. *Schizophr Res* 2018;192:6-8.

65. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). *Enrichment strategies for clinical trials to support: determination of effectiveness of human drugs and biological products guidance for industry.* Rockville (MD): Food and Drug Administration; 2019. Available: www.fda.gov/regulatory-information/search-fda-guidance-documents/enrichment-strategies-clinical-trials-support-approval-human-drugs-and-biological-products (accessed 2020 Oct. 29).

66. Maslej MM, Furukawa TA, Cipriani A, et al. Individual differences in response to antidepressants: a meta-analysis of placebo-controlled randomized clinical trials. *JAMA Psychiatry* 2020;77:1-12.

67. Sanford JA, Nogiec CD, Lindholm ME, et al. Molecular Transducers of Physical Activity Consortium (MoTrPAC): mapping the dynamic responses to exercise. *Cell* 2020;181:1464-74.

68. Paharik AE, Schreiber HL, Spaulding CN, et al. Narrowing the spectrum: the new frontier of precision antimicrobials. *Genome Med* 2017;9:110.

69. Gao M, Warner J, Yang P, et al. On the Bayesian derivation of a treatment-based cancer ontology. *AMIA Jt Summits Transl Sci Proc* 2014;2014:209-17.

70. Garralda E, Dienstmann R, Piris-Gimenez A, et al. New clinical trial designs in the era of precision medicine. *Mol Oncol* 2019;13:549-57.

71. Wu W, Zhang Y, Jiang J, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nat Biotechnol* 2020;38:439-47.

72. Chen LS, Horton A, Bierut L. Pathways to precision medicine in smoking cessation treatments. *Neurosci Lett* 2018;669:83-92.

73. Chambers DA, Norton WE. The adaptome: advancing the science of intervention adaptation. *Am J Prev Med* 2016;51(Suppl 2):S124-31.

74. US Food and Drug Administration. Code of federal regulations, title 21, vol. 5. Silver Spring (MD): US Food and Drug Administration; 2019. Available: www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=314.126 (accessed 2020 Nov. 3).

75. Alexopoulos GS, Arean P. A model for streamlining psychotherapy in the RDoC era: the example of "Engage." *Mol Psychiatry* 2014;19:14-9.

76. Nurmi J, Knittle K, Ginchev T, et al. Engaging users in the behavior change process with digitalized motivational interviewing and gamification: development and feasibility testing of the precious app. *JMIR Mhealth Uhealth* 2020;8:e12884.

77. Gonul S, Namli T, Huisman S, et al. An expandable approach for design and personalization of digital, just-in-time adaptive interventions. *J Am Med Inform Assoc* 2019;26:198-210.

78. Addicott MA, Pearson JM, Sweitzer MM, et al. A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology* 2017;42:1931-9.

79. Salehie M, Tahvildari L. Self-adaptive software: landscape and research challenges. *ACM T Auton Adapt Sys* 2009;V(N):1.

80. Barbour DL, DiLorenzo JC, Sukesan KA, et al. Conjoint psychometric field estimation for bilateral audiometry. *Behav Res Methods* 2019;51:1271-85.

81. Caulfield KA. Is Accelerated, high-dose theta burst stimulation a panacea for treatment-resistant depression? *J Neurophysiol* 2020;123:103.

82. Rogosa D. Myths about longitudinal research. In: Schaie KW, Campbell RT, Meredith WM, Rawlings SC, editors. *Methodological issues in aging research.* New York: Springer; 1988:171-209.

83. Myin-Germeys I, Kasanova Z, Vaessen T, et al. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry* 2018;17:123-32.

84. Solhan MB, Trull TJ, Jahng S, et al. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. *Psychol Assess* 2009;21:425-36.

85. Nunnally JD, Bernstein IH. *Psychometric theory.* New York: McGraw; 1994.

86. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1-32.

87. Bos FM, Schoevers RA. aan het Rot M. Experience sampling and ecological momentary assessment studies in psychopharmacology: a systematic review. *Eur Neuropsychopharmacol* 2015;25:1853-64.

