American College of Neuropsychopharmacology

www.nature.com/npp

REVIEW ARTICLE

How should we design future mechanistic and/or efficacy clinical trials?

Maurizio Fava[1] ✉

The emergence of new molecular targets, together with the development of new approaches to neuropsychiatric diseases, involving psychedelics as well as gene and cell therapies, are creating the need to improve the efficiency of mechanistic and/or efficacy clinical trials. This review article will discuss a number of issues that have hampered our ability to detect therapeutic signals, from excessive placebo/sham response rates to the imprecision of diagnostic and outcome assessments. In addition to reviewing the limitations of current efficacy and mechanistic neuropsychiatric clinical trials, this review presents some of the methodological approaches that may improve the overall performance of our neuropsychiatric trials, including the adoption of novel study designs such as the sequential parallel comparison design and independent confirmation of the appropriateness of subjects' enrollment. In addition, this review will discuss several designs that make mechanistic clinical trials more precise.

Neuropsychopharmacology; https://doi.org/10.1038/s41386-023-01600-9

INTRODUCTION

We are currently experiencing a tremendous growth of neuroscience research, which is leading to the development of novel therapies for neuropsychiatric disorders. The emergence of both novel molecular targets and approaches to neuropsychiatric diseases has re-energized the field of neuropsychiatric treatment development, while creating the need to improve the efficiency of mechanistic and/or efficacy clinical trials. Multiple methodological issues have greatly limited our ability to detect therapeutic signals, from excessive placebo/sham response rates to the imprecision of diagnostic and outcome assessments. These issues have affected some disorders (e.g., pain, mood and anxiety disorders, schizophrenia) more than others (e.g., epilepsy, multiple sclerosis). In addition to reviewing the limitations of current efficacy and mechanistic neuropsychiatric clinical trials, this review will also present some of the methodological approaches that may improve the overall performance of our neuropsychiatric trials, including the adoption of novel study designs such as the sequential parallel comparison design. This review focuses primarily on work carried out by our research group over the past 30 years and is not intended to serve as a systematic or comprehensive review of all the methods that have been developed by the field to de-risk neuropsychiatric clinical trials.

LIMITATIONS OF CURRENT EFFICACY CLINICAL TRIALS IN NEUROPSYCHIATRIC DISORDERS

Efficacy clinical trials in neuropsychiatric disorders are limited by a number of factors which diminish the investigators' ability to detect therapeutic signals. These factors can vary across indications of the trials and across geographical areas (some trials have better outcome in the US, others outside the US). Here are the major factors:

(1) Excessive reliance on subjective endpoints: through the use of clinician-rated and/or self-rated measures, many neuropsychiatric trials depend on the ability of these measures to capture symptomatic and functional changes. Such measures may be valid and reliable, but may lack sensitivity, in that they may perform less well than others in discriminating between the effects of active treatment and placebo. In addition, in the case of the use of clinician-rated measures, variability in level of training and experience among raters may lead to poorer inter-rater reliability.
(2) Relative lack of ecological validity and generalizability: this is a common problem, particularly in registration trials, which tend to exclude many comorbid conditions primarily because of the concern that these conditions may affect their safety and tolerability reports. An analysis of the largest clinical trial in major depressive disorder (MDD) utilizing effectiveness trial inclusion criteria, the STAR*D study, showed that, among 2855 participants, only 22.2% met typical entry criteria for phase III clinical trials (efficacy sample) and 77.8% did not (non-efficacy sample), suggesting that phase III trials do not recruit representative treatment-seeking MDD patients [1]. Similarly, a review of placebo-controlled antidepressant efficacy trials published over 20 years to determine whether there has been a change in the symptom severity inclusion criterion threshold showed that results of these trials may not be applicable to less severely depressed patients who make up at least half of the patients treated in routine clinical practice [2].
(3) Inappropriate enrollment of subjects: the strong enrollment bias among investigators may lead to the inclusion of subjects who may not meet the inclusion criteria. A common aspect of this bias is the score inflation that occurs when a severity of illness threshold is applied to

[1]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. ✉email: mfava@mgh.harvard.edu

Received: 28 February 2023 Revised: 10 April 2023 Accepted: 24 April 2023
Published online: 26 May 2023

enter a trial. A study reported by Mundt et al. [3] showed that, in an MDD trial, the distributions of baseline scores of the clinicians' assessments of severity of illness were severely skewed, with a correlation with an independent self-rated method of severity of illness severity being only 0.50 at baseline, and the independent scores averaging 6.19 points less than clinician scores ($p < 0.001$). However, by study end point, both assessment methods were more concordant (correlation: 0.75; independent scores averaging just 1.41 points lower than the clinician scores. Diagnostic misclassification may also lead to the inclusion of subjects who do not have the psychiatric or neurological disorder under investigation. Such misclassification may be in some cases the result of the poor specificity of the diagnostic instruments used (thereby leading to a high rate of false positives) or of the less than adequate inter-rater reliability in the administration of diagnostic instruments, or of clinician biases to enroll. "Professional patients", or duplicate subjects, are also a common problem in clinical trials, as they may contribute to "noise" in clinical trials.

