Supplementary Online Content

Rutherford BR, Pott E, Tandler JM, Wall MM, Roose SP, Lieberman JA. A meta-analysis of placebo response in antipsychotic clinical trials. *JAMA Psychiatry*. Published online October 8, 2014. doi:10.1001/jamapsychiatry.2014.1319.

eAppendix

eReferences

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix

1. SEARCH STRATEGY

The index terms "schizophrenia—drug therapy," "schizophrenia—drug effects," and "antipsychotic agents," in addition to the class and individual generic names of antipsychotic medications were combined using the 'or' operator. Included antipsychotic medications were either FDA-approved at the time the study was conducted or else subsequently received FDA-approval. Limiting these results to humans, English language articles, publication year 1960 or later, age group ≥ 18, and publication types including clinical trials, comparative studies, controlled clinical trials, meta-analysis, multi-center study, randomized controlled trial, or review, yielded 9,057 journal articles. Three authors (BRR, EP, and JT) conducted a preliminary review to rule out those which were obviously not clinical trials, resulting in 2,326 titles. These were then sequentially examined from titles to abstracts and finally paper texts to determine whether they met inclusion or exclusion criteria (see **Figure 1**).

2. DATA EXTRACTION

We classified medication dosing as 'effective' or 'low dose' by utilizing a modified version of the classification strategy employed by Woods et al (2005) in their study of control groups in schizophrenia trials (e.g., we considered daily effective doses to be risperidone ≥ 6mg, olanzapine ≥ 10mg, quetiapine ≥ 300mg, ziprasidone ≥ 120mg, aripiprazole ≥ 15mg, haloperidol ≥ 5mg) (1).

In addition to our primary method of calculating standardized mean change, we followed Agid et al (2013) in calculating standardized change scores by dividing the mean change for a given treatment cell by its pre-treatment standard deviation, and then we compared the results obtained with the first method (2). Calculation of standardized mean change using pre-treatment standard deviations was a secondary method, because the majority of older studies did not provide information on variability that could be used to calculate standard deviations. Finally, we separated studies by the outcome measure used (primarily the PANSS and BPRS) and individually analyzed these subgroups which all used the same outcome measure, again comparing the results obtained with the above methods.

3. CLINICAL CHARACTERISTICS OF INCLUDED PATIENTS AND METHODOLOGICAL FEATURES OF STUDIES.

As shown in **Table 2**, study duration [$F(3,292) = 7.21$, $p < 0.001$] and drop-out rate [$F(3,258) = 13.82$, $p < 0.001$] significantly differed based on treatment assignment. Effective dose medication was associated with significantly decreased drop-out relative to placebo ($t = 4.98$, df 219, $p < 0.001$) and low dose medication ($t = 4.37$, df 208, $p < 0.001$), while intramuscular medication was associated with decreased drop-out relative to low dose medication ($t = 2.04$, df 39, $p = 0.048$). Intramuscular medication cells had significantly longer duration compared to placebo ($t = -3.41$, df 61, $p = 0.001$), low dose medication ($t = -4.83$, df 47, $p < 0.001$), and effective dose medication ($t = -4.29$, df 230, $p < 0.001$). The number of study visits ($F(3,292) = 1.63$, $p = 0.183$), mean age ($F(3,274) = 1.70$, $p = 0.167$), PANSS score ($F(3,150) = 1.41$, $p = 0.243$), baseline BPRS score ($F(3,92) = 2.31$, $p = 0.082$) were not significantly different between groups.