88. Moore RC, Depp CA, Wetherell JL, et al. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. *J Psychiatr Res* 2016;75:116-23.

89. Depp CA, Moore RC, Dev SI, et al. The temporal course and clinical correlates of subjective impulsivity in bipolar disorder as revealed through ecological momentary assessment. *J Affect Disord* 2016;193:145-50.

90. Goldberg RL, Piccirillo ML, Nicklaus J, et al. Evaluation of ecological momentary assessment for tinnitus severity. *JAMA Otolaryngol Head Neck Surg* 2017;143:700-6.

91. Ramsey AT, Wetherell JL, Depp C, et al. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. *J Technol Hum Serv* 2016;34:209-23.

92. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. *J Psychiatr Res* 2009;43:987-96.

93. Depp CA, Moore RC, Perivoliotis D, et al. Technology to assess and support self-management in serious mental illness. *Dialogues Clin Neurosci* 2016;18:171-83.

94. Riddle M. Science of Behavior Change Working Group. News from the NIH: using an experimental medicine approach to facilitate translational research. *Transl Behav Med* 2015;5:486-8.

95. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013;11:126.

96. Carlson EB, Field NP, Ruzek JI, et al. Advantages and psychometric validation of proximal intensive assessments of patient-reported outcomes collected in daily life. *Qual Life Res* 2016;25:507-16.

97. Moskowitz DS, Young SN. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *J Psychiatry Neurosci* 2006;31:13-20.

98. Davidson CL, Anestis MD, Gutierrez PM. Ecological momentary assessment is a neglected methodology in suicidology. *Arch Suicide Res* 2017;21:1-11.

99. Burke LE, Shiffman S, Music E, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. *J Med Internet Res* 2017;19:e77.

100. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016;41:1691-6.

101. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ Digit Med.* 2019;2:88.

102. Gale CR, Westbury L, Cooper C. Social isolation and loneliness as risk factors for the progression of frailty: the English Longitudinal Study of Ageing. *Age Ageing* 2018;47:392-7.

103. Cohen AB, Mathews SC. The digital outcome measure. *Digital Biomarkers*. 2018;2:94-105.

104. Miranda-Dominguez O, Mills BD, Carpenter SD, et al. Connectotyping: model based fingerprinting of the functional connectome. *PLoS One* 2014;9:e111048.

105. Wang D, Buckner RL, Fox MD, et al. Parcellating cortical functional networks in individuals. *Nat Neurosci* 2015;18:1853-60.

106. Gratton C, Laumann TO, Nielsen AN, et al. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* 2018;98:439-52.

107. Gordon EM, Laumann TO, Gilmore AW, et al. Precision functional mapping of individual human brains. *Neuron* 2017;95:791-807.

108. Sylvester CM, Yu Q, Srivastava AB, et al. Individual-specific functional connectivity of the amygdala: a substrate for precision psychiatry. *Proc Natl Acad Sci U S A* 2020;117:3808-18.

109. Newbold DJ, Laumann TO, Hoyt CR, et al. Pulses of spontaneous activity drive functional connectivity changes in disused brain networks [poster]. Organization for Human Brain Mapping; 2019 June 12; Rome.

110. Keren H, O'Callaghan G, Vidal-Ribas P, et al. Reward processing in depression: a conceptual and meta-analytic review across fMRI and EEG studies. *Am J Psychiatry* 2018;175:1111-20.

111. Hamilton JP, Farmer M, Fogelman P, et al. Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biol Psychiatry* 2015;78:224-30.

112. Sylvester CM, Corbetta M, Raichle ME, et al. Functional network dysfunction in anxiety and anxiety disorders. *Trends Neurosci* 2012; 35:527-35.

113. Watanabe T, Sasaki Y, Shibata K, et al. Advances in fMRI real-time neurofeedback. *Trends Cogn Sci* 2017;21:997-1010.