(4) Excessive placebo response rates: this phenomenon has affected many neuropsychiatric conditions such as pain [4], MDD [5, 6] and schizophrenia [6, 7], where we have clearly seen rising placebo response rates in modern trials. A major contributing factor to excessive placebo response rates is represented by increased levels of expectations. Patients' high expectations of improvement related to exposure to study treatment may stem from a variety of factors, including any previous exposure to treatments similar to the treatment under investigation, similarities between study treatment and prior treatments which had been effective for that individual, emergence of physical and somatic symptoms on placebo interpreted by the patient as evidence for assignment to active treatment, the presence of an active, standard comparator in the study design, general knowledge of the efficacy of the study treatments, positive regard toward the institution(s) where the study is carried out or toward the study clinicians, and underlying cultural assumptions that biological/device/drug treatments are effective [8]. Similarly, clinicians' expectations of patients' improvement may derive from factors such as similarities between study treatment and prior treatments deemed to be effective for that individual, emergence of physical and somatic symptoms on placebo interpreted as evidence for patient's assignment to active treatment, the use of a standard comparator in the study design, strong beliefs that the study treatments are effective (e.g., through personal clinical experience), enthusiasm about the clinical trial, and desire to have a positive trial (e.g., to detect a signal of therapeutic activity) [8]. Papakostas and Fava [9] investigated whether the likelihood of receiving active treatment or placebo, a proxy of the degree of expectation of improvement, would itself influence clinical trial outcome. The data from 182 MDD clinical trials were pooled ($n = 36{,}385$), with response rates for drug and placebo being 53.8 and 37.3%. A meta-regression (random-effects) established that the probability of receiving placebo was an independent predictor of the risk ratio of responding to antidepressants vs. placebo. Specifically, a greater probability of receiving placebo predicted greater antidepressant-placebo "efficacy separation" [9].

LIMITATIONS OF CURRENT MECHANISTIC CLINICAL TRIALS IN NEUROPSYCHIATRIC DISORDERS

One of the major challenges in evaluating possible mechanisms related to the therapeutic effects of neuropsychiatric treatments is the inability to dissect what constitutes the actual effect of the treatment vs. what represents non-specific/placebo-like effects within a given patient. In fact, nested within the response to active neuropsychiatric therapy there is typically a placebo response that could be substantial. As previously pointed out [10], one may classify treated patients in a neuropsychiatric clinical trial based on each participant's propensity to respond to a given type of treatment. The "D−P−" population comprises patients who are not responsive to both active treatment (D) and inactive, placebo treatment (P). In neuropsychiatric trials in nonresistant populations, the D−P− group typically represents 30–50% of the populations. The "D+P+" population comprises patients who are responsive to either active (D) or placebo (P) treatments and represents the intrinsic placebo response rate of the population under investigation. The placebo/sham response rate varies across neuropsychiatric disorders but may be as high as 50% [11]. The third population ("D+P−") comprises patients who are responsive to active treatment but not to placebo/sham, and therefore represents the most informative group of patients. Therefore, in a mechanistic study of a neuropsychiatric treatment, the "D+P+" group is nested within the active treatment response and therefore either blunts the signal or even interferes with the true signal. The extent of such confounding effect can be very large when the drug-placebo difference in response rates is relatively small.

THE PLACEBO EFFECT AS A MAJOR CONFOUND IN NEUROPSYCHIATRIC TRIALS

The progressive rise in placebo/sham response rates over the past few decades has posed a significant challenge to neuropsychiatric clinical trials. The higher the placebo response rates, the smaller the effect sizes for active treatments. In fact, in a meta-analysis of monotherapy, FDA-approved antidepressant trials in MDD by Iovieno and Papakostas [12], a higher placebo response rate correlated with a lower risk ratio of responding to antidepressant vs. placebo ($p < 0.001$) and correlated with higher antidepressant response rates ($p < 0.001$), with the number needed to treat for response being ~4, 6, and 9 in trials with placebo response rates <30%, ≥30% and <40%, and ≥40%, respectively. This work implies that a trial's ability to detect a therapeutic effect is a function of how well managed the placebo response is. It is therefore essential that investigators adopt strategies aimed at reducing the placebo response and the following sections will describe some of these strategies.

THE IMPORTANCE OF CAREFUL SUBJECT SELECTION

As mentioned above, "professional patients", or duplicate subjects, are a common problem in clinical trials and may threaten the integrity of these studies. A number of digital platforms have been developed to identify duplicate subjects and allow investigators to exclude these subjects from trials. Furthermore, to avoid the issues of diagnostic misclassification and severity of illness grade inflation discussed above, it is critical to ensure that the right patients get into the trial. Patients enrolled in neuropsychiatric clinical trials may present with a heterogeneous group of symptoms representing several syndromes or subtypes, subsumed under the same diagnosis in the DSM-5 classification system [13]. Thereby, enrolled patients may not have the "valid" illness characteristics of interest to the particular study, as a "valid" patient should have the primary symptoms that the novel treatment is supposed to affect. We developed operational criteria to delineate a more symptom-specific and ecologically valid approach to the identification of the "valid" patient for clinical trials through an independent interview, called the SAFER interview. In a study from our group [14] we assessed whether these remotely performed multifaceted, centralized structured SAFER interviews can potentially enhance signal detection by ensuring that enrolled patients meet eligibility criteria by including a pooled analysis of nine studies that utilized the SAFER

interviews. We found that, overall, 15.33% of patients who had been deemed eligible at research sites were not eligible after the structured interviews, with the most common reason was that patients did not meet the study requirements in terms of treatment history. In these MDD trials utilizing the SAFER interviews as a tool to confirm eligibility, placebo response rates ranged between 13.0 and 27.3%, below the 30 to 40% average in antidepressant clinical trials, suggesting a benefit of the quality assurance provided by these interviews [14]. The main limitation of the SAFER approach concerns its use in conditions where a remote interview may not capture fully the clinical picture or in patients whose neuropsychiatric histories are quite recent and uncertain in nature.