4. CORRELATIONS BETWEEN STUDY VARIABLES AND PUBLICATION YEAR

We examined variables related to year of publication in order to investigate what accounts for the associations between publication year and medication/placebo response. Study duration ($r = -0.26$, $p < 0.001$) significantly decreased over time, while the number of treatment arms ($r = 0.31$, $p < 0.001$), number of study visits ($r = 0.18$, $p = 0.002$), sample size ($r = 0.43$, $p < 0.001$), and drop-out rate ($r = 0.49$, $p < 0.001$) significantly increased over time. Treatment setting became significantly more outpatient-oriented from 1960 to the present (Chi-square for decade of publication year by treatment setting = 70.4, df 4, $p < 0.001$). Excluding one outlying recent study enrolling severely ill patients (3), baseline symptom severity of antipsychotic trial participants decreased significantly from 1960 to the present ($N = 282$, $r = -0.13$, $p = 0.030$). To illustrate these significant differences, the average 1960s RCT was primarily inpatient, 14.2 ± 7.2 weeks in duration, had 2.7 ± 0.8 treatment arms, 4.5 ± 2.1 study visits, enrolled 17.9 ± 8.5 patients per treatment cell, and had a drop-out rate of 17.2 ± 26.0. In contrast, the average 2000s RCT was primarily outpatient, 8.4 ± 4.5 weeks in duration, had 3.3 ± 1.4 treatment arms, 5.9 ± 2.2 study visits, enrolled 124.4 ± 79.8 patients per treatment cell, and had a drop-out rate of 39.5 ± 15.1. Thus, over time antipsychotic trials have grown larger (requiring more study sites), become shorter in duration, randomized proportionately more subjects to active medication vs. placebo, offered more intensive follow-up, and entailed much higher drop-out.

5. REPEATED MULTILEVEL MODELS USING DIFFERENT METHODS OF CALCULATING STANDARDIZED MEAN CHANGE

To investigate the robustness of these findings across different methods of standardizing mean treatment change, we repeated the above analyses after recalculating standardized mean change using the standard deviation statistics for each treatment cell. Only 51/106 (48.1%) of the sample provided pre-treatment standard deviations or information on variability that could be used to calculate standard deviations. We found the results obtained by computing standardized mean change by dividing the pre-post mean difference by the pre-treatment standard deviation were highly correlated with the results of our preferred method of calculating mean change ($r = 0.80$, $p < 0.001$). Additionally, the overall pattern of results obtained by using pre-treatment standard deviations was similar to the above. Dividing the sample into subgroups based on the rating scale used resulted in 48/106 (45.3%) supplying PANSS data and 53/106 (50.0%) supplying BPRS data (18/106 provided data using both scales). Repeating the multilevel models in each subgroup did not change the overall pattern of results.

6. RELEVANCE OF RESULTS FOR CLINICAL TRIAL DESIGN

The divergent patterns over time of placebo response and effective medication response are surprising and appear inconsistent with the "assumption of additivity" upon which interpretations of drug-placebo differences in RCTs are based (4). That is, medication response in an RCT is typically construed as a linear combination of the specific medication effect with placebo response (i.e., placebo response is the same in the medication and placebo groups). Based on additivity, one would predict medication response would increase over time in parallel with placebo response, since the component of medication response attributable to placebo is increasing while the specific medication effect remains constant. Our findings suggest there may be a more complex relationship between medication response and placebo response than simple additivity. Further research is required to elucidate the precise nature of this relationship, but one possible explanation is that a ceiling is placed on medication response in a given study by the proportion of the sample which is non-responsive to medications (e.g., due to misdiagnosis, treatment refractoriness, etc.). If these factors differentially affect medication response relative to placebo response and have been increasing over time, this may result in a declining ceiling on medication response while allowing placebo response to increase.

Single-blind placebo lead-ins did not improve signal detection in the studies we examined. Although the presence of lead-ins significantly decreased the average pre-post treatment change observed ($t = -2.29$, $df\ 93$, $p = 0.024$), there were no interactions with treatment assignment to suggest that placebo response was preferentially reduced. Lead-ins appear to reduce pre-post change similarly in both the drug and placebo groups and thus may not be beneficial from a study design perspective, particularly given the implications of lead-ins with respect to required sample sizes.

eReferences

1. Woods SW, Gueorguieva RV, Baker CB, Makuch RW. Control group bias in randomized atypical antipsychotic medication trials for schizophrenia. Arch Gen Psychiatry. 2005;62(9):961-970.
2. Agid O, Siu CO, Potkin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. Am J Psychiatry. 2013;170(11):1335-1344.
3. Pandina GJ, Lindenmayer JP, Lull J, et al. A randomized, placebo-controlled study to assess the efficacy and safety of 3 doses of paliperidone palmitate in adults with acutely exacerbated schizophrenia. J Clin Psychopharmacol. 2010;30(3):235-244.
4. Kirsch I. Are drug and placebo effects in depression additive? Biol Psychiatry. 2000;47(8):733-735.