114. Lioumis P, Zomorrodi-Moghaddam R, Hadas I, et al. Combined transcranial magnetic stimulation and electroencephalography of the dorsolateral prefrontal cortex. *J Vis Exp* 2018;138:e57983.

115. Lenz M, Vlachos A. Releasing the cortical brake by non-invasive electromagnetic stimulation? rTMS induces LTD of GABAergic neurotransmission. *Front Neural Circuits* 2016;10:96.

116. Kerwin LJ, Keller CJ, Wu W, et al. Test-retest reliability of transcranial magnetic stimulation EEG evoked potentials. *Brain Stimul* 2018;11:536-44.

117. Farzan F, Vernet M, Shafi MM, et al. Characterizing and modulating brain circuitry through transcranial magnetic stimulation combined with electroencephalography. *Front Neural Circuits* 2016;10:73.

118. Hamaker EL, Asparouhov T, Brose A, et al. At the frontiers of modeling intensive longitudinal data: dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behav Res* 2018;53:820-41.

119. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44-56.

120. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380:1347-58.

121. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 2019;49:1426-48.

122. Hicks JL, Althoff T, Sosic R, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ Digit Med* 2019;2:45.

123. Song XD, Sukesan KA, Barbour DL. Bayesian active probabilistic classification for psychometric field estimation. *Atten Percept Psychophys* 2018;80:798-812.

124. Larsen TJ, Malkomes G, Barbour DL. Accelerating psychometric screening tests with Bayesian active differential selection. *arXiv* 2020:2002.01547. Available: https://arxiv.org/abs/2002.01547 (accessed 2020 Nov. 3).

125. Piccirillo ML, Rodebaugh TL. Foundations of idiographic methods in psychology and applications for psychotherapy. *Clin Psychol Rev* 2019;71:90-100.

126. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016;47:20-33.

127. Gribkoff VK, Kaczmarek LK. The need for new approaches in CNS drug discovery: why drugs have failed, and what can be done to improve outcomes. *Neuropharmacology* 2017;120:11-9.

128. Daly EJ, Trivedi MH, Janik A, et al. Efficacy of esketamine nasal spray plus oral antidepressant treatment for relapse prevention in patients with treatment-resistant depression: a randomized clinical trial. *JAMA Psychiatry* 2019;76: 893-903.

129. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;9:203-14.

130. Hesser H, Andersson G. The role of anxiety sensitivity and behavioral avoidance in tinnitus disability. *Int J Audiol* 2009;48.

131. Gottlieb S. Breaking down barriers between clinical trials and clinical care: incorporating real world evidence into regulatory decision making [lecture]. Bipartisan Policy Center conference; 2019 Jan. 28; Washington (DC). Available: www.fda.gov/news-events/speeches-fda-officials/breaking-down-barriers-between-clinical-trials-and-clinical-care-incorporating-real-world-evidence (accessed 2020 Oct. 29).

132. Frank E, Pong J, Asher Y, et al. Smart phone technologies and ecological momentary data: is this the way forward on depression management and research? *Curr Opin Psychiatry* 2018;31:3-6.

133. Kelly MA, Cyranowski JM, Frank E. Sudden gains in interpersonal psychotherapy for depression. *Behav Res Ther* 2007;45:2563-72.

134. Sommer C, Zuccolin D, Arnera V, et al. Building clinical trials around patients: evaluation and comparison of decentralized and conventional site models in patients with low back pain. *Contemp Clin Trials Commun.* 2018;11:120-6.

135. Gerull KM, Kallogjeri D, Piccirillo ML, et al. Feasibility of intensive ecological sampling of tinnitus in intervention research. *Otolaryngol Head Neck Surg* 2019;161:485-92.

136. Wagner LI, Schink J, Bass M, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer* 2015;121:927-34.

137. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007;45(Suppl 1):S3-11.

138. Rothrock N, Barnard C, Bhatt S, et al. Evaluation of the implementation of PROMIS CAT batteries for total joint arthroplasty in an electronic health record. *Qual Life Res* 2018;27:S30.