CENTRALIZED RATINGS

Over the past few decades, we have seen the frequent adoption of centralized ratings to address the problems of clinician bias and measurement error in trials of MDD, schizophrenia, and Alzheimer's Disease. The use of centralized raters who are remotely linked to sites and/or patients and interview patients via videoconferencing or teleconferencing has been suggested as a way to improve interrater reliability and interview quality. A study comparing the effect of site-based and centralized ratings on patient selection and placebo response in subjects with MDD found that site-based raters' depression scores were significantly higher than centralized raters' at baseline and postbaseline but not at endpoint and that the mean placebo change for site raters was significantly greater than the mean placebo change for centralized raters [15]. Another MDD study comparing different approaches to the assessment of depression severity, including centralized ratings, found that patient self-ratings had greater depression severity at baseline than either site-based ratings or centralized ratings, but significantly lesser depression severity at the end of double-blind treatment than either site-based or centralized ratings [16]. However, the greater importance of the proper subject selection over centralized ratings has been evidenced by a study by Ratti et al. [17]. This study evaluated the efficacy in MDD of a selective NK1 antagonist orvepitant (GW823296) through two identical trials conducted in the US (Studies 733 and 833). In these trials, randomized patients with MDD assigned to double-blind treatment with orvepitant 30 mg/day, orvepitant 60 mg/day or placebo (1:1:1) were evaluated with the SAFER interview to confirm study eligibility in Study 733, while all the efficacy assessments were administered by the site raters, whereas in Study 833 all the assessments were performed by central raters, but there was no SAFER interview to independently confirm the inclusion criteria [17]. While Study 733 ($n = 328$) demonstrated efficacy on the primary endpoint for both doses of the active treatment (estimated drug-placebo differences of 30 mg: −2.41, 95% confidence interval (CI) (−4.50 to −0.31) $p = 0.0245$; 60 mg: −2.86, 95% CI (−4.97 to −0.75) $p = 0.0082$), Study 833 ($n = 345$) did not show statistical significance for either dose of the active treatment (estimated drug-placebo differences of 30 mg: −1.67, 95% CI (−3.73 to 0.39) $p = 0.1122$; 60 mg: −0.76, 95% CI (−2.85 to 1.32) $p = 0.4713$) [17]. These results support the hypothesis that careful subject selection, such as in the case of one aided by SAFER interview, followed by site ratings during the trial, provides a greater chance of detecting a therapeutic signal than an approach of central ratings throughout the trial with highly reliable raters.

THE IMPORTANCE OF INCORPORATING NOVEL OUTCOME MEASURES

When we examine some of the gold standards of the clinician-rated measurements of illness severity in conditions such as anxiety, depression, and pain, one is struck by the observation that most of these scales were developed in the 1950s, 1960s, 1970s, and 1980s as in the case of the Hamilton Rating Scale for Anxiety (HAM-A; [18]), the Hamilton Rating Scale for Depression (HAM-D; [19]), the Montgomery-Åsberg Depression Rating Scale (MADRS; Montgomery and Asberg, [20]), and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [21]. While the diagnosis and the characterization of these disorders has markedly evolved over time, the reliance on instruments that were developed decades ago and that reflect the nosology at the time they were developed may create issues when assessing improvement following novel drug or device treatments. An example of this is in the field of MDD, where there is now robust evidence that cognitive impairments, such as diminished memory, concentration, and executive function, are important symptoms of this disorder [22] and are often inadequately treated with standard antidepressants [23]. As novel drug treatments for MDD target neurogenesis and synaptogenesis and may therefore have distinct pro-cognitive effects, the fact that both the HAM-D and the MADRS minimally assess cognitive impairment, while focusing, at least in the case of the HAM-D, on symptoms that are far less common in MDD such as lack of insight and hypochondriacal concerns, may hamper one's ability to estimate the antidepressant effects of these new treatments. In fact, in a Phase 2 trial assessing the antidepressant efficacy in MDD of a neurogenesis-promoting compound, NSI-189, showed that 40 mg/day of this compound was significantly more efficacious than placebo with the Symptoms of Depression Questionnaire (SDQ), a well-validated 44-item self-rating depression measure capturing our more modern nosology of MDD [24], whereas the drug-placebo difference was not statistically significant on the MADRS [25]. The Major Depression Inventory (MDI) was developed as a well-validated clinician-rated instrument to cover the universe of depressive symptoms in DSM-IV major depression and in ICD-10 depression (mild, moderate, severe), with its summed total score being used as a measure of depression severity, outperforming older self-rating scales for depression [26]. It is therefore important that regulatory agencies begin to recognize the fact that more modern instruments to measure symptom changes in neuropsychiatric trials may be better suited to capture the true benefits of novel treatments than the old gold standards.

DESIGNS THAT MINIMIZE THE PLACEBO EFFECT

A number of methodologies and designs have been developed and implemented over the years to minimize the placebo effect. Some of these designs have focused on improving the precision of the trial assessments, others on improving the quality of the subjects recruited for the trial, others on the actual design of the trial.

(1) Rater training and monitoring

Since the lower the test-retest reliability of the assessments, the greater this artifactual response rate to placebo can be, rater training programs have proliferated over the past four decades, and such programs have become a standard in the pharmaceutical industry. However, there is very little or no evidence of their usefulness [8]. Although ongoing monitoring and feedback to raters has also been shown to decrease error rates and improve internal consistency of ratings [27], once again the usefulness of this approach in diminishing placebo response remains to be fully established.

(2) Independent assessment of subject eligibility

As mentioned earlier, the importance of the proper subject selection through independent assessments has been confirmed by a study by Ratti et al. [17], where the trial that utilized the independent SAFER interview to confirm study eligibility outperformed the trial where all the

assessments were performed by central raters. It is clear that reliance solely on the site raters to determine subjects' eligibility is quite risky, given the possibility of strong enrollment biases among them. The independent adjudication of the appropriateness and validity of the enrolled subject confers greater confidence that the inclusion criteria are truly met [16].

(3) Single-blind placebo lead-in

Placebo lead-in phases were initially designed to screen out patients who were likely to respond to placebo during the double-blind phase. However, such lead-in phases have been typically single-blinded periods of 1 or 2 weeks, during which the patients are unknowingly treated with placebo/sham, but their clinicians are aware of the placebo/sham treatment. The main limitation of this design is that it may potentially lead to clinician biases in underestimating improvement during the lead-in phase, and to verbal/non-verbal communications of low expectations of improvement to patients (e.g., emphasizing strongly to patients that no clinical effect is expected in the first weeks of treatment) [8]. Consistent with this view, an analysis of 75 double-blind trials which had been conducted among patients with MDD and had been published between January 1981 and December 2000, showed no statistically significant association between the proportion of responders to placebo in studies of patients with MDD and the presence of a lead-in phase [11]. These findings are consistent with those by Trivedi and Rush [28] who showed that the average placebo response rate of studies that used a placebo lead-in phase did not differ significantly from that of studies that did not use a placebo lead-in. They are also consistent with the findings of a more recent systematic review and meta-analysis suggesting that randomized clinical trials (RCTs) using single-blind placebo run-in periods yield smaller within-group changes across both placebo and drug groups compared with RCTs without the run-in periods [29]. Since the reduction in effect size across groups was equivalent in magnitude, studies using single-blind placebo run-in periods do not observe larger drug-placebo differences, suggesting that they do not increase trial sensitivity [29]. The authors concluded their review by stating that "given the resources and probable deception required and risk to external validity, the practice of using single-blind placebo run-in periods in RCTs of antidepressants should be ended [29]."

(4) Double-blind placebo lead-In

The double-blind fixed or variable placebo lead-in period was designed to address the limitations of the single-blind placebo lead-in. In these designs, both patients and clinicians are blinded to the presence and, in the case of its variable form, the length of the placebo lead-in period. Although all patients continue in the study (including placebo lead-in responders), the primary efficacy analysis prospectively excludes double-blind placebo lead-in responders. Faries et al. [30] conducted, using a variable double-blind placebo lead-in, two MDD trials of duloxetine, an FDA-approved antidepressant, and found that, while one trial resulted in an increased drug-placebo treatment difference, there was no effect on the treatment difference in the other study. Accordingly, a number of subsequent failed trials using the double-blind placebo lead-in [31–33] have markedly diminished the enthusiasm for this design. Interestingly, a comparison of the failed Forward-3 trial of the buprenorphine/samidorphan combination utilizing the double-blind placebo lead-in [33] with the three trials of the same combination utilizing the sequential parallel comparison design (SPCD) [34, 35] shows a markedly lower placebo response rate during the first stage of the Forward-3 trial, utilizing the double-blind placebo lead-in, compared to the first stage of the other three SPCD trials, and markedly greater drug-placebo difference in the three SPCD trials in the second stage compared to the Forward-3 trial, where the drug-placebo difference was negligible (0.2 points on the MADRS) [33–35]. The failure to enhance the drug-placebo difference with the double-blind placebo lead-in could be attributed to the functional unblinding of clinicians who may end up guessing the presence of a double-blind lead-in based on the absence of adverse events deemed related to drug treatment during the first period. In addition, the cost of implementing trials that may not improve the ability to detect a therapeutic signal and, at the same time, exclude a priori from the analysis responders to this double-blind lead-in phase that remain in the study is significant and might make this design unlikely to be adopted in future neuropsychiatric trials. In addition, as in the case of the single-blind placebo lead-in period, this approach also requires deception, making it perhaps less acceptable to sites.

(5) Sequential parallel comparison design (SPCD)

Twenty years ago, David Schoenfeld and I developed the sequential parallel comparison design (SPCD), a novel study design aimed at reducing both the placebo response rate and the sample size requirement, thereby markedly lowering the expense and time required to evaluate the efficacy of new therapeutic compounds [8]. In an SPCD trial, there are two stages of treatment, typically of equal duration. The first phase involves an unbalanced randomization between placebo and active treatment with more patients randomized to placebo (the typical randomization ratio for placebo/active is 3:1 or 2:1). As described in Fava et al. [8], at the end of Stage 1, placebo non-responders are re-randomized to either active treatment or placebo during Stage 2, whereas placebo responders during Stage 1 stay on placebo during Stage 2 and patients treated with drug during Stage 1 continue to stay on drug during Stage 2, although the data of these two groups in Stage 2 are not utilized (see Fig. 1).

Stage 1 of SPCD is aimed at comparing drug and placebo in a standard parallel comparison design fashion, with drug-placebo differences being expected to be standard, while generating a large cohort of placebo non-responders for the second stage (this is why the randomization ratio favors placebo). Stage 2 is aimed at comparing drug and placebo only among Stage 1 placebo non-responders, with drug-placebo differences being expected to be greater and placebo responses being markedly lower, as these patients on Stage 2 have already "failed placebo". In fact, Table 1 reports the changes in depression severity on placebo in Stages 1 and 2 of eleven SPCD trials in MDD that have been completed since the introduction of the design (randomized $n > 30$) and whose results have been either reported on clinicaltrials.gov or published in the literature, showing that the reduction in the degree of improvement on active treatment (51%) is markedly less than the reduction on placebo (37%). Such difference is likely to be an underestimate because some of the active treatments evaluated in those eleven trials were not necessarily effective, as in the case of riluzole. Since the lower the placebo response rates, the greater the effect sizes for active treatment, as shown in the meta-analysis of monotherapy, FDA-approved antidepressant trials in MDD by Iovieno and Papakostas [12], this marked reduction in the degree of improvement on placebo in Stage 2 is meant to enhance the ability to detect a therapeutic signal. A variant of SPCD involves the re-randomization of placebo responders as well, as implemented in an SPCD trial of the combination of dextromethorphan and quinidine in the treatment of agitation in Alzheimer's Disease patients [36]. Another variant of

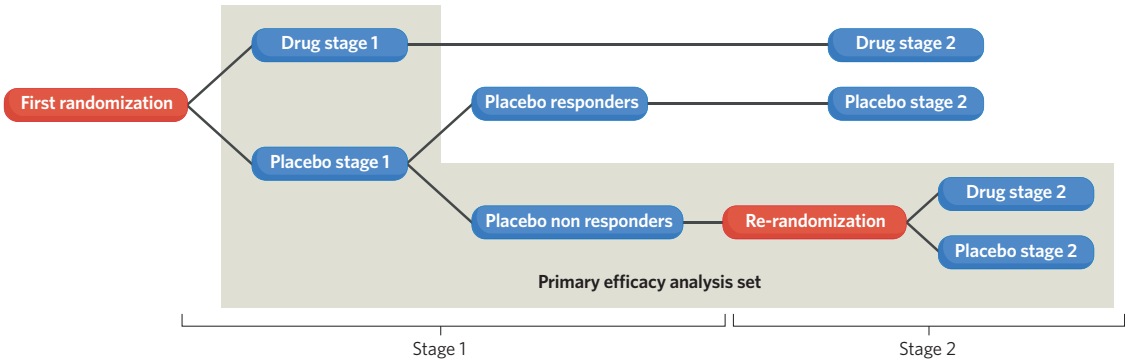

Fig. 1 Sequential Parallel Comparison Design (SPCD). Typical structure of SPCD, with two stages of equal duration and re-randomization at the end of Stage 1 of placebo non-responders to either staying on placebo or going into active drug. The primary efficacy analysis set includes all subjects in Stage 1 and only placebo non-responders in Stage 2.

Table 1. Degree of change on placebo vs. drug in Stage 2 vs. Stage 1 in MDD trials.

Clinical Trial.Gov #	Placebo MADRS change in Stage 1 (S1)	Placebo MADRS change in Stage 2 (S2)	Percent change S2/S1 placebo	Percent change S2/S1 drug
NCT01500200	9.9	2.1	21%	66%
NCT02158533	11.1	2.2	20%	30%
NCT02218008	8.1	2.1	26%	33%
NCT03188185	11.4	4.2	37%	34%
NCT00683852	8.1	3.3	41%	68%
NCT01998958	9.7	6.8	70%	82%
NCT02695472	10.8	2.0	19%	29%
NCT01204918	5.3	3.9	73%	66%
	HAMD change in S1	**HAMD change in S2**		
NCT03018340	7.5	2.1	28%	29%
NCT00321152	6.3	2.1	33%	72%
NCT00321152	4.4	1.7	39%	51%

Average percent change Stage 2/Stage 1: 37% for placebo 51% for drug.

SPCD involves having the patients assigned to drug in Stage 1 go on placebo, as implemented in an MDD study of the combination of buprenorphine and samidorphan [34]. When these two variants are combined, there is the opportunity to have an ABBA crossover trial nested within the SPCD trial, with the latter typically being a secondary or exploratory analysis (given the risk of carry-over effects for crossover trials).

A clear limitation of this design is that it may not be applicable to drugs or devices which require a significant dose titration or ramp up period or whose effects take 12 weeks or longer, given the need to double the duration of the trial. Since in conditions like pain, schizophrenia, and anxiety and mood disorders, 4 weeks of treatment are typically considered adequate, SPCD may be particularly well-suited for these populations, with the overall duration of the trial likely to be acceptable to patients. The great majority of the SPCD trials have been in Phase 2 programs, with only three trials thus far being in Phase 3 programs. A summary of a March 25, 2016 Workshop with the FDA focused on SPCD to discuss some of the regulatory issues is available on the Mass General Hospital website (https://mghcme.org/sprig/). One of the issues raised in that workshop was the type 1 error. It has been subsequently shown that the type I error rate is preserved for SPCD trials for both continuous [37, 38] and binary [39] outcomes.

The SPCD analysis pools the data from both phases in order to maximize power and reduce the required sample size. Data from all eligible subjects randomized in Stage 1 are utilized at least once, whereas data in Stage 2 are utilized only from placebo non-responders and are pooled with those from Stage 1. Over the past 20 years, many biostatisticians have reviewed SPCD and have recognized that there are a number of efficient methods of aggregating the outcome data that take into account the potential correlation of observations from subjects included in more than one stage, and that there are a number of valid test statistics that preserve the type 1 error rate. A large number of analytical methods for SPCD trials have been proposed by authors from academia, industry and FDA for both categorical [8, 37] and continuous data [38, 40, 41]. A variant of the SPCD design was mentioned in the original article by Fava et al. [8] and has been implemented in several multi-center trials, with pre-randomization to three sequences: drug-drug, placebo-drug, and placebo-placebo [42, 43].

DESIGNS THAT MAKE MECHANISTIC CLINICAL TRIALS MORE PRECISE

As mentioned earlier in this review, the "D+P+" population, which comprises patients who are responsive to either active (D) or placebo (P) treatments and represents the intrinsic placebo response rate of the population under investigation, is nested within the active treatment response and therefore either blunts the signal or even interferes with the true signal. Therefore, a mechanistic clinical trial examining possible changes in specific biomarkers and their relationship to clinical improvement is confounded by the fact that the degree of symptomatic improvement is not always caused by the treatment itself, but instead may be in part or wholly due to non-specific/placebo-like effects. Three approaches could be used to minimize the confounding effects of the placebo effects nested within the response to active treatment:

(1) To focus the study on a population that is relatively resistant to standard therapies, that is relatively enriched in the "D−P−" patients who are not responsive to both active treatment (D) and inactive, placebo treatment (P). Their history of resistance to treatment could be considered a proxy for lesser likelihood to respond to placebo and therefore one could assume that changes in biomarkers would actually reflect the true mechanism by which the treatment works. The main issue with this approach is that clinical trials on resistant populations have shown robust placebo responses, as in the case of augmentation trials in major depressive disorder (MDD) [44].

The presence of a substantial proportion of "D+P+" patients who are responsive to either active (D) or placebo (P) treatments will once again undermine the ability of this design to provide more precise assessments of the mechanisms involved in the response to active treatment.

(2) To adopt a study trial where patients are crossed over from placebo to active treatment, such as a simple ABBA crossover design or the SPCD. By focusing on patients who have not responded to placebo in the first period/stage of either the crossover design or SPCD, one can assume that having failed to respond to placebo could be considered a proxy for lesser likelihood to response to placebo in the second period/stage, therefore hypothesizing that changes in biomarkers would actually reflect the true mechanism by which the treatment works. The main issue with using a crossover design for this purpose is that those treated with placebo in the first period/stage are crossed over only to active treatment, not allowing an estimate of the placebo effects nested within the response to active treatment, since response to placebo can occur in a delayed fashion. Utilizing the SPCD for this purpose instead allows one to estimate the contribution of the degree by which placebo effects are nested within the response to active treatment, given the randomization of placebo non-responders at the end of the first period/stage to either active treatment or placebo/sham. For example, in an augmentation trial using SPCD in major depressive disorder (MDD), the Montgomery-Asberg Depression Rating Scale (MADRS) response rates were 5% for placebo and 50% for the active treatment (2 mg and 2 mg of the buprenorphine/samidorphan combination) [34]. Therefore, changes in biomarkers related to improvement in an active treatment so robustly greater than the one observed on placebo suggests that the changes on active treatment are primarily mechanistically related to the treatment under investigation.

(3) To pharmacologically manipulate the placebo effects nested within the active response so that they are minimized. A study by Peciña et al. [45] evaluated with a neuroimaging paradigm the neurochemical mechanisms underlying the formation of placebo effects in patients with MDD and found that reductions in depressive symptoms after 1 week of active placebo treatment, compared with the inactive placebo treatment, were associated with increased placebo-induced μ-opioid neurotransmission in a network of regions implicated in emotion, stress regulation, and the pathophysiology of MDD, namely, the subgenual anterior cingulate cortex, nucleus accumbens, midline thalamus, and amygdala. Their findings suggest co-administration of the active treatment with a μ-opioid antagonist may minimize the placebo effects nested within the active response. This approach has not been tested yet, but it is certainly very intriguing and represents a possible approach to making mechanistic clinical trials more precise.

DESIGNS THAT ADOPT PRECISION MEDICINE APPROACHES

There is a clear focus in adopting precision medicine approaches to neuropsychiatric trials. Many companies are adopting approaches to psychiatric drug development matching the right patient with the right drug/device, utilizing clinical phenotypes, standard biomarkers, and AI-derived brain biomarkers.

DESIGNS THAT LEVERAGE BIOMARKERS' ABILITY TO IDENTIFY MECHANISMS AND TO DEMONSTRATE THE IMPORTANCE OF SPECIFIC PATHWAYS

The robust advances in the neuroscience of neuropsychiatric disorders have led to the identification of biomarkers that may be tied specifically to the mechanism of action of the treatment. An example of this is the use of positron emission tomography (PET), a medical imaging technique, in clinical trials of novel drugs against Alzheimer's disease (AD). PET data contributed to the conditional approval in 2021 of aducanumab, an antibody directed toward amyloid-beta (Aβ) aggregates, by showing a dose-dependent reduction in brain amyloid after treatment [46]. It has also been suggested that, in parallel to clinical studies, preclinical studies in animal models of Aβ pathology may also benefit from PET as a tool to detect target engagement and treatment effects of anti-Aβ drug candidates [46]. Many other designs utilizing both peripheral and central biomarkers have been developed to demonstrate "target engagement" of novel neuropsychiatric treatments and it is likely that we are going to see a marked growth of these approaches in Phase 1 trials, with the goal of optimizing the target dose.

SUMMARY

This review has outlined some of the major limitations of the current approaches to neuropsychiatric treatment development, with a focus on those factors that negatively affect both efficacy and mechanistic clinical trials in neuropsychiatric disorders. The remarkable growth of placebo responses in these trials is certainly a major culprit, but other factors play a key role as well, such as the excessive reliance on subjective and obsolete endpoints and the inappropriate enrollment of subjects. Another important issue is that standard neuropsychiatric clinical trials do not fully leverage the tremendous growth of neuroscience research, leading to the development of novel therapies for neuropsychiatric disorders. As mentioned earlier, the emergence of new molecular targets, together with the development of new approaches to neuropsychiatric diseases, involving both devices and gene and cell therapies, are creating the need to improve the efficiency of mechanistic and/or efficacy clinical trials. This review has also presented some of the methodological approaches that may improve the overall performance of our efficacy and mechanistic neuropsychiatric trials, including the adoption of novel study designs such as SPCD and independent confirmation of the appropriateness of subjects' enrollment.

REFERENCES

1. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. Am J Psychiatry. 2009;166:599–607. https://doi.org/10.1176/appi.ajp.2008.08071027.
2. Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. Symptom severity and the generalizability of antidepressant efficacy trials: changes during the past 20 years. J Clin Psychopharmacol. 2009;36:153–6. https://doi.org/10.1097/JCP.0000000000000466.
3. Mundt JC, Greist JH, Jefferson JW, Katzelnick DJ, DeBrota DJ, Chappell PB, et al. Is it easier to find what you are looking for if you think you know what it looks like? J Clin Psychopharmacol. 2007;27:121–5. https://doi.org/10.1097/JCP.0b013e3180387820.
4. Tuttle AH, Tohyama S, Ramsay T, Kimmelman J, Schweinhardt P, Bennett GJ, et al. Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. Pain. 2015;156:2616–26. https://doi.org/10.1097/j.pain.0000000000000333.
5. Dunlop BW, Thase ME, Wun CC, Fayyad R, Guico-Pabia CJ, Musgnung J, et al. A meta-analysis of factors impacting detection of antidepressant efficacy in clinical trials: the importance of academic sites. Neuropsychopharmacology. 2012;37:2830–6. https://doi.org/10.1038/npp.2012.153.
6. Dotson S, Mischoulon D, Lee H, Fava M. Rising placebo response rates threaten the validity of antipsychotic meta-analyses. Ann Clin Psychiatry. 2019;31:249–59.
7. Leucht S, Arbter D, Engel RR, Kissling W, Davis JM. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. Mol Psychiatry. 2009;14:429–47. https://doi.org/10.1038/sj.mp.4002136.
8. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom. 2003;72:115–27. https://doi.org/10.1159/000069738.

9. Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. Eur Neuropsychopharmacol. 2009;19:34–40. https://doi.org/10.1016/j.euroneuro.2008.08.009.
10. Fava M. Implications of a biosignature study of the placebo response in major depressive disorder. JAMA Psychiatry. 2015;72:1073–4. https://doi.org/10.1001/jamapsychiatry.2015.1727.
11. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial and growing. JAMA. 2002;287:1840–7.
12. Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. J Clin Psychiatry. 2012;73:1300–6. https://doi.org/10.4088/JCP.11r07485.
13. Targum SD, Pollack MH, Fava M. Redefining affective disorders: relevance for drug development. CNS Neurosci Ther. 2008;14:2–9. https://doi.org/10.1111/j.1527-3458.2008.00038.x.
14. Freeman MP, Pooley J, Flynn MJ, Baer L, Mischoulon D, Mou D, et al. Guarding the gate: remote structured assessments to enhance enrollment precision in depression trials. J Clin Psychopharmacol. 2017;37:176–81. https://doi.org/10.1097/JCP.0000000000000669.
15. Kobak KA, Leuchter A, DeBrota D, Engelhardt N, Williams JB, Cook IA, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol. 2010;30:193–7. https://doi.org/10.1097/JCP.0b013e3181d20912.
16. Targum SD, Wedel PC, Robinson J, Daniel DG, Busner J, Bleicher LS, et al. A comparative analysis between site-based and centralized ratings and patient self-ratings in a clinical trial of Major Depressive Disorder. J Psychiatr Res. 2013;47:944–54. https://doi.org/10.1016/j.jpsychires.2013.02.016.
17. Ratti E, Bettica P, Alexander R, Archer G, Carpenter D, Evoniuk G, et al. Full central neurokinin-1 receptor blockade is required for efficacy in depression: evidence from orvepitant clinical studies. J Psychopharmacol. 2013;27:424–34. https://doi.org/10.1177/0269881113480990.
18. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol. 1959;32:50–5.
19. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56–62. https://doi.org/10.1136/jnnp.23.1.56.
20. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382–9. https://doi.org/10.1192/bjp.134.4.382.
21. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15:1833–40.
22. Pan Z, Park C, Brietzke E, Zuckerman H, Rong C, Mansur RB, et al. Cognitive impairment in major depressive disorder. CNS Spectr. 2019;24:22–9. https://doi.org/10.1017/S1092852918001207.
23. Fava M, Graves LM, Benazzi F, Scalia MJ, Iosifescu DV, Alpert JE, et al. A cross-sectional study of the prevalence of cognitive and physical symptoms during long-term antidepressant treatment. J Clin Psychiatry. 2006;67:1754–9. https://doi.org/10.4088/jcp.v67n1113.
24. Pedrelli P, Blais MA, Alpert JE, Shelton RC, Walker RS, Fava M. Reliability and validity of the Symptoms of Depression Questionnaire (SDQ). CNS Spectr. 2014;19:535–46. https://doi.org/10.1017/S1092852914000406.
25. Papakostas GI, Johe K, Hand H, Drouillard A, Russo P, Kay G, et al. A phase 2, double-blind, placebo-controlled study of NSI-189 phosphate, a neurogenic compound, among outpatients with major depressive disorder. Mol Psychiatry. 2020;25:1569–79. https://doi.org/10.1038/s41380-018-0334-8.
26. Bech P, Timmerby N, Martiny K, Lunde M, Soendergaard S. Psychometric evaluation of the Major Depression Inventory (MDI) as depression severity scale using the LEAD (Longitudinal Expert Assessment of All Data) as index of validity. BMC Psychiatry. 2015;15:190. https://doi.org/10.1186/s12888-015-0529-3.
27. Daniel DG, Busner J, McNamara C. Ongoing monitoring and feedback decreases error rates and improves internal consistency of PANSS ratings in an international clinical trial. Poster presentation at the 2010 Autumn International Society for CNS Clinical Trials and Methodology (ISCTM) Conference (2010), Baltimore, MD, October, 2010.
28. Trivedi MH, Rush AJ. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? Neuropsychopharmacology. 1994;11:33–43.
29. Scott AJ, Sharpe L, Quinn V, Colagiuri B. Association of single-blind placebo run-in periods with the placebo response in randomized clinical trials of antidepressants: a systematic review and meta-analysis. JAMA Psychiatry. 2022;79:42–9. https://doi.org/10.1001/jamapsychiatry.2021.3204.
30. Faries DE, Heiligenstein JH, Tollefson GD, Potter WZ. The double-blind variable placebo lead-in period: results from two antidepressant clinical trials. J Clin Psychopharmacol. 2001;21:561–8. https://doi.org/10.1097/00004714-200112000-00004.
31. Ball SG, Ferguson MB, Martinez JM, Pangallo BA, Nery ES, Dellva MA, et al. Efficacy outcomes from 3 clinical trials of edivoxetine as adjunctive treatment for patients with major depressive disorder who are partial responders to selective serotonin reuptake inhibitor treatment. J Clin Psychiatry. 2016;77:635–42. https://doi.org/10.4088/JCP.14m09619.
32. Targum SD, Cameron BR, Ferreira L, MacDonald ID. An augmentation study of MSI-195 (S-adenosylmethionine) in major depressive disorder. J Psychiatr Res. 2018;107:86–96.
33. Zajecka JM, Stanford AD, Memisoglu A, Martin WF, Pathak S. Buprenorphine/samidorphan combination for the adjunctive treatment of major depressive disorder: results of a phase III clinical trial (FORWARD-3). Neuropsychiatr Dis Treat. 2019;15:795–808. https://doi.org/10.2147/NDT.S199245.
34. Fava M, Memisoglu A, Thase ME, Bodkin JA, Trivedi MH, de Somer M, et al. Opioid modulation with buprenorphine/samidorphan as adjunctive treatment for inadequate response to antidepressants: a randomized double-blind placebo-controlled trial. Am J Psychiatry. 2016;173:499–508. https://doi.org/10.1176/appi.ajp.2015.15070921.
35. Fava M, Thase ME, Trivedi MH, Ehrich E, Martin WF, Memisoglu A, et al. Opioid system modulation with buprenorphine/samidorphan combination for major depressive disorder: two randomized controlled studies. Mol Psychiatry. 2020;25:1580–91. https://doi.org/10.1038/s41380-018-0284-1.
36. Cummings JL, Lyketsos CG, Peskind ER, Porsteinsson AP, Mintzer JE, Scharre DW, et al. Effect of dextromethorphan-quinidine on agitation in patients with Alzheimer disease dementia: a randomized clinical trial. JAMA. 2015;314:1242–54. https://doi.org/10.1001/jama.2015.10214.
37. Ivanova A, Qaqish B, Schoenfeld D. Optimality, sample size and power calculations for the sequential parallel comparison design. Stat Med. 2011;30:2793–803.
38. Chen Y, Yang Y, Hung H, Wang S. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. Contemp Clin Trials. 2011;32:592–604.
39. Silverman RK, Ivanova A, Fine J. Sequential parallel comparison design with binary and time-to-event outcomes. Stat Med. 2018;37:1454–66. https://doi.org/10.1002/sim.7635.
40. Tamura R, Huang X. An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. Clin Trials. 2007;4:309–17.
41. Doros G, Pencina M, Rybin D, Meisner A, Fava M. A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. Stat Med. 2013;32:2767–89. https://doi.org/10.1002/sim.5728.
42. Fava M, Mischoulon D, Iosifescu D, Witte J, Pencina M, Flynn M, et al. A double-blind, placebo-controlled study of aripiprazole adjunctive to antidepressant therapy among depressed outpatients with inadequate response to prior antidepressant therapy (ADAPT-A Study). Psychother Psychosom. 2012;81:87–97. https://doi.org/10.1159/000332050.
43. Papakostas GI, Shelton RC, Zajecka JM, Etemad B, Rickels K, Clain A, et al. L-methylfolate as adjunctive therapy for SSRI-resistant major depression: results of two randomized, double-blind, parallel-sequential trials. Am J Psychiatry. 2012;169:1267–74. https://doi.org/10.1176/appi.ajp.2012.11071114.
44. Papakostas GI, Shelton RC, Smith J, Fava M. Augmentation of antidepressants with atypical antipsychotic medications for treatment-resistant major depressive disorder: a meta-analysis. J Clin Psychiatry. 2007;68:826–31. https://doi.org/10.4088/jcp.v68n0602.
45. Peciña M, Bohnert AS, Sikora M, Avery ET, Langenecker SA, Mickey BJ, et al. Association between placebo-activated neural systems and antidepressant responses: neurochemistry of placebo effects in major depression. JAMA Psychiatry. 2015;72:1087–94. https://doi.org/10.1001/jamapsychiatry.2015.1335.
46. Syvänen S, Meier SR, Roshanbin S, Xiong M, Faresjö R, Gustavsson T, et al. PET imaging in preclinical anti-Aβ drug development. Pharm Res. 2022;39:1481–96. https://doi.org/10.1007/s11095-022-03277-z.

COMPETING INTERESTS

MF's 3-year disclosures of competing interests—Research Support: AbbVie; Acadia Pharmaceuticals; Aditum Bio Management Company, LLC; Allergan, Alkermes, Inc.; Altimate Health Corporation; Alto Neuroscience, Inc.; Ancora Bio, Inc.; Angelini S.p.A; Aptinyx; Arbor Pharmaceuticals LLC; Avanir Pharmaceuticals Inc.; Axsome; Benckiser Pharmaceuticals, Inc.; BioClinica, Inc.; Biogen; BioHaven; BioShin Limited; Cambridge Science Corporation; Centrexion Therapeutics Corporation; Cerecor; Cybin IRL Limited; Damona Pharmaceuticals; EmbarkNeuro; Eliem Therapeutics LTD; Gate Neurosciences, Inc.; GenOmind, LLC; Gentelon, LLC; Gerbera Therapeutics, Inc.; Happify; Johnson & Johnson; Lundbeck Inc.; Marinus Pharmaceuticals; Medpace, Inc.; Methylation Sciences, Inc.; Millennium Pharmaceutics, Inc.; Minerva Neurosciences; Neuralstem; Neurocrine Biosciences, Inc.; NeuroRX Inc.; Novaremed; Novartis; Otsuka; Pfizer; Premiere Research International; Praxis Precision Medicines; Protagenic Therapeutics, Inc.; Relmada Therapeutics Inc.; Reckitt; Shenox Pharmaceuticals; Stanley Medical Research Institute (SMRI); Taisho; Takeda; University of Michigan;

Vistagen; WinSanTor, Inc.; Xenon Pharmaceuticals, Inc.; National Institute of Drug Abuse (NIDA); National Institutes of Health (NIH); National Institute of Mental Health (NIMH); and PCORI. MF has not done any personal consulting. Any consulting he has done has been on behalf of Massachusetts General Hospital, except for Revival Therapeutics and Sensorium Therapeutics. Speaking/Publishing: lecture given at Global Medical Education, Inc. Mood Disorders Summit, November 2020. Stock/Other Financial Options: Equity Holdings: Compellis; Neuromity; Psy Therapeutics; Revival Therapeutics; Sensorium Therapeutics Royalty/patent, other income: patents for Sequential Parallel Comparison Design (SPCD) (US_7840419, US_7647235, US_7983936, US_8145504, US_8145505), all expired; patent application for a combination of Ketamine plus Scopolamine in Major Depressive Disorder (MDD), licensed by MGH to BioHaven. Patents for pharmacogenomics of Depression Treatment with Folate (US_9546401, US_9540691), licensed by MGH to Nestle. Copyright for the MGH Cognitive & Physical Functioning Questionnaire (CPFQ), Sexual Functioning Inventory (SFI), Antidepressant Treatment Response Questionnaire (ATRQ), Discontinuation-Emergent Signs & Symptoms (DESS), Symptoms of Depression Questionnaire (SDQ), and SAFER; Belvior; Lippincott, Williams & Wilkins; Wolters Kluwer; World Scientific Publishing Co. Pte. Ltd. All lifetime disclosures can be viewed online by navigating to: https://mghcme.org/maurizio-fava-bio-disclosure/.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Maurizio Fava.